Panoramic Image Stitching Algorithm based on SIFT Features

Li Wang[1, a], Wei Wang[1, b], and Boni Liu[1, c]

[1]Department of Electronic Engineering, Xi'an Aeronautical University, Xi'an Shaanxi 710077, China.

[a]wangli871016@163.com, [b]weiiiwang@qq.com, [c]271629953@qq.com

Abstract

A panorama image stitching algorithm based on scale-invariant feature transformation (SIFT) feature points is proposed in this paper. Due to the limited field of view of the camera, multiple images with different angles of coincident areas are required to be stitched together to obtain a complete panoramic image. Firstly, SIFT algorithm is used to extract feature points, and the improved random sampling consistency (RANSAC) algorithm is used to filter feature points and calculate homography matrix. Homography matrix is used for image stitching, and linear weighting function method is used for image fusion. Experimental results show that this method can achieve the stitching of panoramic images effectively.

Keywords

Panoramic image stitching, SIFT feature points, RANSAC algorithm, Homography matrix, Image fusion.

1. Introduction

In view of the contradiction between the resolution and the field of view in the visual system, in order to ensure higher resolution, the scene area corresponding to a single image is usually small, which is not conducive to accurate and comprehensive observation and analysis of the scene. With the development of computer vision technology and the improvement of photographic quality, multiple cameras can be used to collect scenes, and the image stitching technology produces a wider field of view and higher resolution than a single image. Image stitching technology has become a hot spot in the field of image processing [1] [2]. Image stitching technology is to seamlessly connect multiple related overlapping images to obtain a wide-angle panoramic image. There are three main image registration methods: gray information registration method [3], feature registration method [4] and transform domain registration method [5].

Image stitching mainly includes two steps: image registration and mosaic, the core of which is image registration. The scale-invariant feature transformation (SIFT) [4] [6] in the feature registration method is used to describe the local features in the image. The algorithm solves the partial occlusion, rotation, scale scaling, and viewpoint changes of the scene, effectively improves the accuracy of feature matching. The thesis uses SIFT algorithm to extract the feature points, and uses the random sampling consistency (RANSAC) algorithm to filter the matching points and calculate the homography matrix. Finally, the linear weighting function method is used to achieve seamless connection and smooth transition.

The paper is organized as follows. The first part briefly introduces panoramic image stitching technology, the second part presents SIFT feature point extraction process; the third part gives the process of feature point matching using RANSAC algorithm and image fusion, as well as the algorithm flow. The fourth part gives the experimental results, and gives the conclusion in the fifth part.

2. Features Extraction using SIFT

The idea of SIFT feature matching is to firstly perform feature detection in scale space and determine the position of key points. Then, the main direction of the gradient in the neighborhood square of the key point is used as the direction feature of the key point, achieving the independence of SIFT features on scale and direction [7].

2.1 Gaussian Pyramid Constructing

The SIFT operator uses the Gaussian difference function as the convolution kernel to construct the scale space to form the difference of Gaussian (DoG) [8], which can accurately extract the extreme points in the image to be stitched, and maintain the specialness and stability of the features. The Gaussian convolution kernel is the only linear kernel that achieves scale transformation [9], expressed as:

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

Where, (x,y) represents the two-dimensional coordinates of the image, σ represents the smoothness of the image, also known as the scale factor. The value σ corresponds to different layers in the scale space. The smaller the value σ, the higher the resolution layer, and the more detailed image information.

The scale space of the two-dimensional image is expressed as:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{2}$$

Where, $I(x,y)$ represents the grayscale information of the image.

The Gaussian difference operator is composed of Gaussian kernel functions with different scales. The DoG can be described as:

$$\begin{aligned} D(x,y,\sigma) &= \left(G(x,y,k\sigma)-G(x,y,\sigma)\right)*I(x,y) \\ &= \left(G(x,y,k\sigma)*I(x,y)\right)-\left(G(x,y,\sigma)*I(x,y)\right) \\ &= L(x,y,k\sigma)-L(x,y,\sigma) \end{aligned} \tag{3}$$

Where, the coefficient k is the adjacent two scale ratios.

