Heterologous Image Matching Based on Salience Region

Jinlong Zhai, Yang Xu, Xiaohu Yan

Abstract—**To address the matching problem caused by the significant differences in spatial features, spectrum and contrast between heterologous images, a heterologous image matching method based on salience region is proposed in this paper. Firstly, texture features and corner features are detected. Secondly, the epanechnikov kernel function is used to calculate the point with the maximum kernel function value among all feature points, which is then defined as E-point. Thirdly, the salience region is defined and extracted according to the position of the E-point. Finally, six groups of heterologous images are used as data sources, and compared with HAPCG, HOWP, SIFT, and PSO_SIFT algorithms respectively. The results show that the comprehensive matching performance of the proposed method is better than the original algorithm without salience region processing. The root mean square error and the matching time are less than the original method while ensuring a similar number of matching points, which greatly improves the accuracy and speed of heterologous image matching.**

Index Terms—**Heterologous image, Salience region, Texture features, Corner features**

I. Introduction

WITH the development of remote sensing technology, heterologous image matching plays an increasingly important role in satellite navigation, remote sensing and telemetry, disaster warning, pattern recognition, medical image analysis, military and national defense, and other important fields [1]. The methods of obtaining images have also diversified. The images can be obtained through a range of methods such as multi-spectral cameras, synthetic aperture radar, thermal infrared sensors, and LiDAR systems. However, the images obtained from different sources often have large illumination differences, contrast differences, and nonlinear radiation differences, which leads to challenges and difficulties for heterologous image matching. Traditional image matching methods are divided into feature-based method [2] and region-based method [3]. Feature-based matching method first extracts various features from the images to be matched. Then, these features are compared to determine the similarity between the two images. However, due to the large differences in spatial features, field of view angle, and resolution of heterologous images, features are extracted and compared directly, which may lead to poor matching performance and low accuracy [4]. The region-based matching method is based on the entire pixel region, which uses some similarity measure (such as correlation function, and covariance function) to determine the corresponding relationship between two images. It requires calculating the similarity of the entire region, which results in a significant increase in both computational complexity and time consumption. The two methods do not distinguish the importance of information in the image. They perform excessive operations on non-important regions and fail to focus on the significant content of the images. This results in low accuracy and slow speed of heterologous image matching, which is difficult to meet the real-time application system of computer.

To address the matching problems caused by the many irrelevant feature points and wide search range in the traditional matching method of heterologous images. In this paper, we propose a method of matching heterologous images based on salience region. Due to the minimal impact of differences between heterogeneous images on structural and shape features, the method proposed in this paper combines texture features [5] and corner point features [6] to determine the salience regions of images [7]. By extracting features from salience regions, the sensitivity to local information and the matching accuracy of homologous points in heterologous images are enhanced. We focus on the areas that are more likely to have similar features, rather than the whole image. In this way, too many operations in unimportant areas are avoided, which is helpful to reduce the calculation time and improve the accuracy of matching. While ensuring a similar number of matched homologous points, this method achieves simultaneous optimization of the accuracy and speed of heterologous image matching.

II. Methodology

The calculation steps of salience region are as follows.

Firstly, the improved LBP method [8] is utilized to detect texture feature points, while the FAST method is used to detect corner feature points. The epanechnikov kernel function is used to calculate the point with the maximum kernel function value among all feature points, which is then defined as E-point. Finally, the salience region is detected according to the position of the E-point.

Manuscript received March 21, 2024; revised January 4, 2025.

This work is supported by the National Natural Science Foundation of China under Grant 62102268, the Stable Supporting Program for Universities of Shenzhen under Grant 20220812102547001, Shenzhen Polytechnic University Research Fund under Grants 6022312044K, 6023310030K and 6024310045K.

Jinlong Zhai is a postgraduate student of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan 114051, China (phone:86-15184189808, e-mail: 927341159@qq.com).

Yang Xu is a Professor at the School of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan 114051, China (corresponding author to provide phone: 86-13889785726; e-mail: xuyang_1981@aliyun.com).

Xiaohu Yan is a lecturer at the Institute of Applied Artificial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic University, Shenzhen 518055, China (phone: 86-15071175326; e-mail: yanxiaohu@szpt.edu.cn).

Texture features can resist noise interference to a certain extent, and they can extract effective information from heterogeneous images containing noise. The corner feature can resist the interference of illumination change and occlusion, so that the corner feature is more stable and reliable in different source images. It can provide the position, direction, size, and other information of the object in the image.

It has good recognition and high robustness to rotation and scale change of heterologous images. Therefore, this paper combines texture features and corner features to determine the salience region. Here, the homologous points to be matched in the image usually do not appear in the edge regions. The regions outside the salience region will not be calculated in the subsequent image matching.

