

RESEARCH

Open Access



# Heterogeneous scene matching based on the gradient direction distribution field

Qingge Li<sup>1</sup>, Ruitao Lu<sup>1,2\*</sup> , Xiaogang Yang<sup>1</sup>, Siyu Wang<sup>1</sup>, Tong Shen<sup>1</sup>, Wenxin Xia<sup>1</sup> and Zhaoying Wei<sup>3</sup>

\*Correspondence:  
lrt19880220@163.com

<sup>1</sup> Rocket Force University  
of Engineering, Xi'an 710000,  
China

<sup>2</sup> Science and Technology  
on Electro-Optic Control  
Laboratory, Luoyang 471009,  
China

<sup>3</sup> College of Science, Xi'an Shiyou  
University, Xi'an 710065, China

## Abstract

Heterogeneous scene matching is a key technology in the field of computer vision. The image rotation problem is popular and difficult in the field of heterogeneous scene matching. In this paper, a heterogeneous scene matching method based on the gradient direction distribution field is proposed, and distributed field theory is introduced into heterogeneous scene matching for the first time. First, the distribution field of the gradient direction is constructed and fuzzified, and then the effective regions are selected. Then, the distribution field of the main direction is defined to solve the matching errors due to the existence of rotational transformations between heterogeneous source images. Third, the chi-square distance is introduced as a similarity measure. Finally, the hill-climbing method search strategy, which greatly improves the efficiency of the algorithm, is adopted. Experimental results on 8 pairs of infrared and visible heterogeneous images demonstrate that the proposed method outperforms the other state-of-the-art region-based matching methods in terms of the robustness, accuracy, and real-time performance.

**Keywords:** Heterogeneous images, Scene matching, Distribution field, Hill-climbing method

## 1 Introduction

With the rapid progress of informatization worldwide, the demand for image information has become increasingly strong. In recent years, image matching detection technology has become a popular research in the field of computer vision [1–3] and is widely applied to various fields, such as image retrieval [4], image understanding [5], multiagent cooperation [6, 7] and target detection [8, 9]. Under different task conditions (such as climate, light intensity, shooting position, and angle), image information often has to be acquired through different sensors, and these images generally have differences in gray value, resolution, scale, or nonlinear distortion. It is a difficult research task to achieve accurate matching of heterogeneous source images in complex environments.

Over the years, various heterogeneous image matching methods have been proposed. In general, it seems that heterogeneous image matching methods are mainly classified into region-based matching methods, feature-based matching methods, and artificial neural network-based matching algorithms. Region-based matching algorithms directly or indirectly use the grayscale information of a region in an image as the basis of the

feature space and similarity metric, use the similarity metric algorithm to determine the correspondence between the region and the image that is to be matched, and find the best matching position globally. The commonly used region-based matching algorithms are the grayscale correlation method [10–12], maximum mutual information correlation method [13–15], and gradient correlation method [16]. However, these methods cannot solve the matching problem when the image is rotated and scaled.

Feature-based matching methods mainly use extracted local features such as the points [17, 18], lines [19–21], and faces [22] of an image to achieve matching. Regarding local invariant descriptor-based matching methods [23], the common descriptors are scale invariant feature transform (SIFT) [24], speeded-up robust features (SURF) [25], and oriented FAST and rotated BRIEF (ORB) [26]. The SURF algorithm is influenced by SIFT, which greatly increases the matching speed. The ORB algorithm extracts feature points quickly, but it is more sensitive than other algorithms when there are large rotations and scale changes between images. In the same year [27], proposed binary robust invariant scalable keypoints (BRISK), a binary feature description operator that has excellent rotation invariance, scale invariance, and good matching results for images with significant blur. The feature-based method has good adaptability to the geometric deformation, brightness variation and noise effects of images and has high accuracy. However, the manually designed feature descriptors do not describe the detected features well, have weak generalization ability, lack high-level semantic information, and have certain limitations.

Recently, artificial neural network (ANN)-based matching algorithms [28–33] have been rapidly developed. Representative methods include BP neural network-based image matching methods [34, 35], Hopfield network-based image matching methods [36], annealing algorithm-based image matching methods [37], genetic algorithm-based image matching methods [38], and twin network-based matching methods [39–42]. The artificial neural network-based matching algorithm first preprocesses the image using some image representation algorithm and extracts a certain number of image information features as needed. Then, according to the requirements of some constructed neural networks, some initial state information parameters needed by the network are selected and input, and the selected image features are passed to the neural network as the basic input parameters to start the iterative solving process of the neural network algorithm to complete the recognition matching or localization of the baseline and real-time images. However, the neural network has more parameters, the amount of data required for model training is large, the learning process is relatively long, and it may fall into local minima.

In the literature [43], Laura Sevilla-Lara proposed the application of distribution fields (DF) to the field of tracking with good results. The distribution field contains not only the grayscale information of the image but also the grayscale position information. Therefore, the distribution field map is a fusion of position information and grayscale information. Usually, when matching, the image group is blurred. However, the common blurring techniques that are currently used have much important information of the image lost inside, which leads to the failure of matching. The blurring of the distribution field map loses almost no information of the image and is a lossless blurring. In addition, this blurring process increases the robustness of matching, making it possible

to successfully match even if there are small distortions and rotations in the real-time image. Thus, distribution field theory is applied in this paper to the heterogeneous image matching problem and successfully achieves heterogeneous image matching by constructing a distribution field in the gradient direction to describe the heterogeneous image.

Based on the above considerations, a heterogeneous image matching method is proposed in this paper based on the gradient direction distribution field, which introduces the direction distribution field into the heterogeneous image matching process for the first time. By constructing the gradient direction distribution field and defining the main peak of the regional gradient direction histogram as the main direction of the distribution field, the rotation transformation problem of heterogeneous image matching can be solved well. We have conducted a series of experiments on infrared (IR) and visible heterogeneous images, and the results show that our method has good performance in terms of the robustness and detection accuracy.

The rest of the paper is organized as follows. In Sect. 2, we present the framework of our method and the formulation of the proposed method. We conduct a series of experiments on infrared and visible heterogeneous images, and three prior methods are compared to our approach in Sect. 3. The conclusions of this study are presented in Sect. 4.

## 2 Methods

In this paper, distributed field theory is applied to the heterogeneous source image matching problem, focusing on the problem of difficult matching caused by rotational transformation. The images to be matched are infrared images (template images) and visible images (real-time images). First, the gradient direction DFs of the template image and the real-time image are constructed, and the robustness of matching is enhanced by fuzzy filtering. Second, the best-matched real-time sub-image is searched in the real-time image using the hill-climbing method. The real-time sub-images are centered on the hill-climbing nodes or sub-nodes, whose sizes are the same as the template image. Then, the main direction DFs of the template image and the real-time sub-images are obtained separately and described by a one-dimensional vector. Finally, the similarity between the template image and the real-time sub-images is calculated using the chi-square distance and stored in the correlation matrix. The workflow of the proposed matching algorithm is shown in Fig. 1.

