

Heterogeneous scene matching based on gradient direction distribution field

Qingge Li

PLA Rocket Force University of Engineering <https://orcid.org/0000-0002-7476-6389>

Ruitao Lu (✉ lrt19880220@163.com)

<https://orcid.org/0000-0002-7527-4298>

Xiaogang Yang

PLA Rocket Force University of Engineering

Siyu Wang

PLA Rocket Force University of Engineering

Tong Shen

PLA Rocket Force University of Engineering

Wenxin Xia

PLA Rocket Force University of Engineering

Zhaoying Wei

Xi'an Shiyou University

Research

Keywords: Heterogenous images, scene matching, distribution field, hill climbing method

Posted Date: May 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1651222/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at EURASIP Journal on Image and Video Processing on April 27th, 2023. See the published version at <https://doi.org/10.1186/s13640-023-00608-x>.

Abstract

Heterogenous scene matching is one of the key technologies in the field of computer vision. The image rotation problem has been a hot and difficult problem in the field of heterogenous scene matching. In this paper, a heterogenous scene matching method based on the gradient direction distribution field is proposed. The Distributed field theory is introduced into the heterogenous scene matching for the first time. First, the gradient direction distribution field is constructed and fuzzified, and then the effective regions are selected. Then, the concept of main direction distribution field 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. At last, the hill-climbing method search strategy is adopted, which greatly improves the efficiency of the algorithm. Experimental results on 8 pairs of infrared and visible heterogenous images demonstrate that the proposed method outperforms the other state-of-the-art region-based matching methods in terms of robustness, accuracy and real-time performance.

1 Introduction

With the rapid progress of informationization in the world, the demand for image information has become increasingly strong. In recent years, image matching detection technology has become a research hotspot in the field of computer vision and is widely applied to various fields such as navigation guidance, medical diagnosis and image retrieval. Under different task conditions (climate, light intensity, shooting position and angle, etc.), image information often has to be acquired through different sensors, and these images generally have differences in gray value, resolution, scale or different nonlinear distortion, and it is a difficult research task to achieve accurate matching of heterogeneous source images in complex environments.

Over the years, various heterogenous image matching methods have been proposed. In general, it seems that heterogenous 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 of an image as the basis of feature space and similarity metric, and use the similarity metric algorithm to find out the correspondence between the region and the image to be matched, and find the best matching position globally. The commonly used region-based matching algorithms are grayscale correlation method [1–3], maximum mutual information correlation method [4–6], and gradient correlation method [7]. 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 points [8, 9], lines [10–12], and faces [13] of an image to achieve matching. Local invariant descriptor-based matching methods [14], common descriptors are Scale Invariant Feature Transform (SIFT) [15], Speeded-Up Robust Features (SURF) [16], and Oriented FAST and Rotated BRIEF (ORB) [17], where the SURF algorithm is influenced by SIFT, which greatly increases the matching speed. The ORB algorithm extracts feature points fairly quickly, but it is more sensitive than other algorithms when there are large rotations and scale changes between images. In the same year, [18] 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 large blur. The feature-based method has good adaptability to the geometric deformation, brightness variation and noise effects of images, and also 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 [19–24] have been developed rapidly. Representative methods include BP neural network-based image matching methods [25, 26], Hopfield network-based image matching methods [27], annealing algorithm-based image matching methods [28], genetic algorithm-based image matching methods [29], and twin network-based matching methods [30–33]. 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 required. Then, according to the requirements of some constructed neural network, some initial state information parameters needed by the network are selected and input, and also 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, and the learning process is relatively long, and it may fall into local minima.

In the literature [34], 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, but the common blurring techniques currently used have many 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 match successfully even if there are small distortions and rotations in the real-time image. Inspired by this, this paper applies the distribution field theory to the heterogenous image matching problem, and successfully achieves heterogenous image matching by constructing a distribution field in the gradient direction to describe the heterogenous image.

Based on the above considerations, this paper proposes a heterogenous image matching method based on gradient direction distribution field, which introduces the direction distribution field into the heterogenous 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 heterogenous image matching is well solved. We have conducted a series of experiments on infrared (IR) and visible heterogenous images, and the results show that our method has good performance in terms of robustness and accuracy of detection.

