Visible and infrared missile-borne image registration based on improved SIFT and joint features

Song Xue¹, Xiaorong Zhang², Hang Zhang³ and Chuandong Yang³

¹ Department of Weapons Engineering, Army Academy of Artillery and Air Defense, Hefei, Anhui, 230001, China
² Geological Survey of Anhui Province (Anhui Institute of Geological Science), Hefei, Anhui, 230001, China
³ Postgraduate Team, Army Academy of Artillery and Air Defense, Hefei, Anhui, 230001, China

*Corresponding author’s e-mail: xs_xs6688@sina.com

Abstract. For the acquired visible and infrared missile-borne images, in order to provide accurate and reliable image data for the subsequent processing of images to achieve complementary information, we carried out visible and infrared missile-borne image registration research. Firstly, the shape context is used to extract the rough edge features of the image. Then we improve the SIFT gradient definition to overcome the difference in image gray value, and combine with the coherent point drift algorithm to increase the number of correct matching points to achieve global image registration. Experimental results show that, compared with the existing registration methods, the proposed method has a good visual effect of registration, greatly improves the correct registration points and registration accuracy, and can better solve the problem of visible infrared missile-borne image registration.

1. Introduction

As the most commonly used image sources, visible and infrared sensors have been widely used in computer vision, pattern recognition, automatic control and other fields[1]. At present, based on the visible and infrared multi-source guidance, according to its imaging principle, missile borne images with their own advantages can be obtained. Therefore, they are combined to form a "visible light-infrared" fusion image, giving full play to their respective advantages, complementing each other, and improving the ability to detect and strike targets, which has great research value and application prospects[2].

Figure 1. Unregistered visible and infrared images and their fusion image. (a) visible image. (b) infrared image. (c) direct fusion image.
The premise of visible and infrared image fusion is that they need strict registration. Figure 1 shows the real shot image and direct fusion image of the missile-borne platform. Due to the difference of position and lens parameters of visible and infrared camera, and different imaging mechanism, the difference of gray value is significant, and the missile shaking caused by objective environment in the imaging process, the obtained visible light and infrared images have different scales and pixel deviation. Figure 1 shows the unregistered visible and infrared images of the same scene, and their fusion image. It can be found that the fused image has target ghosts. Therefore, it is necessary to register the visible and infrared images.

Image registration is the process of matching two or more of the same scene with different time, different viewing angles or different sensor images with respect to a specific reference image[3]. At present, researchers have done a lot of research on visible and infrared image registration, and the proposed methods can be divided into three categories: calibration parameter based method, region based method and feature based method[4-6]. The feature based method can adapt to various complex environments and has strong robustness. The common features include point, edge and contour. Among the feature-based methods, scale invariant feature transform (SIFT)[7] is the classic algorithm. It has been used successfully in image registration. Speeded up robust features (SURF)[8] can improve the computational efficiency. At present, these methods have been successfully applied to aerial remote sensing images similar to the perspective of the missile borne images. Li et al. introduced a scale-direction joint constraint criterion and proposed a robust SIFT(R-SIFT) remote sensing image registration method[9]; Kupfer et al. proposed a fast mode-seeking (MS-SIFT)[10]. In addition, the researchers also applied point features to the visible and infrared image registration[11-14]. However, the features of visible and infrared image is not consistent, and cannot be fully described. Directly using edge information will also reduce the registration accuracy.

In response to the above problems, the main work of this paper is divided into the following two aspects. Firstly, the shape context is used to establish the initial edge registration of visible and infrared image, which is used to describe the regularity of the difference between visible and infrared image and enhance the stability of registration; secondly, improve the SIFT point matching algorithm to extract joints in key positions, scales, and directions. The feature establishes a new matching criterion, and combines the coherent point drift algorithm to propose a registration method suitable for visible and infrared missile-borne images.

2. Image matching feature rough extraction based on edge

2.1. Problem description
Image edge features can be stored in visible and infrared image pairs[15-16]. Therefore, Canny edge detector is used to extract the binary edge features of the image. In essence, the edge of an image is composed of a set of discrete points inside and outside the scene, Define a set of discrete points: \( P = \{ p_1, p_2, \ldots, p_n \} \), \( p_i \) is the boundary point of the image. For the edge point set of visible image: \( X = \{ x_1, x_2, \ldots, x_n \} \), it is necessary to obtain the "best" matching point set \( Y = \{ y_1, y_2, \ldots, y_n \} \) at the edge of infrared image through transformation relation \( X = T(Y, \theta) \), Where \( \theta \) is the transformation parameter. Experience shows that the use of rich local descriptors can reduce the registration error, and the matching is relatively easier[17].