### 2.1 Description of the distribution fields

A distributed field is a combination of each pixel distributed in the corresponding field, which is a division of pixel points into gray levels. This distribution defines the probability information of a pixel to appear on each feature map. Taking a grayscale image as an example, the grayscale level is 0–256, the 256 grayscale levels can be divided into  $N$  intervals, and the pixel points corresponding to each grayscale interval contain not only grayscale information but also location information.

The distribution field map of an image can be represented as an  $2 + N$  dimensional matrix  $d$ , with the 2 dimensions representing the length and width of the image and the other  $N$  dimensions representing the set number of feature space dimensions. In other

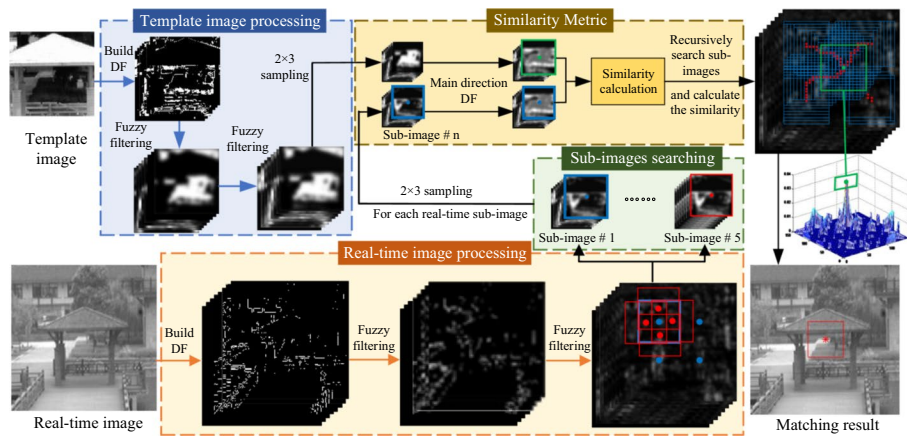


Fig. 1 Flowchart of the proposed method

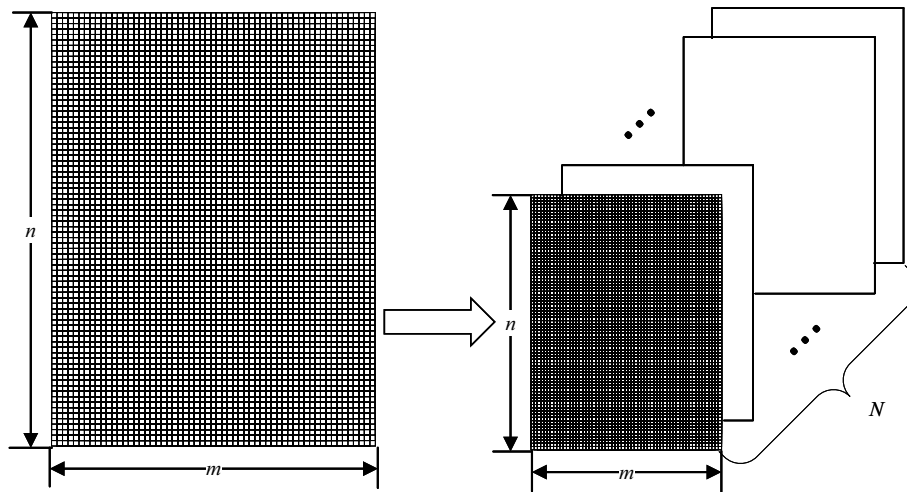


Fig. 2 Schematic diagram of the distribution field

words, if the size of an image is  $m \times n$ , then its distribution field map  $d$  is represented as a  $m \times n \times N$  3-dimensional matrix. The distribution field is shown in Fig. 2.

Calculating the distribution field map of an image is equivalent to calculating the Kronecker delta pulse function at the geometric location of each pixel. It can be formulated by

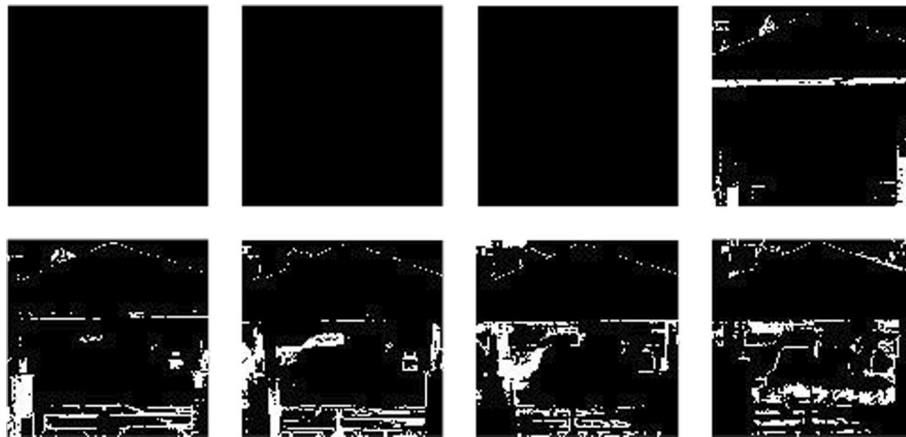
$$d(i, j, k) = \begin{cases} 1 & I(i, j) = k \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where  $I(i, j) = k$  is the gray value of the pixel with the coordinate  $(i, j)$  in the image, and  $d(i, j, k)$  is the value of the pixel with the coordinate  $(i, j)$  in the image on the  $k$  feature layer. It follows that  $d(i, j, k)$  takes the value of 1 or 0 and the values at each position  $(i, j)$  in the  $K$  layer sum to 1:

$$\sum_{k=1}^N d(i, j, k) = 1. \tag{2}$$



**Fig. 3** Example image



**Fig. 4** Eight feature layers of the target

The target shown in Fig. 3 is used as an example to analyze its distribution field. Since the distribution field needs to be fuzzy for each layer and for the convenience of calculation, the field distribution map of the target is calculated in the square area in this paper.

Figure 4 shows the individual feature layers of the target of Fig. 3. To understand the feature layers more intuitively, the 256 Gy levels of the image are compressed to 8, so there are 8 feature layers, with Layers 1 to 4 in the first row from left to right and Layers 5 to 8 in the second row from left to right.