The Local Binary Pattern (LBP) method is computationally efficient and has strong robustness to changes in pixel grayscale values and noise. It can effectively describe features in heterologous images and is insensitive to rotational changes in images. Therefore, according to the characteristics of heterologous images, this paper improves LBP and realizes texture feature detection.

The traditional LBP algorithm compares one pixel in each subregion with eight surrounding pixels. This process generates 2^p patterns, resulting in an excessively large feature space and longer computation time. To reduce the storage space requirements and speed up the computation, we compare it with four pixels above, below, left, and right. Due to the significant color differences between heterologous images, we convert the images to grayscale before performing any operations to reduce the interference and errors caused by color information during image processing. This helps improve the accuracy and speed of image processing while reducing its complexity.

A. Texture feature detection

The calculation steps of texture detection are as follows:

Step 1: Heterologous images are converted into grayscale images. The conversion calculation formula is described in (1).

$$\text{Gray} = R*0.299 + G*0.587 + B*0.114 \tag{1}$$

Where *Gray* is the gray value, and *R*, *G*, and *B* are the color component values of the red, green, and blue channels of the pixel respectively.

Step 2: For a pixel in each subregion, we compare it with the surrounding 4 pixels. If a pixel value is greater than the central pixel value, the value of the point is regarded as 1, and otherwise it is 0. The quantization formula for this step as shown in (2) and (3):

$$LBP(x_c, y_c) = \sum_{p=1}^{4} 2^p s(i(p) - i(c)) \tag{2}$$

$$s(x) = \begin{cases} 1 & if \quad i(p) - i(c) >= 0 \\ 0 & else \end{cases} \tag{3}$$

Where (x_c, y_c) is the central pixel, $i(c)$ is its gray value, p represents the $p-th$ pixel point above, below, to the left, and to the right of the central pixel point, $i(p)$ represents the grayscale value of the $p-th$ pixel within the neighborhood.

Step 3: The LBP value is converted to decimal and define it as the texture value of the point. Then we calculate the histogram of each sub-region and normalize it.

Step 4: The statistical histograms of each sub-region are concatenated into a feature vector. It represents the texture feature vector of the entire image.

Compared with other feature detection algorithms (such as SIFT [10], Harris [11], and DOG), the Features from Accelerated Segment Test (FAST) [9] have the advantage of high computational efficiency. It can process many pixels in a short time. This high efficiency makes FAST very suitable for real-time image processing applications, and it can effectively detect the corners in the image. Therefore, the FAST method is used in this paper to detect corner features of heterologous images.

B. Corner feature detection

The calculation steps of corner detection are as follows:

Step 1: We use point *p* as the center and 3 as the radius, we draw a circle. Then we subtract the pixel values of the 16 points on the circumference of the circle from the pixel value $I(p)$ of the center point. refer to (4).

$$Num = \sum_{x \forall (circle(p))} | I(x) - I(p) | > T \tag{4}$$

Where $I(x)$ represents the pixel value on the edge, $I(p)$ represents the pixel value of the center point, *circle(p)* represents the 16 pixels on the edge of the circle, *T* represents the threshold value defined as 10, and *Num*=12. Through experiments, when *Num*=12, the corner detection performance is the most stable, the fastest, and the effect is better. If 12 consecutive points satisfy $(I(x) - I(p)) > T$ or $(I(x) - I(p)) < -T$, then this point is a candidate corner point.

Step 2: We sum the absolute differences between the pixel values of the 16 pixels on the circle and a reference pixel value. The resulting values are used as the response values for non-maximum suppression. The feature points with lower response values among neighboring feature points are deleted.

Step 3: We define the retained points as corner feature points.

C. Detection of salience region

The calculation steps are as follows:

Step 1: Kernel density estimation objects are initialized. We use epanechnikov as the kernel density estimation function and determine the bandwidth value to be 0.02. x_1, x_2, …, x_n represent the n points in the texture features and corner features. The estimation formula for the kernel density is as follow:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(u) \tag{5}$$

