Multi-Source Image Registration Based on Slic Superpixel Segmentation and Autoencoder

Xie Zhao*, Tianda Yong
School of Electronics and Information, Northwestern Polytechnical University, XiAn, China

*Corresponding author: xz2018261759@mail.nwpu.edu.cn

Abstract. A new Harris-AE feature matching algorithm based on image edge information combined with SLIC superpixel segmentation method is proposed to achieve high-precision registration of visible and infrared images. First, perform edge enhancement and histogram equalization processing on visible light and infrared images, superpixel segmentation technology is used to eliminate sub-images with low information entropy, and then by constructing a multi-scale Gaussian pyramid to detect the Harris corners of the image edge information, using autoencoder neural network to generate feature descriptors corresponding to feature points, and match the feature points of two images through the fast nearest-neighbor algorithm with bidirectional matching strategy. Finally, the RANSAC algorithm is used to purify the matching points, and the spatial geometric change parameters between the two images are estimated by the least-squares solution to complete the registration of visible light and infrared images. Experiments have proved that this algorithm improves the accuracy of matching while shortening the registration time.

Keywords: Superpixel segmentation, edge information, image registration, multi-scale Harris-AE feature detection.

1. Introduction
Infrared and visible light sensors, as the most commonly used image sources, have been widely used in industries [1], military [2], agriculture [3], remote sensing [4] and other fields. However, due to the significant difference in imaging principles between visible light and infrared sensors, the two images often have large differences in grayscale distribution and texture. Ordinary single-source image registration algorithms are not well applicable to Multi-source registration scenario; therefore, it is of great significance to study multi-source image registration algorithms with high accuracy and fast speed.

In the field of single-source image registration, the scale invariant feature transform (SIFT) algorithm has a wide range of applications. Zhang Haitao et al. [5] used the improved SIFT algorithm based on Marr wavelet to complete the remote sensing image registration from the same sensor. However, for the automatic registration between remote sensing images of different sensors with rich texture, especially with significant gray-scale differences, the traditional SIFT algorithm has the problem of low matching accuracy [6]. Literature [7] proposed an image block strategy. For the filtered sub-images, the PCA-SIFT algorithm is used to generate low-dimensional feature descriptors
for image registration, which speeds up the registration while ensuring similar registration accuracy. Literature [8] proposed multi-feature fusion and used the coarse-fine two-step method for multi-source remote sensing image matching, which improved the accuracy of registration. Literature [9] based on FARISFD and ROGFD, proposed an adjustment method for the registration of multiple heterogeneous images to optimize the matching results. Although the above methods implement heterogeneous images, they still have defects such as high computational complexity and redundant feature points. Literature [10] analyzes and compares a variety of feature point extraction algorithms, and proves that Harris feature points have a relatively wide range of applications in multi-source remote sensing image registration. Literature [11] adopts a rough-fine two-step matching strategy, and constructs a triangulation network with Harris feature points in the fine matching stage to achieve high-precision registration of multi-source remote sensing images. However, the algorithm needs to calculate feature points and descriptors twice, which is difficult to generalize because of its high computational complexity. Literature [12] proposed a method of using Harris corner points instead of SIFT extreme points, and using the original SIFT method to generate descriptors for Harris feature points. Experiments show that this method accelerates the matching speed while ensuring the matching accuracy. Literature [13] proposed the use of multi-scale Harris-SIFT feature detection and description method, combined with a two-way matching registration strategy, to achieve the registration of visible light images and infrared images, with a better registration effect.

This paper combines the gray distribution and texture characteristics of visible light images and infrared images, use the edge information response of the image to overcome the gray and texture difference of the two types of images, and combine the edge information to detect Harris feature points, and use the autoencoder network model in deep learning to generate corresponding feature descriptors, at the same time, super pixel segmentation technology is used to speed up the calculation rate of the algorithm. An automatic registration algorithm suitable for visible light and infrared images is proposed in this paper, which improves the efficiency and accuracy of registration, and obtains effective registration results.