As seen from Fig. 4, an image can be represented as a layer distribution field map, but most of the information of the image is not lost. This is the first step in constructing the distribution field map, which is equivalent to redescribing the original image. Next, to prevent the location information from losing its generality, the image needs to be blurred, i.e., Gaussian convolutional filtering is introduced for both horizontal and vertical Gaussian filtering of the distribution field map.

The first transverse filtering is performed, and  $d_s(k)$  is obtained after convolution of the  $k$  feature layer:

$$d_s(k) = d(k) * h_{\sigma_s}, \quad (3)$$

where  $d_s(k)$  denotes the new feature layer after the  $k$  feature layer is convolved with the Gaussian filter;  $d(k)$  is the feature layer before convolution;  $h$  is a two-dimensional Gaussian filter with the standard deviation  $\sigma_s$ ; and  $*$  is the convolution symbol.

Figure 5 shows the effect of convolving each of the eight feature layers with a Gaussian filter with a standard deviation of 9 pixels.

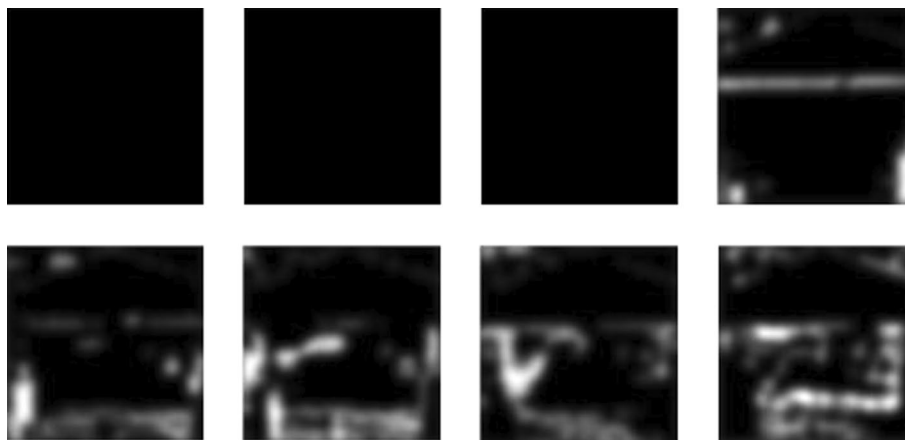
Compared with Fig. 4, it can be seen that before convolution, if the value of a position on the  $k$  feature layer is 1, it indicates that the gray value at this position on the original image falls in the  $k$  interval of  $N$  intervals; after convolution, if the value of a position on the  $k$  feature layer is not 0, it indicates that the gray value at a position near this position on the original image falls in the  $k$  interval of  $N$  intervals. This shows that Gaussian filtering of the feature layer introduces the uncertainty of the position into the field distribution map. This method only loses the exact position information and does not lose the grayscale information in the original image. This will have some effect on the matching error during the matching process and can enhance the robustness of the algorithm, making it possible to successfully match even in the presence of small rotational transformations.

In Eq. (3), if the Gaussian function  $h_{\sigma_s}$  is considered a probability distribution function, then after convolution,  $d_s(k)$  satisfies the properties of  $\sum_{k=1}^N d_s(i, j, k) = 1$  and still satisfies the requirements of the distribution field.

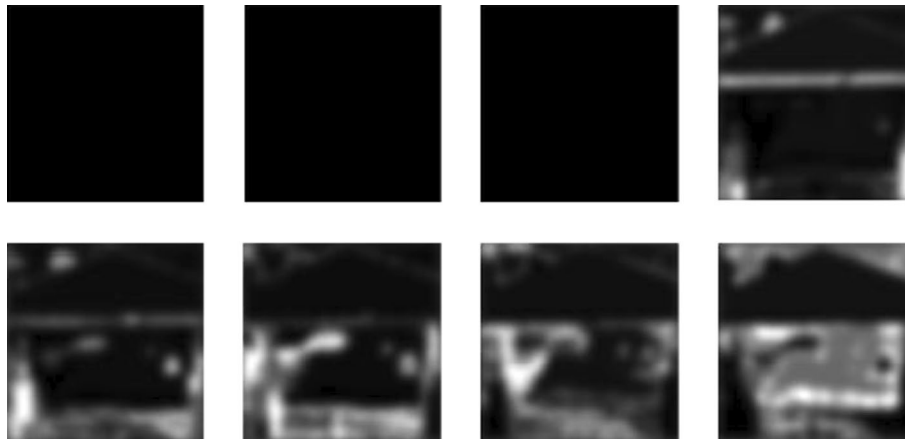
The Gaussian filtering of the  $x$  and  $y$  coordinate directions of each distribution field feature layer increases the uncertainty of the position in the above discussion. Based on the same thinking consideration, Gaussian filtering of the distribution field feature space can be understood as Gaussian filtering of the  $z$  coordinate direction to increase the uncertainty of the features. In this way, theoretically blurring the distribution of grayscale information in a certain layer of the distribution field allows the description of the image to adapt to the motion of subpixels and partial brightness variations, which can enhance the robustness of the algorithm to some extent. Therefore, it is next necessary to filter the feature layer with a one-dimensional Gaussian filter:

$$d_{ss}(i, j) = d_s(i, j) * h_{\sigma_f}, \quad (4)$$

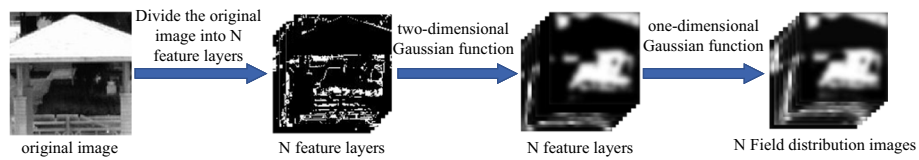
where  $h$  in the above equation is a one-dimensional Gaussian filter with the standard deviation  $\sigma_f$ . The final field distribution obtained from the example image of Fig. 3 is shown in Fig. 6.



**Fig. 5** Feature layers of the image after the first convolution



**Fig. 6** Each feature layer of the small target after the second convolution



**Fig. 7** Field distribution image calculation process

At this point, the field distribution map of an image is calculated, and the calculation is shown in Fig. 7. From the calculation, it can be summarized that the process of calculating the distribution field map is the process of introducing uncertainty into the field distribution map: first, convolution in the direction of the two coordinate axes of the image introduces the uncertainty of the position; second, convolution in the feature space introduces the uncertainty of the grayscale information. In other words, the image represented using the distribution field map is insensitive to smaller position changes and grayscale changes and has good adaptability to position translations, rotations and occlusions within a certain range.