The image $I(x,y)$ is convolved with Gaussian kernels with different scales to obtain a set of scale spaces of the image, and this set of images is used as the first level of the pyramid (octave). Choose the second scale image and down sample it at 2 times to obtain the bottom image of the second level of the pyramid. In the same way, convolution and sampling are performed sequentially upward, thereby forming a complete image of the Gaussian pyramid. The DoG image is formed by subtracting each adjacent Gaussian image, as shown in Fig. 1.

Each pixel of the middle layer needs to be compared with 26 pixels, including 8 pixels in its neighborhood and 18 pixels in the adjacent layer, to determine whether the pixel is a maximum value or a minimum value. The schematic diagram of extreme point detection is shown in Fig. 2.

2.2 Feature Point Location

Since the obtained maximum and minimum points are not all available, they can only be used as rough selection points, and the available feature points should be accurately extracted from these feature points. The SIFT operator selects high-contrast and non-boundary points by establishing a three-dimensional quadratic function.

Assume that the coarse selection point is $P(x,y,\sigma)$, then the Taylor expansion of the DoG function at point P is:

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

The above formula derivates P and equals 0, then formula (4) becomes:

$$\hat{P}=\frac{\partial^2 D^{-1}}{\partial P^2}\cdot\frac{\partial D}{\partial P} \tag{5}$$

Substitute $\hat{p}$ to $D(P)$ to obtain $D(\hat{P})$, which is used to represent the contrast of the coarse selection point. When $D(\hat{P}) > 0.03$ is satisfied, point P is determined to be an accurate feature point.