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & if\,(|u| <= 1) \\ 0 & otherwise \end{cases} \tag{6}$$

$$u = \frac{x - x_i}{h} \tag{7}$$

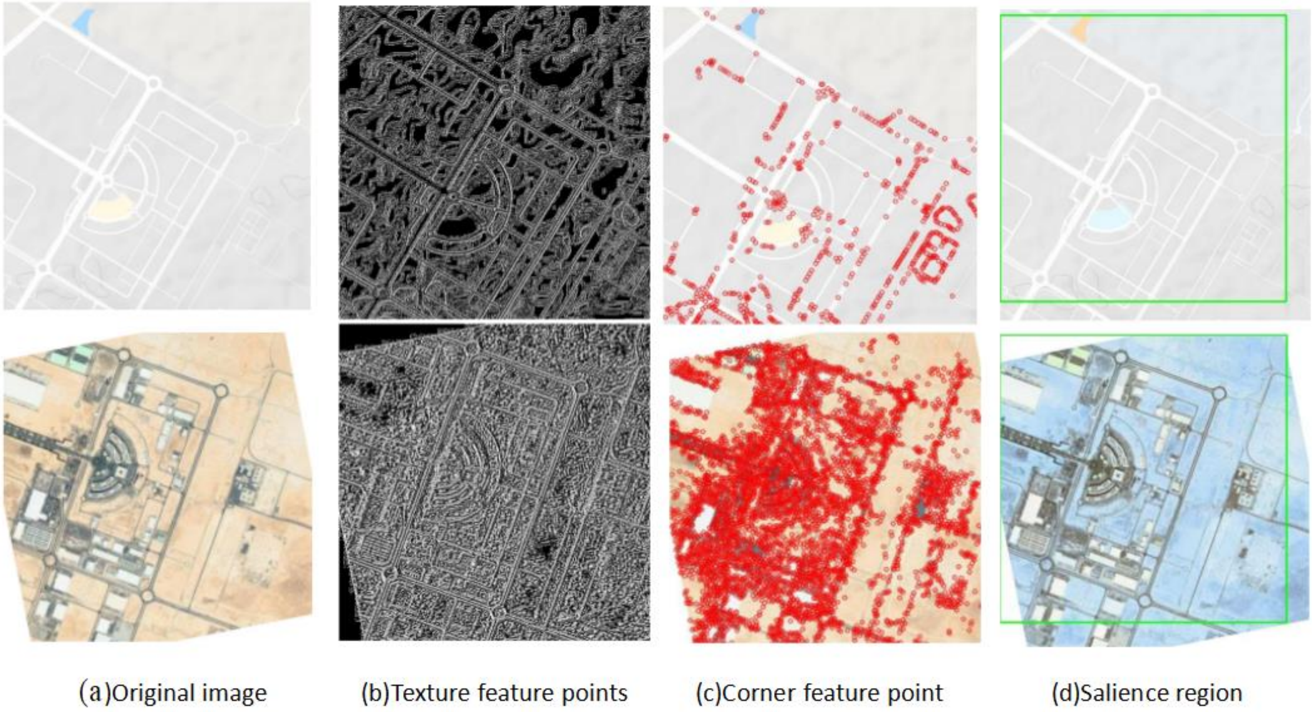

Fig. 1. The extraction effect of salience region

Where $K(u)$ is the epanechnikov function expression, $f(x)$ is the estimated probability density function at x, x_i is other points besides x in the feature points, n is the number of the feature points, and h is the bandwidth.

Step 2: We evaluate the log probability density of each point in Q-points, and calculate the M-point.

Step 3: We define the size of the salience region and detect the salience region according to the coordinates (a,b) of M-point. The size of the salience region is defined based on the length of image l and the width of image w.

Step 4: We define the parameters. x and y are the length and width of the image, respectively. x is r multiplied by l, y is r multiplied by w, t is the edge region discard ratio. In this paper, the ratio r is set to 0.9, and the ratio t is set to 0.05.

Step 5: If $a-1/2*l<0$ and $b-1/2*w<0$, then the top-left coordinate of the salience region is $(a-1/2*x,\ b-1/2*y)$ and the bottom-right coordinate is $(a+1/2*x,\ b+1/2*y)$.

Step 6: If the position of point M is in the top-left corner of the image, then the top-left coordinate of the salience region is $(t*l,\ t*w)$ and the bottom-right coordinate is $(x,\ y)$.

Step 7: If the position of point M is in the top-left corner of the image, then the top-right coordinate of the salience region is $(l-x,\ t*w)$ and the bottom-right coordinate is $((1-t)\ *l,\ y)$.

Step 8: If the position of point M is in the top-left corner of the image, then the bottom-left coordinate of the salience region is $(t*l,\ w-y)$ and the bottom-right coordinate is $(y,\ (1-t)\ *w)$.

Step 9: Otherwise, the bottom-right coordinate of the salience region is $(w-x,\ w-y)$ and the bottom-right coordinate is $((1-t)\ *l,\ (1-t)\ *w)$. The salience regions detected are shown in Fig. 1.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the effectiveness of the method proposed in this paper, we conduct a comparative analysis between the images obtained after extracting salience regions and the original images, using four commonly used methods: HAPCG, HOWP [12], SIFT, and PSO-SIFT. We use Root Mean Square Error (RMSE) and Matching Time (MT) as metrics to evaluate and compare performance. RMSE is a common index used to measure the performance of matching algorithms. RMSE reflects the matching accuracy of a correct match points. The smaller the RMSE value, the higher the precision. The mathematical expression of RMSE is described in (6).

$$RMSE=\sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}\left[\left(x_i'-x_i''\right)^2+\left(y_i'-y_i''\right)^2\right]\right)} \tag{8}$$

Where N is the number of points with the same name, and $\left(x_i'',y_i''\right)$ is the coordinate of the i th truth point $\left(x_i',y_i'\right)$ corresponding to the matching transformation.