### 2.2 Construction of the gradient direction DF

For any 2D image  $I_{x,y}$ ,  $\nabla I_x = \partial I / \partial x$  and  $\nabla I_y = \partial I / \partial y$  are its corresponding horizontal and vertical direction gradients, which can be obtained through common first-order or second-order differential operators, such as the Roberts operator, Sobel operator, and Prewitt operator. In this paper, we do not need to denoise the image [44], but the flat region with a small gradient is regarded as a background susceptible to noise interference, and its gradient direction is defined as 0. The true 0-gradient direction is defined as  $\pi$ , and then the gradient direction is quantized to  $[0,180]$ , which is expressed by the following equation:

$$\theta_{x,y}^I = \begin{cases} \text{angle}(V_{x,y}) & \text{if } (\nabla I_y \neq 0 \cap G_{x,y} > \tau) \\ \pi & \text{if } (\nabla I_y = 0 \cap G_{x,y} > \tau) \\ 0 & \text{if } G_{x,y} < \tau \end{cases}, \tag{5}$$



$$V_{x,y} = \text{sign}(\nabla I_y) \cdot (\nabla I_x + i\nabla I_y), G_{x,y} = |V_{x,y}|, \tag{6}$$

where  $\text{angle}(x)$  is the phase angle finding function of vector  $x$ ;  $\text{sign}(\nabla I_y)$  is the gradient sign in the vertical direction;  $i$  is the complex unit;  $G_{x,y}$  is the gradient amplitude; and  $\tau$  is the gradient amplitude threshold, which is used to distinguish the low amplitude flat region from the effective gradient region. Moreover, this algorithm can take  $\tau \in [0.1 \sim 0.4]$ .

Next, we construct a DF for the gradient direction, because the image rotation will generate a part of the 0 region at the edge. To prevent this region from affecting the matching accuracy and robustness, we only construct a DF for the points with a gradient direction greater than zero. Taking  $N=18$ , we divide  $[1,180]$  into 18 levels equally, and each level corresponds to one layer of DF; that is, any image  $I_{x,y}$  will be represented as 18 layers of DF, and the value of each point in the first layer indicates the probability that the gradient direction of  $I_{x,y}$  is included in the range of  $[1, 10]$ . In addition, we can construct 18 layers of gradient direction DF, that is

$$d(i, j, k) = \begin{cases} 1 & \theta_{ij}^I \in [10 * (k - 1) + 1, 10 * k] \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Finally, the DF feature space is filtered to introduce the ambiguity of the position and the ambiguity of the gray intensity into the distribution field map. This process only loses the exact information and does not introduce the wrong position information into the DF. In the case of smaller deformations, matching still occurs correctly, which enhances the robustness.

### 2.3 Main direction DF

The image principal direction characterizes the orientation of the image content and is a subjective concept in image processing. It can be defined as the texture direction of an image, the direction of a backbone, or the direction of a family of gradient vectors, and this artificially defined direction feature is sufficient as long as it has stable rotational invariance. The principal direction difference between two images characterizes the rotation angle between the images, according to whether the images can be rotationally corrected and then if the search is matched.

The classical gradient direction histogram-based principal direction estimation method is the most widely used. The method counts the gradient direction distribution (histogram) within a rectangular region and defines the most numerous classes of directions (main peaks) as the principal direction of the region.

Similar to the histogram statistics, the main direction of the DF is defined in this paper as the DF feature layer with the largest sum of probabilities of occurrence in the gradient direction, denoted by  $n$ . The calculation process is as follows:

$$dsum_k = \sum_{i=1, j=1}^{i=m, j=n} d_s(i, j, k) \quad k = 1, 2, \dots, 18, \tag{8}$$

$$[mlysum, n] = \max(dsum), \tag{9}$$



where  $\text{dsum}_k$  is the probability statistics of the DF at the  $k$  layer;  $\text{dsum}$  is the matrix storing the probability sum of each DF layer;  $\text{mlysum}$  is the maximum value in  $\text{dsum}$ ; and  $n$  is the value of  $k$  corresponding to the maximum  $\text{dsum}$ .

## 2.4 Similarity metric

The previous section describes how to determine the principal direction  $R$  of the template image and the principal direction  $R'$  of the real-time sub-image, and the approximate rotation angle of the real-time sub-image with respect to the template image can be obtained from the difference  $\nabla R = |R - R'|$  between the two. The template image is rotated  $\nabla R$  and  $\nabla R + 180$  to construct the DF and described by a one-dimensional column vector, denoted as  $x$ . The feature vector of the real-time sub-image is denoted by  $y$ .

There are many methods used to measure the correlation of two feature vectors, such as the Euclidean distance, Marxian distance, parametric and Eulerian distance, which have their own advantages and disadvantages and cannot be fully applied to the method in this paper. Moreover [45], introduces the chi-square distance formula, which can achieve good results in measuring the similarity of two eigenvectors:

$$\chi^2(x, y) = \sum \frac{(x_i - y_i)^2}{(x_i + y_i)}, \quad (10)$$

where  $\chi^2(x, y)$  denotes the chi-square distance of two vectors  $x, y$ , and  $x, y$  are the corresponding elements in the two vectors. From the equation, the chi-square distance calculates the ratio of the variance of the corresponding elements to the sum of the elements, and a smaller ratio indicates that the closer the distance is, the higher the similarity, while the Euclidean distance only considers the difference in the corresponding elements. During the experiment, it was found that the robustness of using the Euclidean distance as the similarity discriminator could not meet the matching requirements, and mismatching occurred, while the chi-square distance could better meet the requirements of the method in this paper.

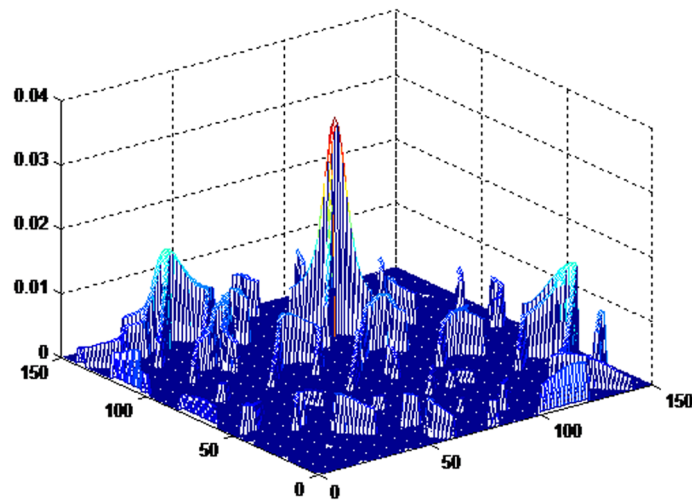
## 2.5 Hill-climbing method

To improve the operation speed and enhance the practicality of the matching algorithm in this paper, the hill-climbing method is used for fast search. The algorithm in this paper uses the chi-square distance as the similarity measure, and a larger value indicates a lower similarity of two images. Thus, the chi-square distance is further processed in this paper to invert the correlation surface. Furthermore, the best matching point can be seen more intuitively.