2. Algorithm Flow
The paper algorithm flow is as follows

(a) In the precessing process, the improved Laplacian operator is used to enhance the edge of the image, and at the same time, increasing the gray-scale similarity of the image through platform histogram equalization. Finally, SLIC superpixel segmentation technology is used to eliminate sub-images with low information entropy.

(b) Construct a Gaussian pyramid, and use Sobel operator to detect edge information of visible light and infrared images.

(c) Use Harris operator to detect feature points in combination with image edge information, and use autoencoder model (AE) to generate corresponding feature descriptors.

(d) Adopt a two-way strategy, use the fast nearest-neighbor algorithm for rough matching of feature points, and use the RANSAC algorithm for second fine matching.

(e) Calculate the affine transformation model parameters by using the registration points, and complete the image registration fusion of the infrared image to be registered through the bilinear difference.
In the multi-source image registration, because the imaging principles of the source image and the image to be registered are different, the single-source image registration algorithm often has a high mismatch rate or even unsuitability in the multi-source field. How to improve the similarity between images by preprocessing images is an important factor affecting the accuracy of subsequent image registration. Combining the imaging principle analysis of visible light image and infrared image, it can be seen that there are two main differences in its characteristics: (1) The grayscale distribution of the two images is different. Because the infrared image is mainly obtained by sending the heat radiated from the object. Therefore, compared with the visible light image, the grayscale distribution of the infrared image has no direct linear relationship with the target reflection characteristics, which is obviously different from the grayscale distribution of the visible light image; (2) When depicting the same area or target, the texture details are not consistent.

This paper uses Laplacian to enhance the edge of the image, and at the same time, uses the platform histogram equalization method to process the image to make the grayscale distribution between the images as close as possible.

### 3.1. Edge Enhancement

In order to effectively extract the edge information of the image and ensure the accuracy of the high-quality edge information, this paper uses an improved 5x5 matrix Laplacian to enhance the edge features of the image [16]. As shown in Figure 2, the Laplacian using a 5x5 matrix has a stronger effect of enhancing the edge of the image, and the image is clear and the edge contour is sharp.

![Fig. 2 Traditional Laplacian operator edge enhancement effect diagram and improved Laplacian operator enhancement effect diagram](image)

### 3.2. Superpixel Segmentation Based on Adaptive Threshold

In order to improve the computational efficiency of the algorithm while preserving the texture characteristics of the source image to the greatest extent, this paper introduces SLIC superpixel segmentation technology in the preprocessing process, the information entropy value of the segmented image blocks can be computed as:
\[ H = - \sum_{i=0}^{n} p_i \log p_i \]  

(1)

By equation (1), smooth low-texture image blocks, such as lakes, deserts, and sky, are eliminated, which shortens the time of image registration and improves the efficiency of registration.

In order to increase the universality of the image segmentation algorithm, we did not use a fixed threshold method to set the threshold entropy to filter the image block, but calculates it through an adaptive entropy threshold. The steps are as follows:

(a) Assuming that the image is divided into \( n \) sub-images, calculate the entropy value \( S_i \) of the \( i \)-th sub-image in turn.

(b) Calculate the average \( S_{av} \) of the \( n \) sets of data obtained in the previous step, Initialization threshold \( S_r = S_{av} \).

(c) Compare the value of \( S_i \) and \( S_{av} \). If three consecutive sub-images are larger than \( S_{av} \), There may be problems with uneven distribution of image entropy and obvious polarization. At this time, \( S_r = S_r + \Delta S \), \( \Delta S = \frac{S_{av}}{n} \). Otherwise, \( S_r = S_r - \Delta S \).