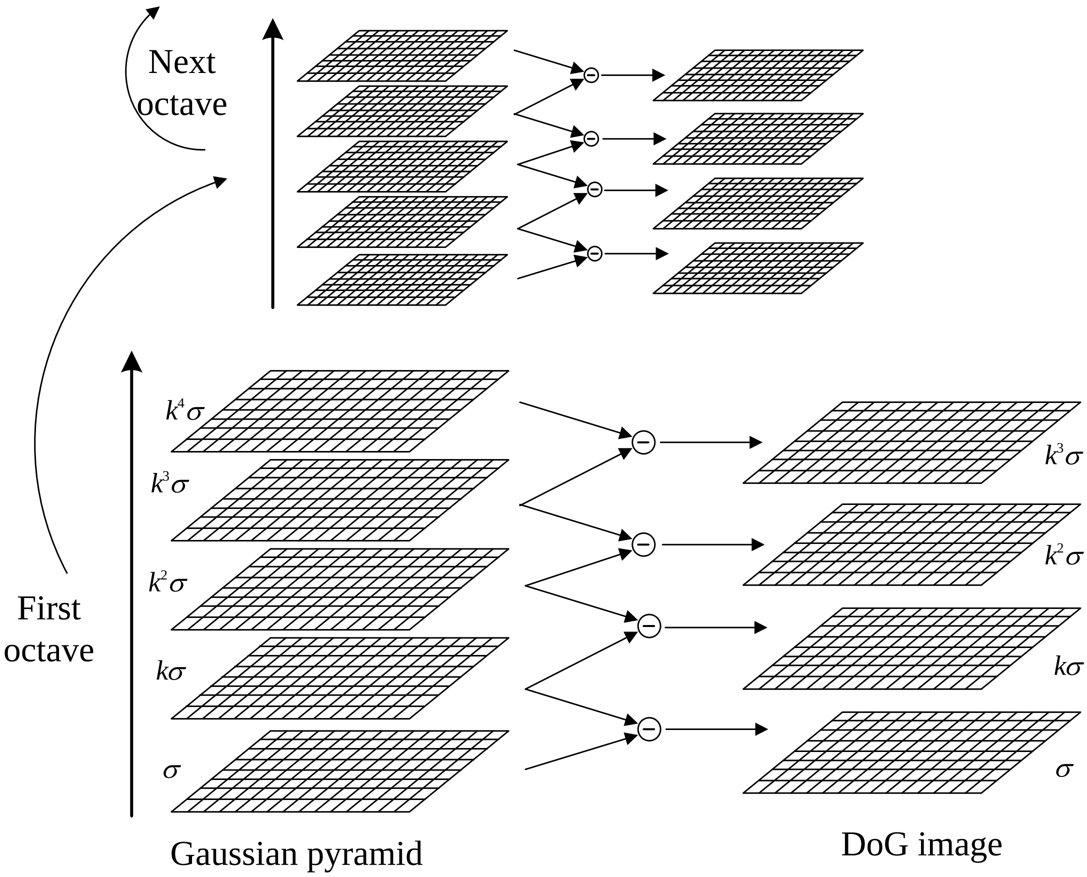


Fig. 1 Gaussian pyramid image and DoG image

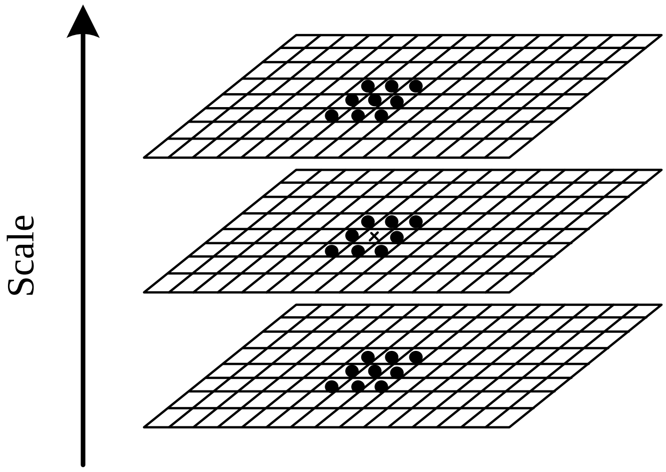


Fig. 2 Extreme point detection in Gaussian difference scale space

2.3 Feature Point Direction Assignment

The rotation invariance of SIFT can be achieved by adding direction information to each feature point, through the gradient magnitudes and gradient direction between the four neighboring pixels of the feature point [9]. The gradient magnitudes of the four neighbor pixels are:

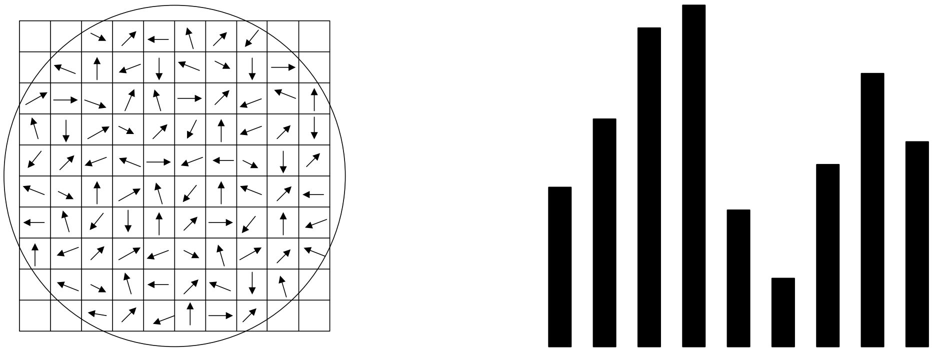
$$t_{x,y} = \sqrt{\left(I_{x+1,y} - I_{x-1,y}\right)^2 + \left(I_{x,y+1} - I_{x,y-1}\right)^2} \tag{6}$$

The gradient direction is:

$$\theta_{x,y} = \tan^{-1}\left(\frac{I_{x,y+1} - I_{x,y-1}}{I_{x+1,y} - I_{x-1,y}}\right) \tag{7}$$

Where, I represents the gray value of four points in the neighborhood of the feature point.