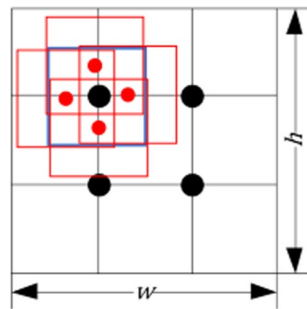
$$\text{dist} = \exp(k\chi^2) \quad \theta \in (0, 360). \quad (11)$$

Inverting the correlation surface diagram using Eq. (11) enables better visualization of the best matching points in the correlation surface diagram. Figure 8 gives a schematic diagram of the correlation surface for the hill-climbing method.

The initial hill-climbing nodes in the DF of the real-time image are shown in Fig. 9. The blue window represents the real-time sub-image, the black nodes are the initial hill-climbing nodes, and the red nodes are the hill-climbing sub-nodes. The hill-climbing



**Fig. 8** Example of the correlation surface to the hill-climbing method

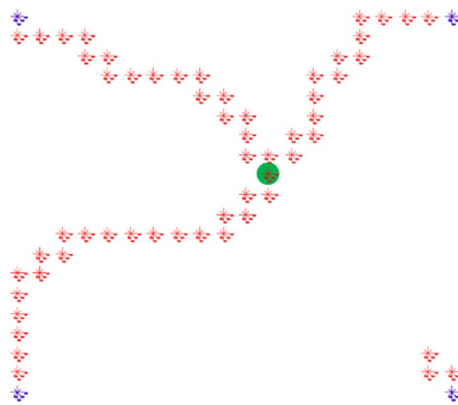


**Fig. 9** Diagram of the initial hill-climbing node

sub-nodes are obtained by expanding the hill-climbing node as the center in four directions: up, down, left, and right. Each hill-climbing node and sub-node corresponds to a real-time sub-image. The hill-climbing method starts from the initial node and combines the 4 adjacent sub-nodes to calculate the similarity of the corresponding 5 real-time sub-images and the template image. The centroid of the real-time subgraph with the highest similarity is used as the next hill-climbing node. With this centroid used as the center, the 4 sub-nodes are re-expanded. This is iteratively calculated until the target matching point is searched. If the target matching point cannot be found, the search is restarted from an intermediate state and proceeds along a suboptimal branching path. Figure 10 shows a schematic diagram of the hill-climbing process.

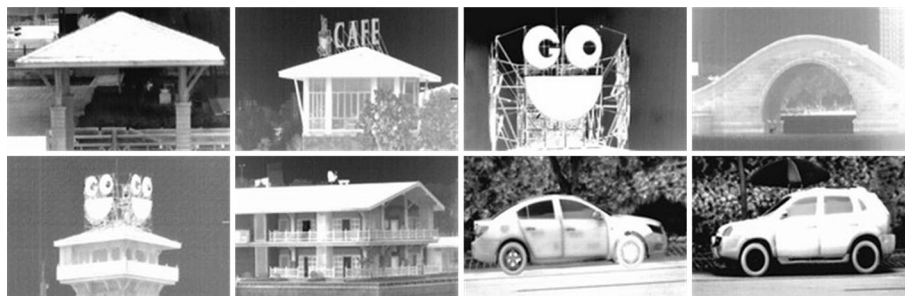
### 3 Results and discussion

To conduct a more comprehensive and objective performance test of the matching algorithm based on the gradient direction DE, the experiments were conducted using eight sets of IR and visible images taken in the field. Moreover, the test images are shown in Figs. 11 and 12, where Fig. 11 shows the IR template image with size  $108 \times 168$  and Fig. 12 shows the visible real-time image with size  $256 \times 256$ . The coordinates of the theoretical matching centroids are shown in Table 1. First, the matching

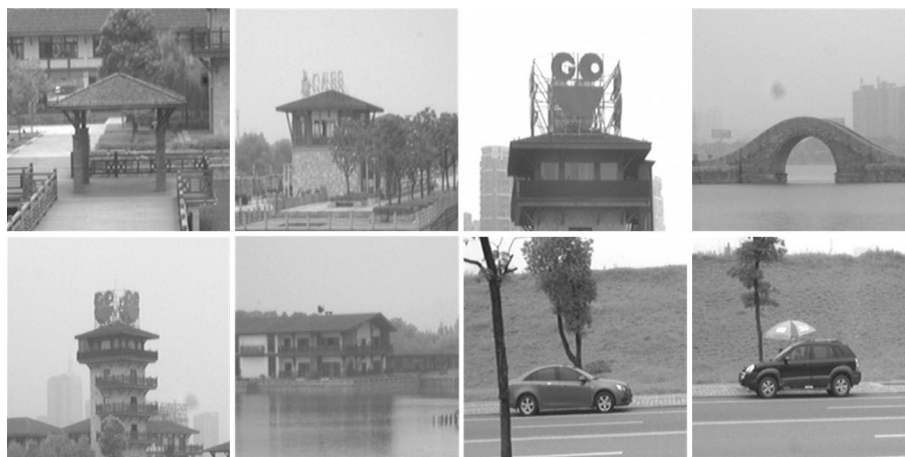


**Fig. 10** Diagram of the hill-climbing process

of this algorithm is tested for the case of the translation transformation only to verify the correctness of the theory of this algorithm; then, to solve the matching problem of rotation transformation, the relationship between the main direction map of the distribution field and the rotation angle is analyzed. Finally, the matching robustness in the case of the random angle transformation is experimentally verified, and the advantages and disadvantages of this algorithm are analyzed by comparing the matching algorithms based on mutual information.