The parameters of the four matching methods are adjusted to the optimal state. The above matching processes are implemented on MATLAB R2022a. The experimental platform processor is Intel(R) Core (TM) i7-9700 CPU 3.00(GHz), RAM is 16(GB), and Windows X64 operating system.

A. Image Data

In this paper, six groups of heterologous images with illumination differences, contrast differences, displacement differences, angle differences, and scale differences are used as data sources (see Fig. 2). Among them, the size of the first group of images is 500×500 pixels and 540×539 pixels. The size of the second group of images is 600×600 pixels. The

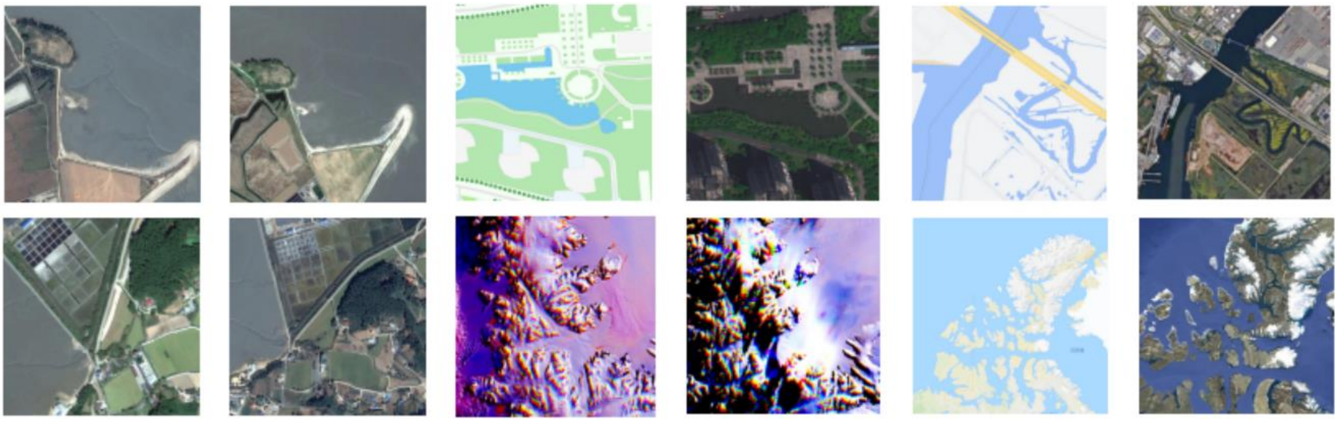

Fig. 2. Heterologous image data

size of the third group of images is 500×500 pixels. The size of the fourth group of images is 500×500 pixels and 540×540 pixels. The size of the fifth group of images is 500×500 pixels. The size of the sixth group of images is 550×550 pixels.

B. *Experimental Results*

Fig. 3 shows the effect of six sets of heterologous images using HAPCG, HOWP, SIFT, and PSO_SIFT methods on the original images (rows 1, 3, 5, and 7), as well as the corresponding results of salience regions of these same images (rows 2, 4, 6, and 8). As can be seen from Fig. 4, although the image size is reduced after the extraction of salience region, the number of matching points is similar to the original method, and there is no significant reduction. This is because the extracted salience region combines texture features and corner features. The feature information in the salience region can better express the important information of the image and is more in line with human visual perception. At the same time, because the matching points with the same name are in the non-significant region outside the salience region, they will not play an important role in the overall image matching.

Table I shows the RMSE and MT values of four methods on images with and without extraction of salience region. We quantitatively compare and analyze the performance of the methods proposed in this paper. As can be seen from Table I, when matching points with the same name is similar, the RMSE and MT by the four methods of HAPCG, HOWP, SIFT, and PSO_SIFT in the six groups of images are all smaller than the RMSE and MT without salience region extraction, which greatly improves the accuracy and speed of heterologous image matching.