The rest of the paper is organized as follows. Section 2 presents the framework of our method and the formulation of the proposed method. We conduct a series of experiments on infrared and visible heterogenous images and three prior methods are compared to our approach in section 3. The study's conclusions are presented in section 4.

2 Methods

In this paper, the distributed field theory is applied to the study of heterogeneous source image matching problem, focusing on the matching difficulties caused by rotational transformation. First, the gradient direction DF is constructed and fuzzy filtered. Secondly, the feature space of the template image is reordered according to the location difference of the main direction DF of the template image and the real-time sub graph. Finally, the similarity is calculated and stored in the correlation matrix. The flow of the proposed algorithm is shown in Fig. 1.

2.1 Calculation of distribution field diagram

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. Take a grayscale image as an example, the grayscale level is 0 ~ 256, and the 256 grayscale levels can be divided into N intervals, and the pixel points corresponding to each interval grayscale contains not only grayscale information but also location information.

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

The distribution field schematic can be represented graphically 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:

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

1

where $I(i, j) = k$ is the gray value of the pixel with coordinate (i, j) in the image, and $d(i, j, k)$ is the value of the pixel with coordinate (i, j) in the image on the 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 layer sum to :

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

2

The target shown in Fig. 3 is used as an example to analyze its distribution field. The target size in the figure is 14×14, because the distribution field needs to be fuzzy for each layer, and also 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 gray 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 can be 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 re-describing the original image. Next, in order to make the location information not lose 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 feature layer:

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

3

where $d_s(k)$ denotes the new feature layer after the feature layer is convolved with the Gaussian filter; $d(k)$ is the feature layer before convolution; h_{σ_s} is a two-dimensional Gaussian filter with standard deviation σ_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.

Comparing with Fig. 4, it can be seen that before convolution, if the value of a position on the feature layer is 1, it indicates that the gray value at this position on the original image falls in the interval of N intervals; after convolution, if the value of a position on the feature layer is not 0, it indicates that the gray value at a position near this position on the original image falls in the interval of N intervals. This shows that Gaussian filtering of the feature layer is introducing uncertainty of 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, which will have some effect on the matching error during the matching process, but can enhance the robustness of the algorithm, making it possible to match successfully even in the presence of small rotational transformations.

In Eq. (3), if the Gaussian function h_{σ_s} is considered as 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:

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

4

in the above equation is a one-dimensional Gaussian filter with standard deviation σ_f . The final field distribution obtained from the example image shown in Fig. 3 is shown in Fig. 6.

At this point, the field distribution map of an image is calculated, and the process of 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: firstly, convolution in the direction of the two coordinate axes of the image introduces the uncertainty of position; secondly, convolution in the feature space introduces the uncertainty of grayscale information. In other words, the image represented by 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 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 by common first-order or second-order differential operators, such as Roberts operator, Sobel operator, Prewitt operator, etc. In this paper, the flat region with 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 π , and then the gradient direction is quantized to $[0, 180]$, which is expressed by the following equation:

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

5

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

6

Where, $\text{angle}(x)$ is the phase angle finding function of vector; $\text{sign}(\nabla I_y)$ is the gradient sign in vertical direction; i is the complex unit; $G_{x,y}$ is the gradient amplitude; τ is the gradient amplitude threshold, which is used to distinguish the low amplitude flat region from the effective gradient region, and 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 0 region at the edge, in order to prevent this region from affecting the matching accuracy and robustness, we only

construct a DF for the points with gradient direction greater than zero. Take $N = 18$, 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 $[1, 10]$. And so on, we can construct 18 layers of gradient direction DF, that is:

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

7

Finally, the DF feature space is filtered to introduce the ambiguity of position and the ambiguity of gray intensity into the distribution field map, which only loses the exact information and does not introduce the wrong position information into the DF, and still matches correctly in the case of smaller deformations, making the robustness enhanced.