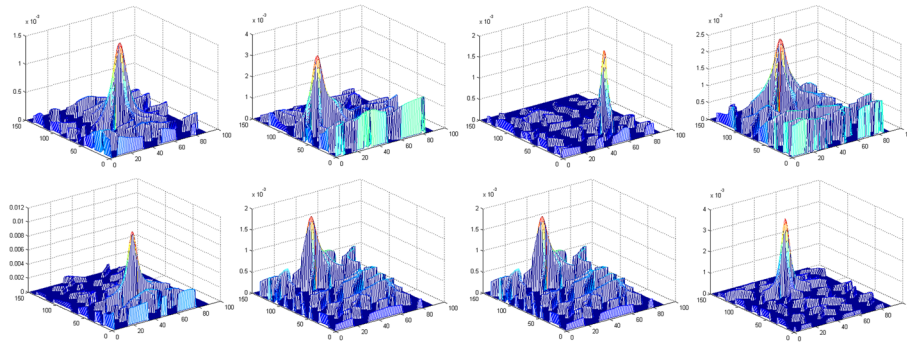
**Fig. 11.** 108 × 168 pixel infrared template images



**Fig. 12.** 256 × 256 pixel real-time visible images

**Table 1** Coordinates of the theoretical matching center point

Group number	1	2	3	4	5	6	7	8
Center point	(126, 129)	(120, 100)	(92, 141)	(159, 126)	(104, 129)	(124, 112)	(172, 122)	(150, 125)



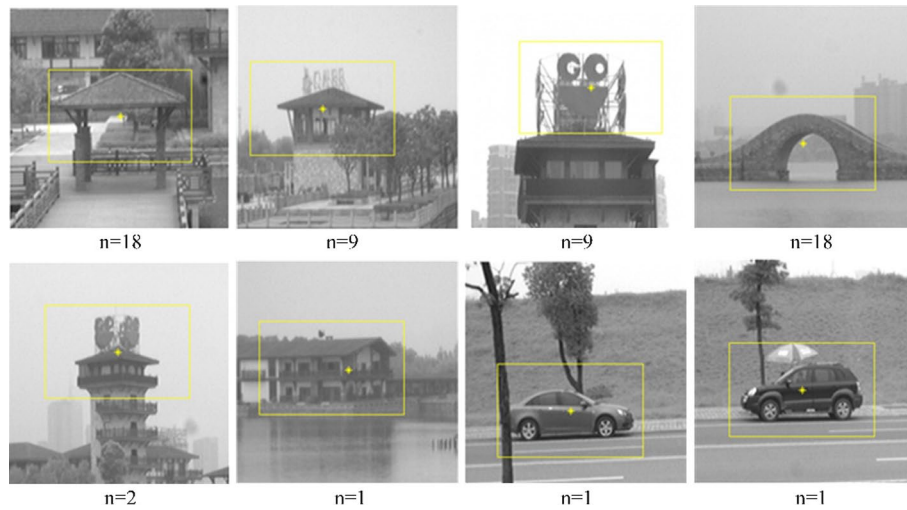
**Fig. 13** Matching correlation surfaces under translational change

### 3.1 Translational transformation matching effect

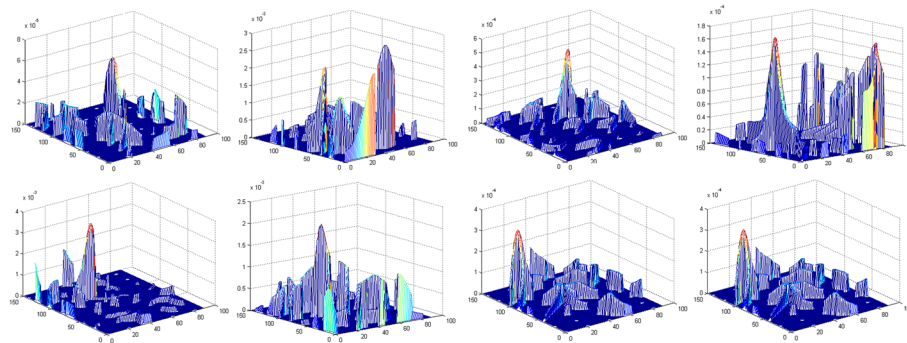
The effectiveness of this algorithm is verified by first testing the matching in the presence of translational transformations only. The matched correlation surfaces are shown in Fig. 13, where the correlation surface plots are shown from left to right and from top to bottom for the groups of images numbered 1–8. It is obvious from the figure that the highest peak of the surface is very prominent, and the peak corresponds to a unique coordinate position. Therefore, it can be concluded from the correlation surface plots that the algorithm in this paper is robust and adaptable and can complete the correct matching between heterogeneous source images. The main direction of the distribution field of the infrared template image is calculated during the experiment and expressed by  $n$ . The matching results are shown in Fig. 14.

### 3.2 Rotation transformation matching effect

Since the distribution field itself is not rotationally invariant and there is usually a certain angular difference between the heterogeneous image sets, solving the rotation transformation is the most critical and challenging challenge in the field of heterogeneous image matching. In this paper, the experiments simulate the actual rotational transformation by rotating the visible real-time images, where each real-time image is rotated randomly by a certain angle  $\theta$ , and  $\theta \in (0, 360)$ . In the matching process, the difference between the main direction of the real-time sub-image and the main direction  $n$  of the template image can be calculated as the index of the lookup table, which is equivalent to this rotational transformation of the template image. Then, the similarity measure is calculated to match the correlation surfaces, as shown in Fig. 15. The results are shown in Fig. 16, where the groups 1–8 are rotated by degrees of 70, 163, -138, -88, 155, 18, -43, and 92, respectively.



**Fig. 14** Matching results under translation transformation

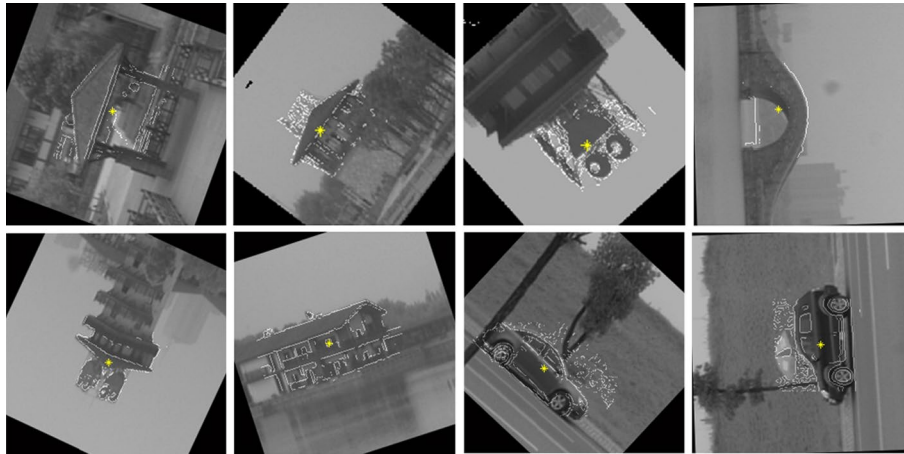


**Fig. 15** Matching correlation surfaces under rotational transformations

### 3.3 Experimental results and comparison

To test the proposed algorithm more comprehensively, the matching method based on the Bayesian mutual information (BayesMI) method, normalized cross correlation (NCC) method, sum of absolute differences (SAD) method, and sum of absolute transformed difference (SATD) method are selected for the experimental comparison. The test images are the 8 sets of heterogeneous images that are shown in Figs. 11 and 12. The visible images in each group are randomly rotated 10 times and then matched with the template image. Finally, the matching success rate, matching error and average elapsed time are counted.