2.2. Edge feature extraction based on Shape Context
Define the shape context cost matching function:

\[
C_{ij} = C(x_i, y_j) = \frac{1}{2} \sum_{k=1}^{n} \frac{(x_i(k) - y_j(k))^2}{x_i(k) + y_j(k)} \tag{1}
\]

Where \( x_i(k) \) is the shape histogram of the edge point \( x_i \) of the visible image, \( y_j(k) \) is the shape histogram of the edge point \( y_j \) of the infrared image, \( C_{ij} \) essentially represents the corresponding
relationship matrix between the edge points of the visible image and the infrared image, and the lower \( C_{ij} \) is, the more similar the corresponding visible and infrared edge points are.

Applying the shape context directly to the visible and infrared image is not good, and only a rough match can be performed. There are two main reasons. One is that the edge characteristics of the visible and infrared image are very different due to the different imaging mechanisms. The clear edges of visible images are often inconspicuous or not reflected in infrared images, and matching errors often occur when seeking the minimum cost matching function; And, matching errors may also occur for image pairs with low scene offset. Figure 2 shows the shape context edge matching of Figure 1(a)(b). In Figure 2 (a) and (b), the correct matching point of point a in the visible edge image should be point B in the infrared edge image, while the application of shape context mistakenly thinks that point C in the infrared edge image is the matching point of point A. It can be seen from the local enlarged figure of Fig. 2(c) that compared with point A, the structure of point B has obvious bending distortion, while point C is more similar to point A. Therefore, to solve this problem, it is necessary to further optimize the matching, eliminate the interference matching features, and seek more accurate matching criteria.

![Figure 2. Shape context edge matching. (a) visible image edge map. (b) infrared image edge map. (c) partial enlarged view at points A, B and C.](image)

### 3. Exact image matching based on joint features

#### 3.1. Basic principles of SIFT

Sift is a kind of local feature descriptor. It has scale invariance and has good adaptability to illumination change, noise distortion and micro angle transformation. Based on Its theory, the scale histogram of all sift key points matching between visible and infrared image is calculated to obtain the mode scale \( s_{mode} \). Similarly, all the matching direction difference histograms are calculated, and the pattern rotation difference \( \Delta \theta_{mode} \) is found. Then let \((x, y)\) and \((x', y')\) denote the coordinates of the corresponding SIFT key points in the visible image and the infrared image, and each pair of corresponding key points defines the following horizontal and vertical displacements:

\[
\Delta x = x - s_{mode} (x' \cos(\Delta \theta_{mode}) - y' \sin(\Delta \theta_{mode}))
\]

\[
\Delta y = y - s_{mode} (y' \sin(\Delta \theta_{mode}) + x' \cos(\Delta \theta_{mode}))
\]

Figure 3 is a histogram of the scale ratio, main orientation difference, horizontal offset and vertical offset of the original SIFT matching key points in Figure 1 (a) and (b). The parameter settings of the histogram are referenced in[10]. It can be seen from Figure 3 that the four histograms cannot clearly show a single pattern. Due to the different formation mechanisms of visible light images and infrared images, the gray values of the same region are different, which leads to significant non-linear intensity difference[9]. This intensity difference will greatly affect the image registration results. And the feature descriptor is not robust to these differences.
3.2. Improved gradient algorithm

The different intensities and contrasts of visible and infrared images will make the gradient magnitude and direction of the same area in the image very different. Therefore, the gradient calculation method of the image is redefined in the Gaussian scale space. First use the Sobel filter to calculate the gradient amplitude of the Gaussian scale space image:

$$G_\sigma = \sqrt{(M_{x,\sigma})^2 + (M_{y,\sigma})^2}$$

(4)

Where $\sigma$ represents the Gaussian space image scale, $M_{x,\sigma}$ and $M_{y,\sigma}$ represent the horizontal and vertical derivatives of the Gaussian scale space image, respectively. Then calculate the new gradient magnitude and gradient direction:

$$G'_{\sigma} = \sqrt{(G_{x,\sigma})^2 + (G_{y,\sigma})^2}$$

(5)

$$R = \arctan\left(\frac{G_{y,\sigma}}{G_{x,\sigma}}\right)$$

(6)

Where $G_{x,\sigma}$ and $G_{y,\sigma}$ represent the horizontal and vertical derivatives of the Gaussian scale spatial gradient amplitude image $G_{\sigma}$ respectively. Previous studies have shown that Gradient Location-Orientation Histogram (GLOH) can increase the robustness and uniqueness of feature descriptors[18]. Therefore, in reference[19], GLOH-based circular neighborhoods and log-polar coordinates are used to replace the square neighborhoods and coordinates of the original SIFT.