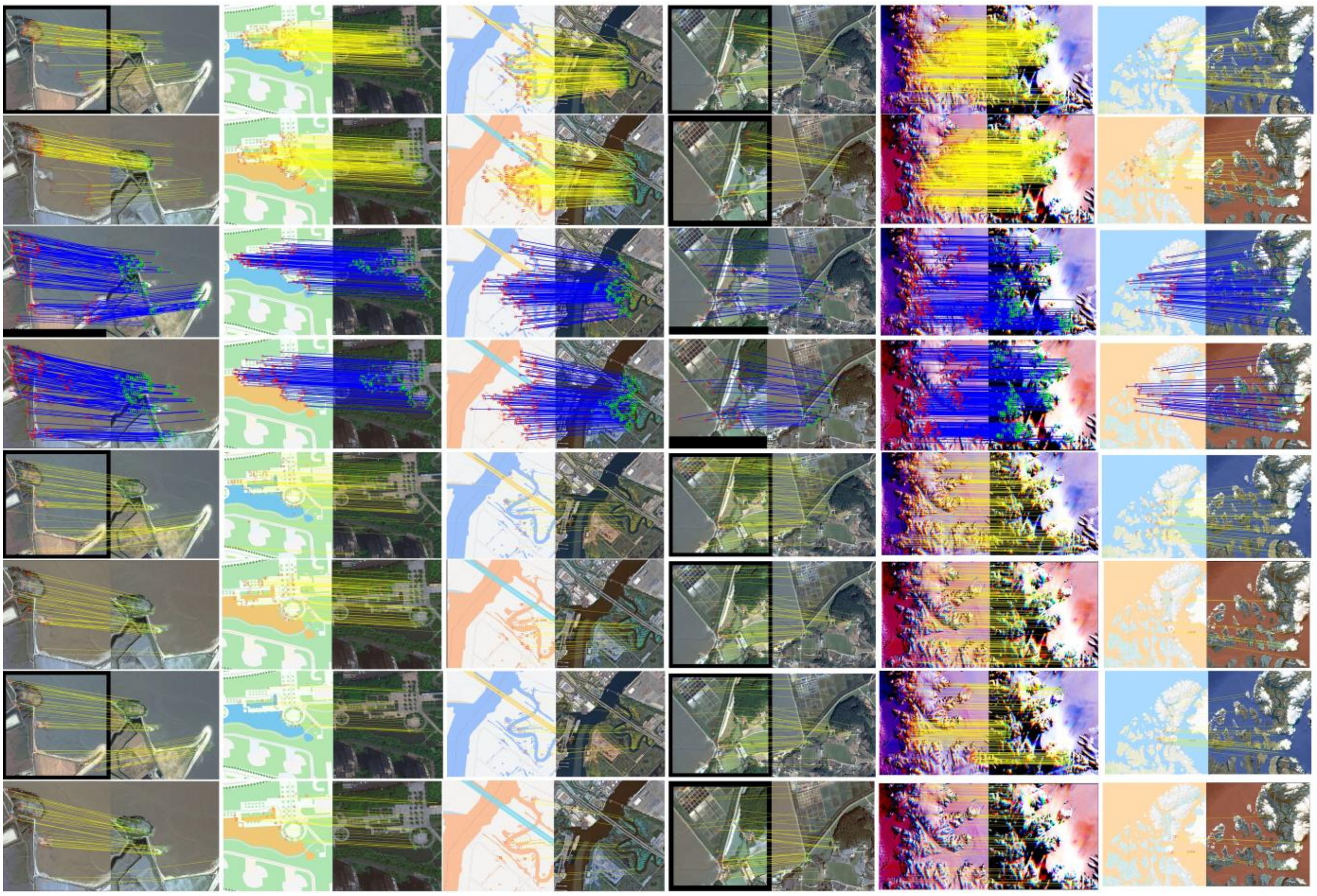

Fig. 3. A comparison chart of the matching results

TABLE I
MATCHING DATA OF THE FOUR METHODS IN THE PROPOSED METHOD AND THE ORIGINAL METHOD

image	index	HAPCG	HAPCG based on salience region	HOWP	HOWP based on salience region	SIFT	SIFT based on salience region	PSO_SIFT	PSO_SIFT based on salience region
1	RMSE(pix) MT(s)	2.0656 11.2358	**1.8964** **7.955**	2.1126 6.055	**1.976** **3.904**	2.6108 4.125	**2.5033** **1.952**	1.8655 7.63	**1.8111** **7.185**
2	RMSE(pix) MT(s)	2.0588 12.672	**2.0299** **10.804**	1.9687 6.35	**1.9332** **4.672**	2.1724 4.066	**2.0512** **3.558**	1.9228 9.248	**1.8925** **7.4**
3	RMSE(pix) MT(s)	2.0621 9.174	**1.9544** **8.51**	1.8611 4.542	**1.857** **3.384**	2.19 2.2411	**1.6079** **2.238**	1.9806 6.102	**1.975** **5.754**
4	RMSE(pix) MT(s)	2.2108 9.323	**1.8486** **8.826**	1.9172 4.722	**1.6382** **3.748**	2.2726 4.266	**2.2066** **2.774**	1.9736 9.82	**1.9607** **6.871**
5	RMSE(pix) MT(s)	1.9329 8.514	**1.9102** **8.041**	1.8868 4.052	**1.7988** **3.801**	1.8589 4.548	**1.8149** **4.583**	1.6957 16.862	**1.6091** **9.16**
6	RMSE(pix) MT(s)	1.9493 11.016	**1.81** **9.495**	1.9799 5.125	**1.7857** **4.248**	1.9594 3.135	**0.59417** **2.129**	1.9652 7.244	**1.7704** **4.941**