From the comparison in Table 2, it can be seen that the proposed algorithm has a higher success rate, smaller average error and less time consumption for the matching problem in the presence of rotational transformations of heterogeneous source images. The BayesMI method has a higher matching success rate but has a poor real-time performance. The SAD method runs the fastest but has a low matching success rate. The NCC method balances the matching success rate, matching error and real time. However, the advantages are not significant. The main reason for the error of the proposed algorithm is that, in the rotation matching process, the rotation angle of the template image, which



**Fig. 16** Matching results under rotational changes

**Table 2** Statistics of the experimental results

Methods	Group	1	2	3	4	5	6	7	8	Average value
Our method	Success rate/%	100	100	90	80	100	90	90	100	94.00
	Average error	2.8	2.6	2.2	3.4	2.6	2.5	3.0	2.3	2.70
	Time spent/s	2.75	2.82	2.76	2.68	2.72	2.92	2.66	2.74	2.76
BayesMI method	Success rate/%	95	0	95	50	30	90	80	90	66.25
	Average error	6.6	11.3	4.3	18.0	5.5	8.4	8.6	9.7	9.05
	Time spent/s	36.26	30.68	44.90	14.35	35.11	32.76	52.49	15.55	32.76
NCC method	Success rate/%	80	0	90	50	70	70	60	30	56.25
	Average error	7.8	15.6	3.2	4.8	4.0	6.8	10.8	16.7	8.70
	Time spent/s	9.13	8.31	12.17	3.94	10.01	8.25	12.18	3.86	9.73
SAD method	Success rate/%	0	0	0	100	0	70	0	90	32.50
	Average error	14.5	7.9	9.4	1.2	13.9	10.4	13.5	8.6	9.90
	Time spent/s	1.47	1.48	1.46	1.47	1.46	1.46	1.51	1.46	1.47
SATD method	Success rate/%	80	90	30	10	100	90	90	60	68.75
	Average error	4.9	1.7	5.4	5.4	4.0	2.7	5.4	8.2	4.70
	Time spent/s	2.84	2.84	2.86	2.86	2.86	2.85	2.85	2.86	2.85

is calculated based on the difference of the main direction, is somewhat different from the rotation angle of the real-time image. For example, the real-time image is rotated by  $48^\circ$ , while the template image is rotated by  $50^\circ$  in the matching process. Then, the similarity metric is performed with the real-time sub-image. Compared with the proposed algorithm, various other algorithms can only solve the heterogeneous matching problem in the case of horizontal displacement. Moreover, the matching results are poor for the case of rotational stretching, while the proposed algorithm can solve the matching problem well in the case of rotation.

#### 4 Conclusion

In this paper, we propose a novel heterogeneous scene matching method based on the gradient direction distribution field. By constructing the gradient direction distribution field to redescribe the heterogeneous images and defining the main direction of the



distribution field, the matching problem between heterogeneous images with rotation transformations is solved. The similarity measure of the chi-square distance combined with the hill-climbing method search strategy improves the matching speed. Compared with the state-of-the-art region-based matching methods, the experimental results show that the proposed matching method has better robustness, accuracy and real-time performance.

#### Abbreviations

DF	Distribution field
SIFT	Scale invariant feature transform
SURF	Speeded-up robust features
ORB	Oriented FAST and rotated BRIEF
BRISK	Binary robust invariant keypoints
ANN	Artificial neural network
IR	Infrared
BayesMI	Bayes mutual information
NCC	Normalized cross correlation
SAD	Sum of absolute differences
SATD	Sum of absolute transformed difference

#### Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

#### Author contributions

Conceptualization: QL and RL; methodology: RL and QL; software: RL and SW; investigation: TS and WX; resources: ZW; writing—original draft preparation: RL and QL; writing—review and editing: XY, SW and TS; visualization: SW; supervision: RL and ZW; project administration: XY; funding acquisition: RL. All authors have agreed to the published version of the manuscript. All authors read and approved the final manuscript.

#### Funding

This work was supported by the National Natural Science Foundation of China (No. 61806209), Natural Science Foundation of Shaanxi Province (No. 2020JQ-490), Chinese Aeronautical Establishment (No. 201851U8012), and Natural Science Foundation of Shaanxi Province (No. 2020JM-537).

#### Availability of data and materials

The data sets used and analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Competing interests

The authors declare that they have no competing interests.

Received: 13 May 2022 Accepted: 30 March 2023

Published online: 27 April 2023

#### References

1. C. Yan, L. Meng, L. Li, J. Zhang, Z. Wang, J. Yin, J. Zhang, Y. Sun, B. Zheng, Age-invariant face recognition by multi-feature fusion and decomposition with self-attention. *ACM Trans. Multimed. Comput. Commun. Appl.* **18**(1s), 1–18 (2022)
2. C. Yan, Y. Hao, L. Li, J. Yin, A. Liu, Z. Mao, Z. Chen, X. Gao, Task-adaptive attention for image captioning. *IEEE Trans. Circuits Syst. Video Technol.* **32**(1), 43–51 (2021)
3. C. Yan, T. Teng, Y. Liu, Y. Zhang, H. Wang, X. Ji, Precise no-reference image quality evaluation based on distortion identification. *ACM Trans. Multimed. Comput. Commun. Appl.* **17**(3s), 1–21 (2021)
4. C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(4), 1445–1451 (2020)
5. C. Yan, L. Li, C. Zhang, B. Liu, Y. Zhang, Q. Dai, Cross-modality bridging and knowledge transferring for image understanding. *IEEE Trans. Multimed.* **21**(10), 2675–2685 (2019)
6. J. Xi, L. Wang, J. Zheng, X. Yang, Energy-constraint formation for multiagent systems with switching interaction topologies. *IEEE Trans. Circuits Syst. I Regul. Pap.* **67**(7), 2442–2454 (2020)
7. L. Junlong, X. Jianxiang, H. Ming, L. Bing, Formation control for networked multiagent systems with a minimum energy constraint. *Chin. J. Aeronaut.* **36**(1), 342–355 (2022)