The segmentation effect is shown in Figure 3:

![Fig. 3 Visible light image after segmentation](image)

4. Hariss Feature Point Detection

4.1. Edge detection effectiveness analysis

In the field of single-source image registration, the registration algorithm with feature points as the key information is the mainstream algorithm. However, simple feature point detection and matching methods mostly rely on the grayscale distribution and local texture features of the image. Direct application of this kind of algorithm in the registration of visible light and infrared images, due to the difference in the imaging principle of the two source images, the gray distribution and texture characteristics are different, which will lead to a significant increase in the mismatch rate in the subsequent feature point matching process, and even match failed.

Literature [14] shows that image edge features are not sensitive to grayscale, contain important information that is easy to identify, and can effectively express image features. As the stable shared information between images, the edge structure is well presented in different bands [15].

4.2. Multi-scale Harris Corner Detection

Harris corner detection algorithm is a classic signal-based feature extraction algorithm proposed by Chris Harris and Mike Stephens in 1988 [16]. Harris corner points have low computational complexity, high stability, and are not easily affected by interference factors such as light and noise. However, Harris corner points have the disadvantage of not having scale invariance.

Aiming at the disadvantage that Harris algorithm does not have scale invariance, we extend Harris corner detection to Gaussian pyramid space, and combines the edge information of image sets to
extract edges Harris corner points with scale information in the structure. Constructs image sets of different scales and sizes according to

\[
G(x_i, y_i, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{ \frac{-(x-x_i)^2 + (y-y_i)^2}{2\sigma^2} \right\}
\]  

(2)

\[L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y)\]  

(3)

\[I(x, y, \sigma)\] is defined as the original image, \(\otimes\) represents convolution operation, \(\sigma\) is the scale space factor, \((x, y)\) represents the pixel position of the image, \(G(x_i, y_i, \sigma)\) represents the two-dimensional convolution kernel with position \((x, y)\) and scale \(\sigma\), the smaller the value of \(\sigma\), the less smoothed the image. Large-scale corresponds to the general features of the image, and small-scale corresponds to the detailed features of the image.

4.3. Calculate the edge feature response in Gaussian scale space

Use the sobel operator to detect the edge features of the image in the scale space, \(A\) is the image to be detected, \(G_x\) and \(G_y\) are the corresponding gray gradient values in the horizontal and vertical directions, can be computed as

\[
G_x = \begin{bmatrix}
-1 & 0 & +1 \\
-2 & 0 & +2 \\
-1 & 0 & +1 \\
\end{bmatrix} \otimes A
\]

\[
G_y = \begin{bmatrix}
+1 & +2 & +1 \\
0 & 0 & 0 \\
-1 & -2 & -1 \\
\end{bmatrix} \otimes A
\]  

(4)

We calculate the size of the gray gradient of the corresponding point as

\[|G| = |G_x| + |G_y|\]  

(5)

This paper uses the OTUS algorithm to divide the gradient magnitude information \(G\) into two parts: the background and the foreground. The foreground part is the image edge information.

4.4. Harris feature point detection based on edge feature

We perform Harris feature point detection based on the detected edge information. The second moment of the multi-scale Harris corner operator is

\[
M(x, y, \sigma) = G(x, y, \sigma) \otimes \begin{bmatrix}
A(\sigma_i) & C(\sigma_i) \\
C(\sigma_i) & B(\sigma_i) \\
\end{bmatrix}
\]  

(6)

And

\[
A(\sigma_i) = (I_x)^2 \otimes G(x_i, y_i, \sigma)
\]

\[
B(\sigma_i) = (I_y)^2 \otimes G(x_i, y_i, \sigma)
\]

(7)

\[
C(\sigma_i) = (I_x \cdot I_y)^2 \otimes G(x_i, y_i, \sigma)
\]
we calculate the Harris corner response function $R(\sigma_i)$ on scale $\sigma_i$ as

$$R(\sigma_i) = \det(M(\sigma_i)) - k \cdot \mu^2(M(\sigma_i))$$  \hspace{1cm} (8)$$

$k$ is the weight coefficient, the value is 0.05, $\det$ is the determinant of the matrix, $\mu$ is the rank of the matrix, when the value of $R$ is greater than the set threshold and is the maximum in the 5x5 field, the current point is the corresponding Harris feature corners.