Figure 3. Histograms of original sift matching key points. (a) scale ratio. (b) main orientation difference. (c) horizontal offset. (d) vertical offset.

Figure 4 shows the histograms of the key point scale ratio, main direction difference, horizontal offset and vertical offset obtained from formula (6) and (7). These histograms show obvious peaking effects. The peak position of the scale ratio histogram is $r = 0.9322$, the peak position of the main orientation difference histogram is $\Delta \theta = 1.23^\circ$, and the peak positions of the horizontal and vertical translation histograms are $\Delta x = -97.35$ and $\Delta y = -99.13$, respectively.
3.3. Joint feature matching

Generally, the similar transformation of image registration is divided into scale, translation and rotation to achieve feature matching by defining position, scale and direction information. The feature point sets obtained by improving the SIFT gradient to obtain the edge feature extraction of visible light and infrared images are \(E_1, \ldots, E_M\) and \(E'_1, \ldots, E'_M\) respectively. Reference [23] defines that the position, scale and main orientation of the \(i\)th feature point in the visible image are \((x_i, y_i, \sigma_i, \theta_i)\) respectively, and in the infrared image are \((x'_i, y'_i, \sigma'_i, \theta'_i)\) respectively. Then the position transformation error between two points is:

\[
e_p(p_i, p'_j) = \| (x_i, y_i) - T((x'_i, y'_i), \mu) \| \tag{7}
\]

Where \(T\) is the transformation model and \(\mu\) is the transformation parameter. Reference [24] defines scale error and main orientation error:

\[
e_s(p_i, p'_j) = 1 - r' \frac{\sigma'_i}{\sigma_i} \tag{8}
\]

\[
e_o(p_i, p'_j) = \text{abs}((\theta_i - \theta'_i) - \Delta \theta'_i) \tag{9}
\]

Where \(r^*\) is the peak position of scale ratio histogram and \(\Delta \theta'_i\) is the peak position of principal direction difference histogram. The research of this paper is visible and infrared missile-borne image, and the image acquisition equipment has been calibrated so the image has the same scale ratio, horizontal displacement, vertical displacement, and the change of rotation angle in the main direction is small. Therefore, a new reduce orientation position scale Euclidean distance (ROPSED) is defined:

\[
R_{O,PSED} = \frac{(1 + e_p(p_i, p'_j))(1 + e_s(p_i, p'_j))}{1 + e_o(p_i, p'_j)} ED(p_i, p'_j) \tag{10}
\]

Where ED is the Euclidean distance of the corresponding descriptor between two points. The specific matching process is as follows:

3.3.1. Initial matching. The shape context of the visible image and the infrared image is calculated to obtain the attribute matrix and generate the rough matching key points. The ratio of the distance between the nearest neighbor and the second nearest neighbor is used for matching, and the threshold value of the ratio is set to \(d\). Pattern position \(R^*, \Delta \theta^*_i, \Delta x^*_i, \Delta y^*_i\) are obtained by joint features matching. According to \(R_{O,PSED}\), position transformation error, scale error, principal direction error and initial transformation parameters \(\theta\) are obtained.

3.3.2. Rematching. Match the key points of the feature again and define the transformation relationship \(P'_k = P_k(T, \theta)\). Establish the probability density function of the Gaussian mixture model:

\[
p(P_k) = \sum_{m=1}^{M} \frac{1}{M} g(x|m) \tag{11}
\]

\[
g(x|m) = \frac{1}{(2\pi \sigma^2)^{d/2}} \exp \left(-\frac{|x-M|^2}{2\sigma^2}\right) \tag{12}
\]

Where \(m = 1, 2, \ldots, M\). In order to avoid the influence of outliers, uniform distribution is introduced: \(g(x|M+1) = 1/N\). Define the weight parameter \(\omega\) to establish the probability density function:

\[
g(x) = \omega \frac{1}{N} + (1-\omega) \sum_{m=1}^{M} \frac{1}{M} g(x|m) \tag{13}
\]

Establish a maximum likelihood function for each point:

\[
L(\theta, \sigma^2) = - \sum_{n=1}^{N} \log \sum_{m=1}^{M} p(P_n) g(x|m) \tag{14}
\]
Where \( n = 1, 2, \ldots, N \). In reference [23], the expectation maximization algorithm is used to calculate the parameters alternately until convergence, and the objective function is established:

\[
Q(\theta, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} P_{E}^{\text{old}}(m|x_n) \|x_n - T(y_m, \theta)\|^2 + \frac{N_R D}{2} \log \sigma^2 \tag{15}
\]

Where \( N_R = \sum_{n=1}^{N} \sum_{m=1}^{M} P_{E}^{\text{old}}(m|x_n) \leq N \), \( P_{E}^{\text{old}}(m|x_n) \) is the posterior probability distribution of the old parameter value (superscript "old").