8. R. Lu, X. Yang, W. Li, J. Fan, D. Li, X. Jing, Robust infrared small target detection via multidirectional derivative-based weighted contrast measure. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2020)
9. R. Lu, X. Yang, X. Jing, L. Chen, J. Fan, W. Li, D. Li, Infrared small target detection based on local hypergraph dissimilarity measure. *IEEE Geosci. Remote Sens. Lett.* **19**, 1–5 (2020)
10. X. He, Research on fast grayscale image matching algorithm, Doctoral dissertation, MA dissertation, Hefei University of Technology, 2012
11. Z. Shuang, J. Gang, Q. Yu-ping, Gray imaging extended target tracking histogram matching correction method. *Procedia Eng.* **15**, 2255–2259 (2011)
12. Z. Song, Research on image alignment techniques and their applications, Doctoral dissertation, Doctoral dissertation, Fudan University, 2010
13. B. Cui, J.-C. Créput, NCC based correspondence problem for first-and second-order graph matching. *Sensors* **20**(18), 5117 (2020)
14. X. Wan, J.G. Liu, S. Li, H. Yan, Phase correlation decomposition: the impact of illumination variation for robust subpixel remotely sensed image matching. *IEEE Trans. Geosci. Remote Sens.* **57**(9), 6710–6725 (2019)
15. L. Yong, Research on image mosaic algorithm based on mutual information, 2016
16. O. Angah, A.Y. Chen, Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy. *Autom. Constr.* **119**, 103308 (2020)
17. W. Shi, F. Su, R. Wang, J. Fan, A visual circle based image registration algorithm for optical and SAR imagery, in 2012 IEEE international geoscience and remote sensing symposium, IEEE, p. 2109–2112, 2012
18. Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors, in Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004, vol. 2, IEEE, 2004
19. J. López, R. Santos, X.R. Fdez-Vidal, X.M. Pardo, Two-view line matching algorithm based on context and appearance in low-textured images. *Pattern Recogn.* **48**(7), 2164–2184 (2015)
20. L. Zhang, R. Koch, An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency. *J. Vis. Commun. Image Represent.* **24**(7), 794–805 (2013)
21. S. Suri, P. Reinartz, Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Trans. Geosci. Remote Sens.* **48**(2), 939–949 (2009)
22. M. Hasan, M.R. Pickering, X. Jia, Robust automatic registration of multimodal satellite images using CCRE with partial volume interpolation. *IEEE Trans. Geosci. Remote Sens.* **50**(10), 4050–4061 (2012)
23. W. Qiu, X. Wang, X. Bai, A. Yuille, Z. Tu, Scale-space sift flow, in IEEE winter conference on applications of computer vision, IEEE, p. 1112–1119, 2014
24. J. Zhang, G. Chen, Z. Jia, An image stitching algorithm based on histogram matching and sift algorithm. *Int. J. Pattern Recognit. Artif. Intell.* **31**(04), 1754006 (2017)
25. H. Bay, T. Tuytelaars, L.V. Gool, *Surf: Speeded Up Robust Features* (Springer, New York, 2006), pp.404–417
26. E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to sift or surf, in 2011 international conference on computer vision, IEEE, p. 2564–2571, 2011
27. S. Leutenegger, M. Chli, R.Y. Siegwart, Brisk: binary robust invariant scalable keypoints, in 2011 international conference on computer vision, IEEE, p. 2548–2555, 2011
28. X. Shen, C. Wang, X. Li, Z. Yu, J. Li, C. Wen, M. Cheng, Z. He, Rf-net: an end-to-end image matching network based on receptive field, in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, p. 8132–8140, 2019
29. B.D. De Vos, F.F. Berendsen, M.A. Viergever, H. Sookooti, M. Staring, I. Išgum, A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **52**, 128–143 (2019)
30. G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca, Voxelmorph: a learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **38**(8), 1788–1800 (2019)
31. Y. Ono, E. Trulls, P. Fua, K.M. Yi, LF-NET: learning local features from images, in Advances in neural information processing systems, vol. 31, 2018
32. V. Balntas, E. Johns, L. Tang, K. Mikolajczyk, Pn-Net: conjoined triple deep network for learning local image descriptors (2016), arXiv preprint [arXiv:1601.05030](https://arxiv.org/abs/1601.05030)
33. X. Han, T. Leung, Y. Jia, R. Sukthankar, A.C. Berg, Matchnet: unifying feature and metric learning for patch-based matching, in Proceedings of the IEEE conference on computer vision and pattern recognition, p. 3279–3286, 2015
34. Y. Yuan, Research on feature point matching method based on bp neural network, Doctoral dissertation, MA dissertation, Xi'an University of Science and Technology, 2013
35. G. Dongyuan, Y. Xifan, Z. Qing, Development of machine vision system based on BP neural network self-learning, in 2008 international conference on computer science and information technology, IEEE, p. 632–636, 2008
36. Ł. Laskowski, J. Jelonekiewicz, Y. Hayashi, Extensions of hopfield neural networks for solving of stereo-matching problem, in Artificial intelligence and soft computing: 14th international conference, ICAISC 2015, Zakopane, Poland, June 14–18, 2015, Proceedings, Part I 14, Springer, p. 59–71, 2015
37. W. Mahdi, S.A. Medjahed, M. Ouali, Performance analysis of simulated annealing cooling schedules in the context of dense image matching. *Computación y Sistemas* **21**(3), 493–501 (2017)
38. Z. Wang, H. Pen, T. Yang, Q. Wang, Structure-priority image restoration through genetic algorithm optimization. *IEEE Access* **8**, 90698–90708 (2020)
39. C. Ostertag, M. Beurton-Aimar, Matching ostraca fragments using a siamese neural network. *Pattern Recogn. Lett.* **131**, 336–340 (2020)
40. Y. Liu, X. Gong, J. Chen, S. Chen, Y. Yang, Rotation-invariant siamese network for low-altitude remote-sensing image registration. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* **13**, 5746–5758 (2020)
41. W. Liu, C. Wang, X. Bian, S. Chen, S. Yu, X. Lin, S.-H. Lai, D. Weng, J. Li, Learning to match ground camera image and uav 3-d model-rendered image based on siamese network with attention mechanism. *IEEE Geosci. Remote Sens. Lett.* **17**(9), 1608–1612 (2019)

42. L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H. Torr, Fully-convolutional siamese networks for object tracking, in *Computer vision—ECCV 2016 workshops: Amsterdam, The Netherlands, October 8–10 and 15–16, 2016, proceedings, Part II 14*, Springer, p. 850–865, 2016
43. L. Sevilla-Lara, E. Learned-Miller, Distribution fields for tracking, in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, p. 1910–1917, 2012
44. C. Yan, Z. Li, Y. Zhang, Y. Liu, X. Ji, Y. Zhang, Depth image denoising using nuclear norm and learning graph model. *ACM Trans. Multimed. Comput. Commun. Appl.* **16**(4), 1–17 (2020)
45. X. Gao, F. Sattar, R. Venkateswarlu, Multiscale corner detection of gray level images based on log-gabor wavelet transform. *IEEE Trans. Circuits Syst. Video Technol.* **17**(7), 868–875 (2007)

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---