The main direction of the feature points can be represented by Fig. 3. A window with a constant range is established, composed of the feature points and its neighborhood. Each pixel is represented by an arrow, and the direction and length of the arrow represents the gradient direction and gradient magnitude corresponding the pixel. Count the gradient direction in the 0-360 degree, and establish a histogram of the gradient direction to indicate a new direction at interval of 10 degree, resulting in 36 new directions. Take the gradient direction represented by the peak value as the main direction of the feature point, and use the secondary peak direction in the gradient direction histogram as the auxiliary direction of the feature point to improve the robustness of the SIFT operator.



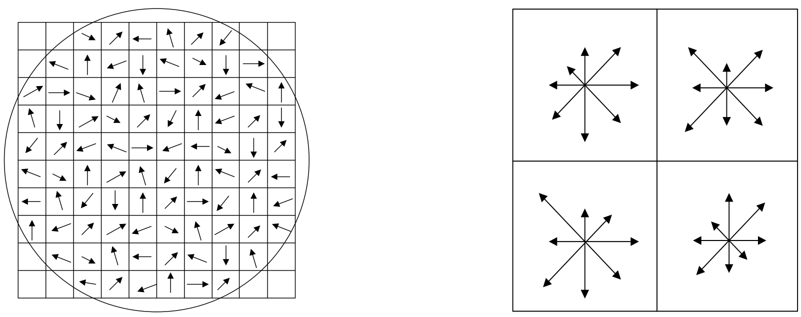
(a) The gradient magnitude and direction of feature points in the neighborhood

(b) The histogram of the gradient direction

Fig. 3 Schematic diagram of the main directions of feature points

2.4 Feature Point Description

After determining the feature points, each feature point has coordinates, magnitude and direction, and a descriptor needs to be established for the feature points. Take a feature point as the center position, take a 16×16 large area, and then divide the large area into 16 small areas with 4×4, and finally establish the gradient direction histogram of 16 small areas along 8 directions according to the method of feature point direction assignment. Arrange the gradient direction histogram, then a 128-dimensional feature vector can be generated, also known as SIFT feature descriptors. The schematic diagram of feature descriptors is shown in Fig. 4.



(a) The gradient magnitude and direction of feature points in the neighborhood

(b) The gradient direction of the sub-region

Fig. 4 Schematic diagram of feature point descriptors

3. Image Stitching Algorithm

3.1 Feature Point Matching

In order to match the feature points, the method of finding the feature point distance is usually adopted, such as the Mahalanobis distance and the Euclidean distance. Here, the SIFT operator selects the Euclidean distance as the basis for measuring the similarity of the feature points [10]. SIFT feature point extraction is performed on the reference image and the image to be stitched, and a 128-dimensional feature point descriptors is calculated.

The steps of coarse matching are: (1) One feature point of the reference image is P, calculate the Euclidean distance between the point P and each feature point of the image to be stitched. (2) Find the two points Q and O, which are corresponding to the minimum distance and the second smallest distance, the Euclidean distance is recorded as $M(P,Q)$ and $M(P,O)$. (3) Compare $M(P,Q)/M(P,O)$ with the threshold T for whether or not to match. If $M(P,Q)/M(P,O)<T$, it indicates that the points P and Q are successfully matched.

It can be seen from the coarse matching process that the setting of the decision threshold determines the number of feature points that match successfully. However, the feature point pairs that match successfully may be mismatched. If these mismatch feature point pairs are not removed, the final stitching result will be affected. In order to improve the stitching effect, RANSAC algorithm is used to remove the mismatch points, and the homography matrix is optimized [11] [12]. The steps are as follows: (1) Extract 6 feature points, the limitation is: any 3 points cannot be collinear, otherwise re-select. (2) Homography matrix is calculated using the least square method with the selected feature points. (3) Calculate the Euclidean distance between the image point and the projection point (corresponding point calculated by the homography matrix above). If the distance is less than a certain threshold, it would be selected as the inner point; otherwise it is the outer point. Record the number of interior points. (4) Select the set of points with the largest number of interior points, and recalculate the homography matrix using the least square method as the optimal homography matrix. This method can improve the effectiveness of feature point matching.