Define the transformation relationship of the center position as:

\[
T(\theta e, e_x, e_y) = \theta e_x y + e_x + e_y,
\]

and establish the objective function:

\[
Q(e, e_x, e_y, \sigma^2) = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} P_{E}^{\text{old}}(m|x_n) \|x_n - \theta e_x y - e_x - e_y\|^2 + \frac{N_R D}{2} \log \sigma^2 \tag{16}
\]

Minimiz formula (16) is to solve the minimum error, and the minimum error is brought into formula (10) to solve the minimum distance, and the point set corresponding to the minimum distance is recorded as \( \{R_{E}^{*}\} \).

3.3.3. Outlier removal. There will be some false correspondences in \( \{R_{E}^{*}\} \). The method proposed in MS-SIFT[10] is used. Let \( (x_1, y_1) \) and \( (x_2, y_2) \) denote the coordinates of corresponding keypoints in set \( \{R_{E}^{*}\} \). The horizontal and vertical displacements of corresponding keypoints are defined as:

\[
\begin{align*}
\Delta x_1 &= x_1 - r^* (x_2 \cos(\Delta \theta^*) - y_2 \sin(\Delta \theta^*)) \\
\Delta y_1 &= y_1 - r^* (x_2 \sin(\Delta \theta^*) + y_2 \cos(\Delta \theta^*))
\end{align*}
\tag{17}
\]

Then, most of the outliers are eliminated according to the following logical filter:

\[
\left| \Delta x^* - \Delta x_1 \right| \geq \Delta x_{th}, \left| \Delta y^* - \Delta y_1 \right| \geq \Delta y_{th}
\tag{18}
\]

where \( \Delta x_{th} \) and \( \Delta y_{th} \) denote the thresholds of horizontal and vertical differences respectively. Finally use ROPSED again to get the correct set of key points.

4. Experimental results and analysis

4.1. Test image

The experiments verify the two kinds images. The first is the missile borne simulation platform visible and infrared image; The second is aerial visible and infrared image similar to the missile-borne image type. The images are shown in the column of Figure 5(a)(b), and the numbers are 1~4. Root Mean Square Error(RMSE) is selected as the criterion.
4.2. Test results and analysis

The registration image and checkerboard mosaiced image are shown in Fig. 5(c) and (d). It can be seen that the region and edge of the same position of the image are clearly overlapped, which has a good visual effect, showing the accuracy of the algorithm.

In order to increase the persuasiveness of the algorithm, this paper compares the proposed method with SIFT-FSC, SURF and MS-SIFT to test the RMSE, matching point and time costs of four groups images in Figure 5. The test results are shown in Table 1.

| Method       | Figure 5(a)1 400×300 | Figure 5(a)2 300×250 | Figure 5(a)3 300×250 | Figure 5(a)4 300×250 |
|--------------|-----------------------|----------------------|----------------------|----------------------|
|              | RMSE                  | Point                | Time(s)              | RMSE                  | Point                | Time(s)              | RMSE                  | Point                | Time(s)              |
| SIFT-FSC     | ×                     | ×                    | ×                    | 0.2917                | 5                    | 7.128                | 0.4886                | 6                    | 7.016                | ×                    | ×                    |
| SURF         | ×                     | ×                    | ×                    | ×                     | ×                    | 0.7633               | 30                   | 1.7                  | ×                    | ×                    | ×                    |
| MS-SIFT      | 0.4465                | 5                    | 10.4                 | 0.6329                | 7                    | 6.5                  | 0.6769                | 12                   | 5.3                  | 0.7152               | 9                    | 5.1                  |
| Proposed method | 0.5872               | 45                   | 17.1                 | 0.5389                | 422                  | 25.2                 | 0.5407                | 595                  | 41.4                 | 0.5500               | 282                  | 43.0                 |

From Table 1, we can see that the proposed method significantly enhances the number of correct feature matching points compared with the comparison algorithm, but increases the operation time of the algorithm accordingly. Because of the realization of SURF based on C++, the efficiency is very high. For some images, the RMSE value is slightly higher, but it also achieves sub-pixel accuracy, which proves the effectiveness of the algorithm. Figure 6 shows the feature points matching results in Figure 5 (a)(b) using the proposed method, which are represented by lines.
5. Conclusion
Aiming at the problem of visible and infrared missile-borne image registration, this paper proposes an improved gradient and joint feature matching image registration method. The test results show that the algorithm is better than the existing methods in registration accuracy and correct matching points, and has good visual effect, which can provide certain data support for the design of missile-borne image equipment. In the future, more real images will be collected to enhance the efficiency of the algorithm and improve the real-time performance of the algorithm.