5. Feature Description Based on Autoencoder

In order to better describe the local features of feature points and improve the accuracy of multi-source image registration, we use autoencoders to automatically extract feature point descriptors to overcome the knowledge limitations of manually defining descriptors. The autoencoder is a neural network whose output value is equal to the input value using the backpropagation algorithm. It compresses the input into a latent space representation, and reconstructs the output through this representation. In this letter, we use the self-reconstruction characteristic of the autoencoder to realize the generation of the feature point descriptor. In order to reduce the computational complexity of the algorithm while ensuring the accuracy of the descriptor description, we define the encoding layer (Encoder) and the decoding layer (Decoder) as a two-layer network, and the hidden layer is a 96-dimensional descriptor vector. The network model is shown in Figure 4:

![Autoencoder network structure](image)

**Fig. 4** Autoencoder network structure

Define the loss function as:

$$J(W,h) = \left[ \frac{1}{2} \| h_{W,h}(x) - x \| \right] + \frac{\lambda}{2}w^2$$  \hspace{1cm} (9)$$

$h_{W,h}(x)$ is the output image of the self-encoding network, $x$ is the input image, and $\lambda$ is the regularization parameter, which is set to 0.03. The $L_2$ regularization method is used to avoid network overfitting.

The circle has good rotation invariance characteristics, after the image is rotated and transformed, the local features around the feature points will not change significantly. In order to ensure the rotation invariance of the feature descriptors, we use the feature point partial circular image to construct the feature descriptor.
As shown in Figure 5, taking the feature point as the center and 12 pixels as the radius, take the local circle image and convert it into a feature vector, and pass it into the autoencoding network to generate a feature descriptor; Compared with the SIFT algorithm that generates feature descriptors through statistical histograms, the method in the paper increases the local perception field of descriptors, and generates descriptors through the self-learning ability of the autoencode network which effectively overcomes the knowledge limitations of artificially designed descriptors. And compared with the original 128-dimensional descriptor, the 96-dimensional descriptor generated by the self-encoding network reduces the computational complexity of the algorithm. Finally, the generated descriptor is normalized to reduce the influence of light.

6. Feature Point Matching
Because the rich feature points and corresponding feature vectors in the multi-source images increase the mismatch rate, we use a combination of coarse and fine matching methods to screen matching point pairs.

Aiming at the shortcomings of the fast nearest-neighbor algorithm in the process of multi-source image registration, we adopt two-way nearest neighbor/second nearest neighbor matching strategy for rough matching of feature points. First, perform traversal search and match from the feature points of image A to the feature points of image B, and then reverse the traversal search and match from image B to image A. Only when the matching results are consistent in the two detection processes can it be retained as a pair of registration points, otherwise it will be deleted, and further filtered by the RANSAC algorithm, and the least square method is used to calculate the affine transformation parameters to complete the registration.

7. Experimental Results and Analysis
In order to prove the effectiveness of the algorithm in this paper, a comparative experiment of feature point matching is carried out on two sets of infrared images and visible light images. The simulation platform hardware environment is: Inter(R)Core(TM)i5CPU, 2.4GHz, 8G memory PC, the development tool is Windows10 64-bit operating system, and the simulation programming environment is MATLAB-R2016b The images used in the experiment come from the data set TNO_Image_Fusion_Dataset, as shown in Figure [6], where the first group (a) (b) is the visible light and near-infrared images corresponding to the field scene, the size is 575×475; the second group (c) (d) is the visible light and near-infrared image corresponding to the park scene, the size is 640×480.
Firstly, the advantages of this method in the detection time of feature points compared with the traditional method of SIFT are proved through experimental comparison. Using the superpixel segmentation and edge response based on the visible and infrared image registration algorithm and the standard SIFT algorithm proposed in the paper to compare the visible image and the infrared image in Figure [6]. The statistical results of the feature point detection time of visible light and infrared images are shown in Table I. It can be seen that using the method in this paper to detect feature points has the advantages of short time consumption and high calculation efficiency:

| Angle Ratio | Number of feature points | Time of feature points detection/s |
|-------------|--------------------------|-----------------------------------|
|             | SIFT                     | The method                        | SIFT                  | The method                        |
| Fig.6(a)    | 0 1                      | 2457 986                          | 1.488 1.233           |
| Fig.6 (b)   | 0 1                      | 1940 677                          | 0.822 0.684           |
| Fig.6 (c)   | 0 1                      | 932 524                           | 1.844 1.105           |
| Fig.6 (c)   | 0 1                      | 417 311                           | 1.209 0.763           |

Then the experiment proves the effectiveness of the method proposed in this paper for the registration of visible light and infrared images. In order to quantitatively analyze the registration accuracy, the RMSE between the registration points is calculated to measure the registration accuracy as

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (E_x^i)^2 + (E_y^i)^2}
\]

Where \( n \) is the number of matching points, \( E_x^i \) and \( E_y^i \) are the registration errors of the i-th matching point in the horizontal and vertical directions respectively, \( L_x^i, L_y^i \) represents the horizontal and vertical coordinates of the i-th matching point on the reference image, \( R_x^i, R_y^i \) represent the horizontal and vertical coordinates of the i-th matching point on the reference image, \( \varphi(\bullet) \) and \( \omega(\bullet) \) are the affine transformation model.

In each group of experiments, the images to be registered are rotated by 10° and zoomed by 0.8 times, and the feature points of each group of test images in Figure [6] and the transformed images to be registered are matched by SIFT and the method of this paper. The correct rate and time statistics are shown in Table II. From the data in the table, we can see that the algorithm in this paper can get a
higher correct matching rate and matching accuracy. The total running time of this method is about 20% lower than that of SIFT, and it can still achieve high matching accuracy and matching accuracy when there are obvious angle and scale transformations in the image to be registered.

### Tab. 2 Comparison of the Registration Parameters of the Two Methods

| Image | Algorithm | Angle | Ratio | Matches | Correct Matches | Percent of correct matches | RMSE | Total time |
|-------|-----------|-------|-------|---------|----------------|---------------------------|-------|-------------|
| Group 1 | SIFT | 0° | 1 | 344 | 136 | 39.5% | 0.567 | 26.247s |
| | Propose | 0° | 1 | 71 | 57 | 80.2% | 0.456 | 16.672s |
| | SIFT | 10° | 1 | 270 | 78 | 28.9% | 0.560 | 27.104s |
| | Propose | 10° | 1 | 51 | 38 | 74.5% | 0.426 | 20.961s |
| | SIFT | 0° | 0.8 | 210 | 63 | 30.0% | 0.667 | 21.182s |
| | Propose | 0° | 0.8 | 53 | 41 | 77.3% | 0.608 | 17.329s |
| Group 2 | SIFT | 0° | 1 | 105 | 44 | 41.9% | 0.533 | 13.206s |
| | Propose | 0° | 1 | 43 | 35 | 81.4% | 0.387 | 12.877s |
| | SIFT | 10° | 1 | 92 | 36 | 39.1% | 0.588 | 13.223s |
| | Propose | 10° | 1 | 62 | 50 | 80.6% | 0.439 | 13.785s |
| | SIFT | 0° | 0.8 | 69 | 35 | 50.7% | 0.582 | 13.216s |
| | Propose | 0° | 0.8 | 27 | 21 | 77.7% | 0.562 | 11.117s |

Figure [7] is the visualization result of the first set of test image registration. It can be seen that the algorithm in this paper can effectively complete the registration of visible light images and infrared images, and the registration time is shorter and the registration accuracy is higher.
Fig. 7 Results of registration for Fig.6(a)(b) Conclusion

In this letter, the Harris corner point detection is performed on the edge response points of visible light and infrared images, and combined with Harris-AE network model to generate feature descriptors, which effectively overcomes the problem of different gray distributions and inconsistent texture features of the two source images. The preprocessing step of segmentation ensures the processing
efficiency of the algorithm. Experiments have proved that this algorithm improves the accuracy of matching while shortening the registration time.