3.2 Image Fusion

If the images are directly stitched after the feature points are matched, the overlapping area will have cracks with obvious brightness changes. Therefore, the linear weighting function method is used to eliminate this phenomenon, to realize the seamless connection and smooth transition of the image.

The calculation formula of image fusion by linear weighting function method is:

$$g(x,y)=\begin{cases} f_1(x,y) & (x,y)\in f_1 \\ \alpha f_1(x,y)+\beta f_2(x,y) & (x,y)\in (f_1 \cap f_2) \\ f_2(x,y) & (x,y)\in f_2 \end{cases} \tag{8}$$

Where, $f_1(x,y)$ and $f_2(x,y)$ are the reference and stitched image. α and β are the weights of the overlapping area, for the specific value method, please refer to [13].

3.3 Algorithm Flowchart

The flow chart of image stitching algorithm based on SIFT features is shown in Fig. 5. The implementation process of the algorithm can be summarized as:

Input: image sequence

Output: stitched image

Step1: Read in the image and extract the SIFT feature points of each image;

Step2: Utilize the Euclidean distance for coarse match;

Step3: Use the RANSAC algorithm to filter the feature points and calculate the homography matrix;

Step4: Use homography matrix for image transformation;

Step5: Use the linear weighting function method to fuse the current image with the reference image;
Step6: Determine whether it is the last image, if so, output the stitching result; otherwise, go to Step2.

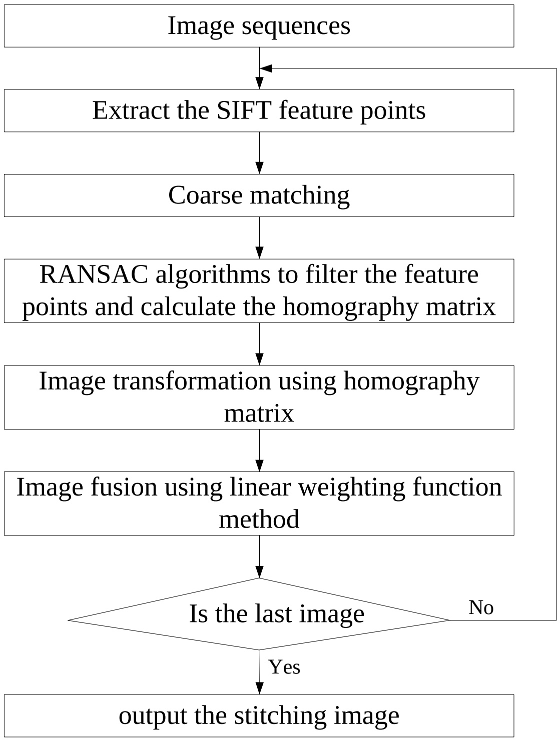
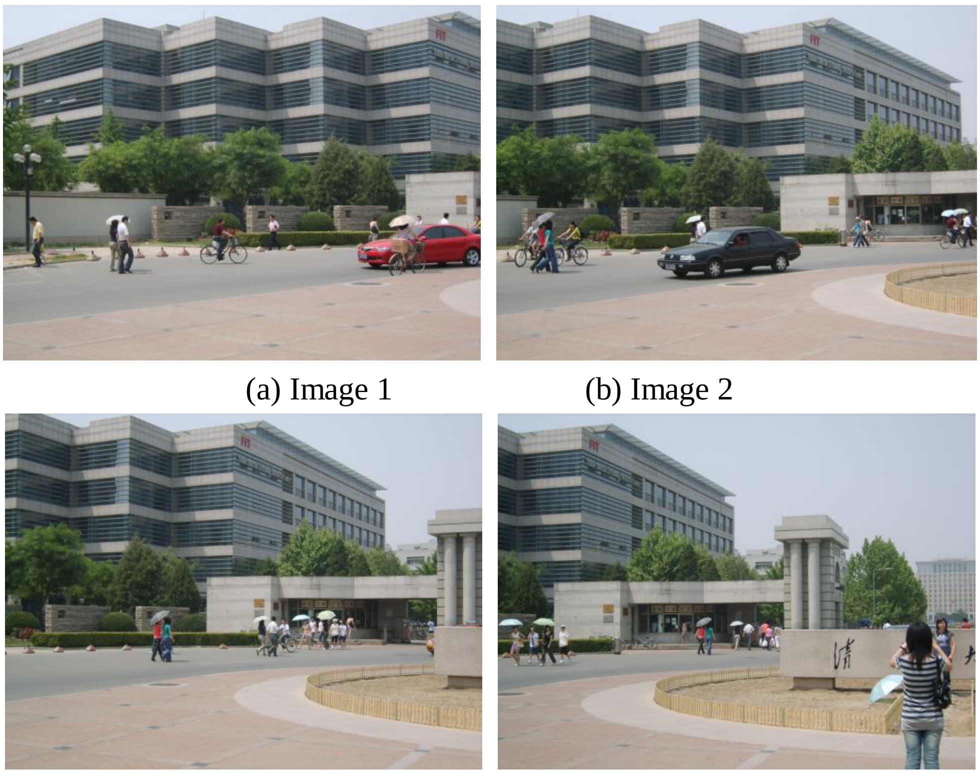