References
[1] Jiang Z T, Liu X Y, Wang Q. (2019) Visible and infrared image registration algorithm based on saliency and ORB. Laser & Infrared, 49: 251-256.
[2] Xue S, Qian L Z, Zhang H, Yang C D (2021). Review of Key Technologies for Target Detection of Multi-source Terminal Missile-based Fusion Image. http://kns.cnki.net/kcms/detail/61.1234.TJ.20201224.1357.002.html.
[3] Nag S. (2017) Image registration techniques: A survey. arXiv preprint, 28: 1712.07540.
[4] Chen S W, Zhang S X, Yang X G. (2017) Registration of visual-infrared images based on ellipse symmetrical orientation moment. Chinese Journal of Engineering, 39: 1107-1113.
[5] Zhao H S, Zhang Z Y. (2020) Power Equipment Infrared and Visible Images Registration Based on Cultural Wolf Pack Algorithm. Acta Optica Sinica, 40: 1-14.
[6] Jiang Q, Liu Y D, Fang J. (2020) Registration method for power equipment infrared and visible images based on contour feature. Chinese Journal of Scientific Instrument, 41: 252-260.
[7] Lowe G D. (2004) Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 60: 91-110.
[8] Bay H, Tuytelaars T, Gool L V. (2008) SURF: Speeded up robust features. Comput. Vis. Image Understand, 110: 346–359.
[9] Li Q, Wang G, Liu J. (2009) Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. IEEE Geoscience & Remote Sensing Letters, 6:287-291.
[10] Kupfer B, Netanyahu N S, Shimshoni I. (2014) An Efficient SIFT-Based Mode-Seeking Algorithm for Sub-Pixel Registration of Remotely Sensed Images. IEEE Geoscience & Remote Sensing Letters, 12: 379-383.
[11] Liu C, Cui T, He C L. (2016) Infrared-Visual Image Registration Based on High-Curvature Feature Matching. Command Information System and Technology, 7: 13-17.
[12] Wang H, Wei M. (2017) CCD-IR image registration based on adaptive feature point detection. Journal of Image and Graphics, 22: 179-187.
[13] Klimaszewski J, Kondej M, Kawecki M. (2013) Image Processing and Communications Challenges. Springer Berlin Heidelberg, Berlin.
[14] Ma J, Zhao J, Ma Y. (2015) Non-rigid visible and infrared face registration via regularized Gaussian fields criterion. Pattern Recognition, 48: 772-784.
[15] Ma J, Chen C, Li C. (2016) Infrared and visible image fusion via gradient transfer and total variation minimization. Information Fusion, 31: 100-109.
[16] Coiras, Enrique. (2000) Segment-based registration technique for visual-infrared images. Optical Engineering, 39: 282-289.

[17] Jones D G, Malik J. (1992) Computational framework for determining stereo correspondence from a set of linear spatial filters. Elsevier, 10: 699-708.

[18] Mikolajczyk K, Schmid C. (2005) A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27: 1615-1630.

[19] Dellinger F, Delon J, Gousseau Y. (2013) Sar-sift: a sift-like algorithm for sar images. IEEE Transactions on Geoscience & Remote Sensing, 53:453-466.

[20] Wu Y, Ma W P, Gong M G. (2015) A novel point-matching algorithm based on fast sample consensus for image registration. IEEE Geoscience and Remote Sensing Letters, 12: 43-47.

[21] Ma W P, Wen Z L, Wu Y. (2016) Remote Sensing Image Registration With Modified SIFT and Enhanced Feature Matching. IEEE Geoscience & Remote Sensing Letters, 14: 3-7.

[22] Li Q L, Wang G Y, Liu J G. (2009) Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. IEEE Geoscience & Remote Sensing Letters, 6: 287-291.

[23] Myronenko A, Song X B. (2010) Point Set Registration: Coherent Point Drift. IEEE Transactions on Pattern Analysis & Machine Intelligence, 32: 2262-2275.