For the traditional rectangular region, when it is rotated by a certain angle, the four vertex regions of the rectangular frame will lose more or less information in the original image according to the size of the rotation angle, and for a real-time subgraph, some irrelevant image information will be introduced while the original image information is lost. Therefore, the application of the traditional rectangular region to solve the rotation problem inevitably affects the accuracy and robustness of matching. In order to describe the image features more accurately, the rotation-invariant region needs to be selected as the effective region for computing the feature description.

Among all shapes, only a circle can satisfy the rotation invariant property, so the definition domain of feature description is set as a circle in this paper. As shown in Fig. 8, the center of the image is taken as the rotation axis, and the largest inwardly connected circle with the axis as the center of the circle is taken as the valid region.

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 which the images can be rotationally corrected and then search 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 class 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 . The

calculation process is as follows:

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

8

$$[mlaysum, n] = \max(dsum)$$

9

where $dsum_k$ is the probability statistics of the DF at the layer; $dsum$ is the matrix storing the probability sum of each DF layer, $mlaysum$ is the maximum value in $dsum$, and is the value of corresponding to the maximum $dsum$.

2.4 Similarity Metric

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

There are many methods used to measure the correlation of two feature vectors, such as 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. [35] introduces the Chi-square distance formula, which can achieve good results in measuring the similarity of two eigenvectors.

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

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, it can be seen that the Chi-square distance calculates the ratio of the variance of the corresponding elements to the sum of the elements, and the smaller the ratio indicates the closer the distance, the higher the similarity, while the Euclidean distance only considers the difference of the corresponding elements. During the experiment, it was found that the robustness of using Euclidean distance as the similarity discriminator could not meet the matching requirements and mismatching occurred, while Chi-square distance could better meet the requirements of the method in this paper.

The Chi-square distance is further processed so that it is inverted, i.e., its minimum value is expected to become a maximum value, allowing a better visualization of the best match when drawing the relevant surface plot.

$$dist = \exp(k\chi^2) \quad \theta \in (0, 360)$$

11

In order 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 adopts Chi-square distance as the similarity measure, and its larger value indicates the lower similarity of two images, so this paper uses Eq. (11) to invert the correlation surface map, which can more intuitively see the best matching point obtained by hill climbing search. Figure 9 gives a schematic diagram of the correlation surface of the hill-climbing method.

3 Results And Discussion

In order to conduct a more comprehensive and objective performance test of the matching algorithm based on the gradient direction DF, the experiments were conducted using eight sets of IR and visible images taken in the field, and the test images are shown in Fig. 10, numbered 1, 2, 3, ..., and 8 from left to right and from top to bottom, where Fig. 10(a) is the IR template image with size 108×168 and Fig. 10(b) is the visible real-time image with size 256×256 . the coordinates of the theoretical matching centroids are shown in Table 1. First, the matching of this algorithm is tested for the case of translation transformation only to verify the correctness of the theory of this algorithm; then, in order 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 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.

Table 1
Coordinates of the theoretical matching center point.

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

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. 11, where the correlation surface plots are shown from left to right and from top to bottom for 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. 12.

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 heterogenous image sets, solving the rotation transformation is the most critical and challenging challenge in the field of heterogenous 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 θ , and $\theta \in (0, 360)$. In the matching process, the difference between the main direction of the real-time subgraph 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, and then the similarity measure is calculated to match the relevant surfaces as in Fig. 13. The results are shown in Fig. 14, where groups 1–8 are rotated by degrees 70, 163, -138, -88, 155, 18, -43, and 92, respectively.

3.3 Experimental results and comparison

In order to test the proposed algorithm more comprehensively, the matching method based on Bayes 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 heterogenous images shown in Fig. 10. 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.

Table 2
Statistics of 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
	Average error	2.8	2.6	2.2	3.4	2.6	2.5	3.0	2.3	2.7
	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	8.7
	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	32.5
	Average error	14.5	7.9	9.4	1.2	13.9	10.4	13.5	8.6	9.9
	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	4.7
	Time spent /s	2.84	2.84	2.86	2.86	2.86	2.85	2.85	2.86	2.85

From the comparison in the table, it can be seen that the proposed algorithm has a higher success rate, smaller average error and less time consuming for the matching problem in the presence of rotational transformation of heterogeneous source images. The BayesMI method has a higher matching success rate but has 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, but the advantages are not outstanding. 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 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 degrees, while the template image is rotated by 50 degrees in the matching process and then the similarity metric is performed with the real-time subgraph. Compared with the proposed algorithm, various other algorithms can only solve the heterogeneous matching problem in the case of horizontal displacement, and the matching results are poor for the case of rotational stretching, while the proposed algorithm can well solve the matching problem in the case of rotation.