Fig. 5 flow chart of image stitching algorithm based on SIFT features

4. Experimental results

The four images shown in Fig. 6 are taken by the same camera, and there is a certain correlation and continuity between the images.



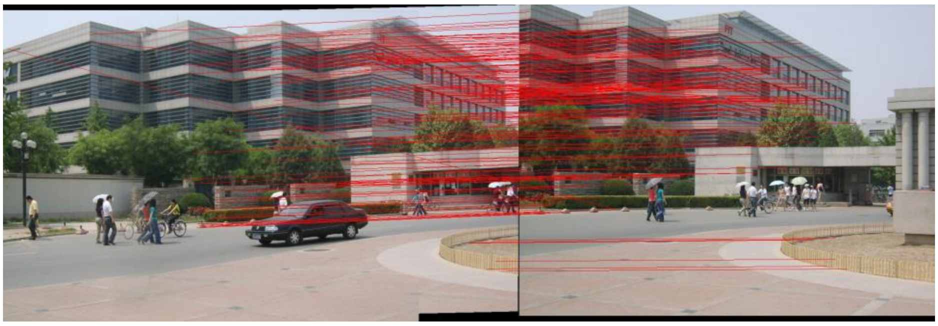
(a) Image 1 (b) Image 2

(c) Image 3 (d) Image 4

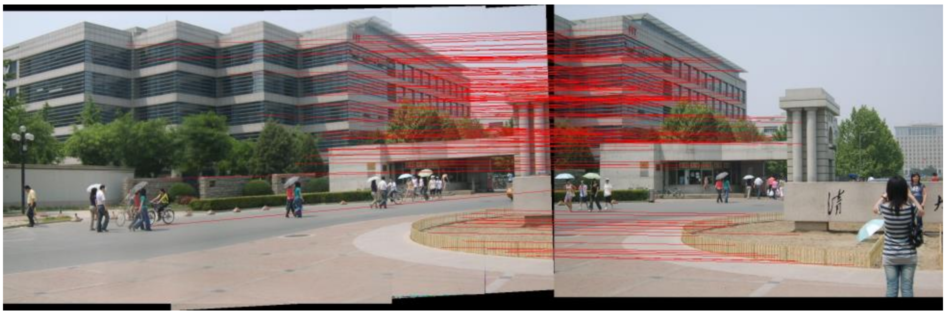
Fig. 6 Four images to be stitched

The result of feature matching is shown in Fig. 7, where Fig. 7 (a) is the matching result of image 1 and image 2, Fig. 7 (b) is the matching result of the stitched result and Image 3, Fig. 7 (c) is the matching result of the stitched result and Image 4.