As shown in the chart above, although a large number of feature points are detected in the whole image before improvement, irrelevant feature points occupy a large amount of computing time, which will seriously affect the image matching quality. After the improvement, the feature points are gathered on the target object, which can minimize the matching interference of the background region to the target region.

The findings indicate that prior to the improvements, the matching pairs generate from feature points in the image background significantly affected the quality of matches. Moreover, there are numerous erroneous matches, such as those between feature points in the background and those in the target area. This situation has a considerable impact on the overall matching results.

After implementing improvements, there is a notable reduction in the number of detected feature points compared to before. Nevertheless, it is important to highlight that almost all of the feature points detected after improvement were located within the target area. Consequently, these post-improvement detected feature points exhibited higher quality than their pre-improvement counterparts and more accurately reflected the true information regarding the target.

To further confirm the salience region detection effect of the proposed method on different types of heterologous images, this study selected five commonly used heterogeneous data types as data sources: electronic navigation map (Map), point cloud depth map (DSM), infrared image (Infrared), night light image (Night), and synthetic aperture radar (SAR). Among them, Fig. 4, 5, 6, and 7 respectively show the experimental comparison results of HAPCG, HOWP, SIFT, and PSO-SIFT algorithms on five types. The results show that the proposed method has better comprehensive matching performance than the original method on most types, indicating that this method has a high degree of universality and is suitable for complex heterologous image matching.

The first row of each group of images did not undergo salience region detection, while the second row of each group of images underwent salience region detection. The RMSE and Total Time on the original image are marked as RS-1 and MT-1, respectively. The RMSE and MT on significant regions are represented as RS-2 and MT-2.

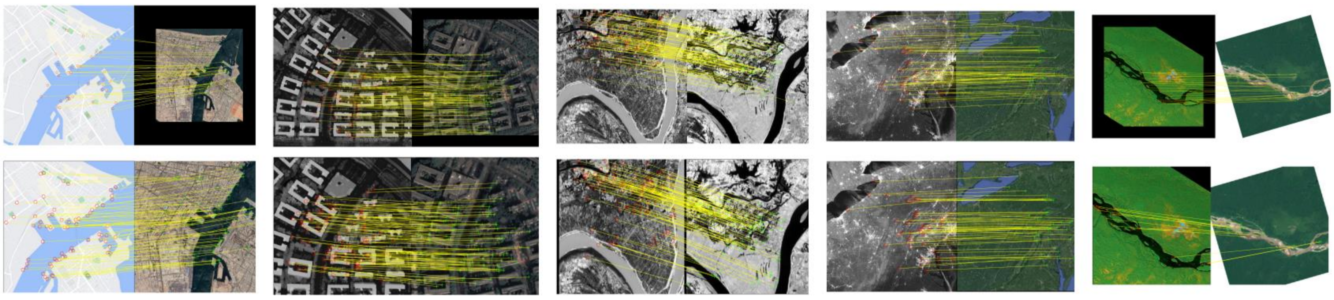

Fig. 4. Comparison effect diagram of PSO-SIFT algorithm

TABLE II
COMPARISON RESULTS OF PSO-SIFT ALGORITHM

Image pair	Map		DSM		Infrared		Night		SAR	
	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2
PSO-SIFT	1.854/ 7.175	**1.677/ 5.151**	1.8776/ 10.419	**1.966/ 7.594**	2.0962/ 4.6816	**1.9722/ 10.656**	1.826/ 8.086	**1.7366/ 6.633**	2.1755/ 6.638	**1. 7314/ 4.357**