Acknowledgments
This research was supported by Youth Program of the National Natural Science Foundation of China, by the General Program of National Natural Science Foundation of China and by Youth Program of the Natural Science Foundation of ShaanXi Province, China, under grant numbers 61806166, 62076204 and 2020JQ-197. I would like to extend my sincere gratitude to Professor Tian DaYong for his support and guidance.

References
[1] Ze-Tao J, Xiao-Yan L, Qi W. Visible and infrared image registration algorithm based on saliency and ORB[J]. Laser & Infrared, 2019.
[2] Li Hang, Lin Yi, Sun Erman. Research on the Method of UAV Multi-source Reconnaissance Image Fusion[J]. Ship Electronic Engineering, 2019.
[3] Zheng-Hua M A, Xiao-Jun L, Li-Ming X U, et al. Multi-source image registration for fruit recognition in close background [J]. Computer Engineering and Design, 2018.
[4] Yong-Li H U , Liang W , Rong L , et al. A Coarse-to-Fine Registration Method for Satellite Infrared Image and Visual Image [J]. Spectroscopy & Spectral Analysis, 2013, 33(11):2968-2972.
[5] Zhang Haitao, Jin Yan, Liu Wanjun. SIFT Remote Sensing Image Registration Algorithm based on Marr Wavelet [J].Remote Sensing Technology and Application, 2019,34(3):622-629.
[6] Hao Z, Yuan-Xin Y E, Lei W. A Multi-source Remote Sensing Image Registration Method Using Point Features and Mutual Information [J]. Remote Sensing Information, 2015.
[7] Hongmin Z, Jianshuang Z, Yongtao L, et al. A Fast SIFT Registration Method Based on Image Region Segmentation [J]. Infrared Technology, 2017.
[8] Liu Huan,Xiao Genfu, et al. Initial-fine Combination and Multi-feature Fusion for Multi-source Remote Sensing Image Registration [J]. Remote Sensing Information, 2018, 033(006):61-70.
[9] Zhang Yan, Sun Shi-yu, et al. Volume Multi-Source Images Registration Based on Improvement Model Fitting and Adjustment [J], Transactions of Beijing Institute of Technology. 2018, 38(10):75-80.
[10] AANAS H,DAHL A L,PEDERSEN K S, Interesing interest points [J]. International Journal of Computer Vision,2012,97(1):18-35.
[11] AANAS H,DAHL A L,PEDERSEN K S, Interesing interest points [J]. International Journal of Computer Vision,2012,97(1):18-35.
[12] Bin-Liang J , Man-Man F . Research on the multi-source remote sensing image registration based on the improved algorithm of SIFT [J]. Laser & Infrared, 2011.
[13] Rui T , Zhang S A , Zhou Y , et al. Registration of Multi-sensor Images Based on Harris Corner with SIFT Description[J]. Opto-Electronic Engineering, 2012, 39(8):26-31.
[14] Qiang W . A Research of Contour - based Registration Technique for Multi - Source Images [J]. Computer & Digital Engineering, 2002.
[15] Jin-Sha Y, Zhen-Bing Z , Qiang G , et al. Review and prospect on infrared/visible image registration [J]. Laser & Infrared, 2009, 39(7):693-699.
[16] Harris, Christopher G., and Mike Stephens. “A COMBINED CORNER AND EDGE DETECTOR.” Proc. 4th Alvey Vision Conf., Manchester, U.K., Aug. 1988, 1988, pp. 147–151.