(a) matching result of image 1 and image 2



(b) matching result of the stitched result and Image 3



(c) matching result of the stitched result and Image 4

Fig. 7 Matching results

The final stitching result is shown in Fig. 8.

Fig. 8 Image stitching result

5. Conclusion

This paper proposes a panoramic image stitching algorithm based on SIFT features. Improved RANSAC algorithm is used to calculate the homography matrix to ensure the robustness, so that the image matching reaches sub-pixel accuracy. In order to eliminate the seam after image stitching, an improved linear weighting function method is adopted to make the stitched image achieve a smooth transition. The experimental results show that the algorithm can perform panoramic stitching on the image sequence and obtain a better stitching effect.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant number 61901350); Higher Education Research Project of Xi'an Aeronautical University (grant number 2019GJ1006) and Science Research Fund of Xi'an Aeronautics University (grant number 2019KY0208).

References

[1] L. J. Cao, M. Lv: An Image Mosaic Method based on SIFT Feature Matching [J]. Advanced Materials Research, 2012, 433-440: 5420-5424.

[2] Y. T. Yang, Y. Feng, Y. Cao: Fast Method for Image Sequences Mosaic Based on SURF [J]. Computer Technology and Development, 2011, 21(3): 6-9.

[3] H. Lombaert, L. Grady, X. Pennec: Spectral Log-Demons: Diffeomorphic Image Registration with very Large Deformations [J]. International Journal of Computer Vision, 2014, 107(3): 254 -271.

[4] D. G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[5] B. S. Reddy, B. N. Chatterji. An FFT-based Technique for Translation, Rotation, and Scale-Invariant Image Registration [M]. IEEE Press, 1996, 5(8): 1266-1271.

[6] J. F. Dou, Q. Qin, Z. M. Tu: Robust Image Matching based on the Information of SIFT [J]. Optik, 2018, 171(10): 850-861.

[7] M. Wahed, G. S. El-Tawel, A. G. El-Karim: Automatic Image Registration Technique of Remote Sensing Images[J]. International Journal of Advanced Computer Science & Application, 2013, 4.

[8] W. Wen, M. Zhang: Research on Panoramic Image Mosaic Technology Based on SIFT Algorithm [J]. Computer Systems & Applications, 2017, 26(7): 227-231. DOI:10.15888/ j.cnki.csa.005899.

[9] S. Li, C. Cui: Research on Underwater Image Mosaic Technology based on SIFT Algorithm [J]. Internet of Things Technologies, 2018, 8(4): 51-53.

[10] P. Lu, Q. Lu, G. Zou: Research on Time Series Image Mosaic Method Based on Improved SIFT [J]. Computer Engineering and Applications, 2020, 56(1): 196-202. DOI:10.3778 /j.issn. 1002-8331.1809-0317.

[11] W. Huang: Research on Key Technology of Video Image Mosaic [D]. Changchun University of Science and Technology, 2014.

[12] M. Zhang, Y. Wang, X. Hou: Image Stitching Algorithm based on ORB and Improved RANSAC [J]. Computer Science, 2019, 46(z2): 294-298.

[13] D. Zhou, M. He, Q. Yang: A Robust Seamless Image Stitching Algorithm Based on Feature Points [J]. Measurement & Control Technology, 2009, 28(6): 32-36. DOI:10.3969/ j.issn. 1000-8829.2009.06.009.