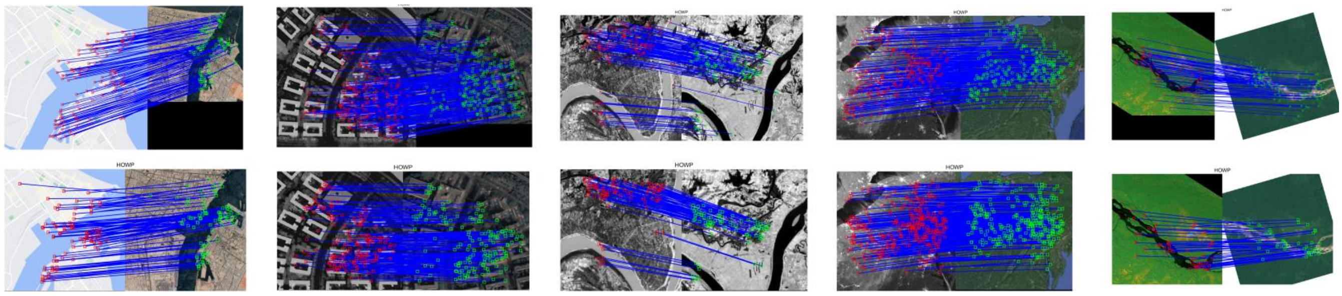

Fig. 5. Comparison effect diagram of HOWP algorithm

TABLE III
COMPARISON RESULTS OF HOWP ALGORITHM

Image pair	Map		DSM		Infrared		Night		SAR	
	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2
HOWP	1.8853/ 3.628	**1.8473/ 2.799**	2.0089/ 5.524	**1.9021/ 3.508**	1.9026/ 4.18	**1.7897/ 2.88**	1.9612/ 5.698	**1.8987/ 3.93**	1.9892/ 6.147	**1. 9321/ 3.375**

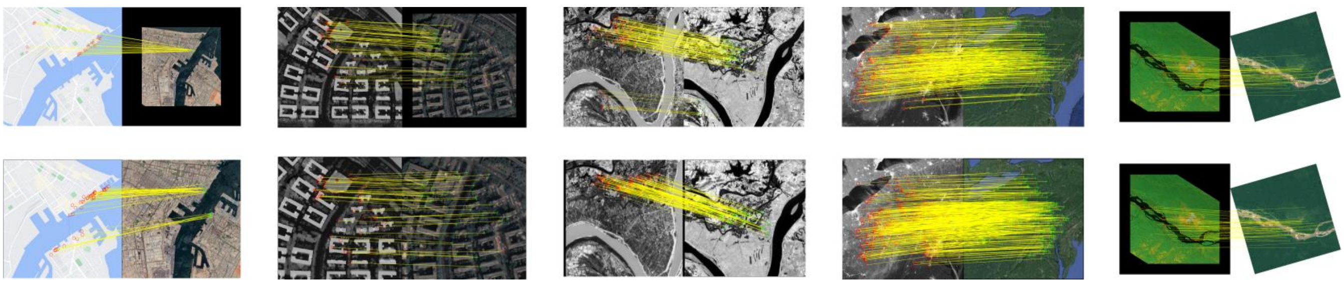

Fig. 6. Comparison effect diagram of HAPCG algorithm

TABLE IV
COMPARISON RESULTS OF HAPCG ALGORITHM

Image pair	Map		DSM		Infrared		Night		SAR	
	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2
HAPCG	1.8349/ 10.239	**1.6824/ 7.13**	1.8776/ 10.419	**1.966/ 7.077**	1.895/ 9.369	**1.6714/ 7.007**	1.9867/ 8.788	**2.0214/ 7.233**	1.9928/ 12.653	**2.0328/ 7.228**

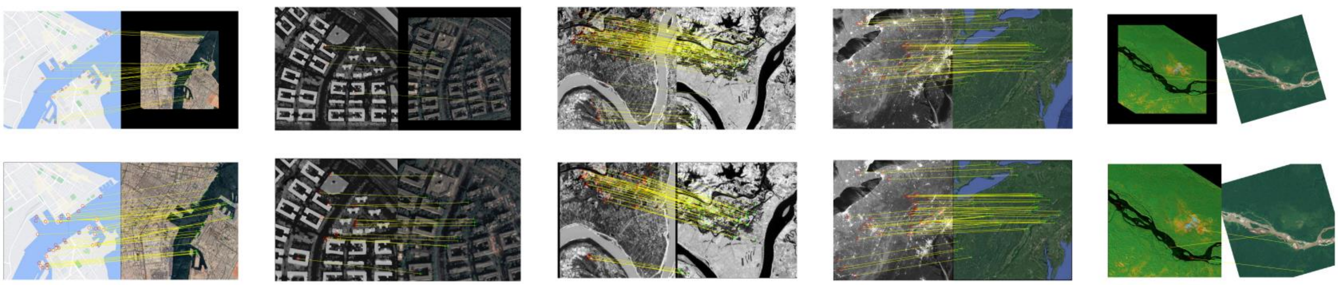

Fig. 7. Comparison effect diagram of SIFT algorithm

TABLE V
COMPARISON RESULTS OF SIFT ALGORITHM

Image pair	Map		DSM		Infrared		Night		SAR	
	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2	RS-1/ MT-1	RS-2/ MT-2
SIFT	2.2653/ 2.214	**1.8573/ 1.889**	1.2937/ 7.587	**1.2243/ 3.398**	1.8263/ 7.247	**1.5679/ 3.122**	1.8178/ 2.941	**1.8481/ 1.924**	1.425/ 3.546	**1.346/ 1.694**

As shown in Fig. 4, 5, 6, and 7, although the image size is reduced after the extraction of salience region, the number of matching points is similar to that of the original method, and there is no significant reduction. This is because the feature information in the extracted salient regions can better express the important information of the image. As can be seen from Table II, III, IV and V, the MT of the two methods in the five types of images is smaller than the MT of those without salience region extraction. Based on all the above situations, we can draw such a conclusive conclusion: By carrying out the operation of salient region extraction, the speed of heterologous image matching can be effectively improved. It should be particularly noted that for those algorithms with relatively long matching times, the improvement effect of the proposed method in terms of time is even more prominent and significant.

IV. Conclusion

To address the matching problem caused by the large differences in spatial characteristics, spectrum, and contrast between heterologous images. In this paper, a heterologous image matching method based on salience region is proposed. The salience region proposed in this paper combines texture features and corner features to better express the important information of an image. It is more distinctive and aligns better with human visual perception. By using four methods HAPCG, HOWP, SIFT, and PSO_SIFT, we conduct detailed experimental analysis on six groups of images with and without salience region extraction. The experimental results show that the proposed method has lower root mean square error and matching time than the original method while the number of matching feature points is similar. The salience regions proposed in this paper can express the core content of the image more effectively. The proposed method greatly improves the speed and accuracy of heterologous image matching and has strong robustness and universality in heterologous image matching.

References

[1] H. Sui, J. Li, J. Lei, C. Liu, and G. Gou, "A Fast and Robust Heterologous Image Matching Method for Visual Geo-Localization of Low-Altitude UAVs," *Remote Sensing*, vol.14, no.22, pp.5879,2022.

[2] B. Karasfi, S.H. Tang, K. Samsudin, A.R.B. Ramli, A. Jalalian, and O. Motlagh, "Unsupervised place recognition for assistive mobile robots based on local feature descriptions," *Proceedings of the Institution of Mechanical Engineers*, vol.225, no.8, pp.1068-1085,2011.

[3] Z. Tang, Z. Ding, R. Zeng, Y. Wang, J. Wen, L. Bian, and C. Yang, "Multi-Threshold Corner Detection and Region Matching Algorithm Based on Texture Classification," *IEEE Access*, vol.7, pp. 128372-128383,2019.

[4] Z. Wang, A. Yu, Z. Dong, B. Zhang, and X. Chen, "Performance Evaluation of Interest Point Detectors for Heterologous Image Matching," *Remote Sensing*, vol.14, no.15, pp.3724-3724,2022.

[5] K. Özkan, A. Mert, and S. Özdemir, " A new proposed GLCM texture feature: modified Rényi Deng entropy," *The Journal of Supercomputing*, vol.79, no.18, pp.21507-21527,2023.

[6] M. Bansal, M. Kumar, and M. Kumar, "2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors," *Multimedia Tools and Applications*, vol.80, no.12, pp.1-19,2021.

[7] C. Li, K. He, D. Xu, D. Tao, X. Lin, and H. Shi, "Superpixel-based adaptive salient region analysis for infrared and visible image fusion,"*Neural Computing and Applications*, vol.35, no.30, pp. 22511-22529, 2023.

[8] R. S. Zhou, P. J. Yin, "LBP Texture Feature Based on Haar Characteristics," *Journal of Software*, vol.24, no.8, pp.1909-1926,2014.

[9] Q. Feng, S. Tao, C. Liu, H. Qu, and W. Xu, "IFRAD: A Fast Feature Descriptor for Remote Sensing Images," *Remote Sensing*, vol.13, no.8, pp.3774-3774,2021.

[10] D. Barath, "On Making SIFT Features Affine Covariant," *International Journal of Computer Vision*, vol.131, no.9, pp.2316-2332,2023.

[11] L. Rasmy, I. Sebari, and M. EMTarid, "Automatic Sub-Pixel Co-Registration of Remote Sensing Images Using Phase Correlation and Harris Detector," *Remote Sensing*, vol,13.no,12. pp.2314-2314,2021.

[12] Y. Zhang, Y. Yao, Y. Wan, W. Liu, W. Yang, and Z. Zheng, "Histogram of the orientation of the weighted phase descriptor for multi-modal remote sensing image matching," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.196, pp. 1-15, 2023.