4 Conclusion

This paper proposes 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 transformation is solved. The similarity measure of 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, experimental results show that the proposed matching method has better robustness, accuracy and real-time performance.

Abbreviations

SIFT: Scale Invariant Feature Transform; SURF: Speeded-Up Robust Features; ORB: Oriented FAST and Rotated BRIEF; BRISK: Binary Robust Invariant Scalable Keypoints; ANN: artificial neural network; DF: distribution fields; IR: infrared; BayesMI: Bayes Mutual Information; NCC: Normalized Cross Correlation; SAD: Sum of Absolute Differences; SATD: Sum of Absolute Transformed Difference.

Declarations

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing financial interests.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grant 61806209, in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JQ-490, in part by the Chinese Aeronautical Establishment under Grant 201851U8012, and in part by the Natural Science Foundation of Shaanxi Province under Grant 2020JM-537.

Authors' Contributions

Conceptualization, Q.L. and R.L.; Methodology, R.L. and Q.L.; Software, R.L. and S.W.; Investigation, T.S. and W.X.; Resources, Z.W.; Writing-original draft preparation, R.L. and Q.L.; Writing-review and editing, X.Y., S.W. and T.S.; Visualization, S.W.; Supervision, R.L. and Z.W.; Project administration, X.Y.; Funding acquisition, R.L. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

The research 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).

References

1. X. He, "Research on fast grayscale image matching algorithm," MA dissertation, Hefei University of Technology, 2012.
2. S. Zhang, G. Jin, and Y.-p. Qin, "Gray Imaging Extended Target Tracking Histogram Matching Correction Method," *Procedia Engineering*, vol. 15, pp. 2255-2259, 2011.
3. Z. Song, "Research on image alignment techniques and their applications," Doctoral dissertation, Fudan University, 2010.
4. B. Cui and J.-C. Créput, "NCC Based Correspondence Problem for First- and Second-Order Graph Matching," *Sensors*, vol. 20, no. 18, p. 5117, 2020.
5. X. Wan, J. G. Liu, S. Li, and H. Yan, "Phase Correlation Decomposition: The Impact of Illumination Variation for Robust Subpixel Remotely Sensed Image Matching," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6710 - 6725, 2019.
6. Y. Li, "Research on image stitching algorithm based on mutual information," MA dissertation, Harbin Institute of Technology, 2016.
7. S. Y. Wen and A. Y. Chen, "Tracking multiple construction workers through deep learning and the gradient based method with re-matching based on multi-object tracking accuracy," *Automation in Construction*, vol. 119, p. 103308, 2020.
8. S. Wei, F. Su, R. Wang, and J. Fan, "A visual circle based image registration algorithm for optical and SAR imagery," in *2012 IEEE International Geoscience and Remote Sensing Symposium*, 2012, pp. 2109-2112.
9. N. Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.
10. J. López, R. Santos, X. R. Fdez-Vidal, and X. M. Pardo, "Two-view line matching algorithm based on context and appearance in low-textured images," *Pattern Recognition*, vol. 48, no. 7, pp. 2164-2184, 2015.
11. L. Zhang and R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 794-805, 2013.

12. S. Suri and P. Reinartz, "Mutual-Information-Based Registration of TerraSAR-X and Ikonos Imagery in Urban Areas," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 2, pp. 939-949, 2010. IEEE
13. M. Hasan, M. R. Pickering, and X. Jia, "Robust Automatic Registration of Multimodal Satellite Images Using CCRE With Partial Volume Interpolation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 4050-4061, 2012.
14. W. Qiu, X. Wang, B. Xiang, A. Yuille, and Z. Tu, "Scale-space SIFT flow," in *2014 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014, pp. 1112-1119.
15. J. Zhang, G. Chen, and Z. Jia, "An image stitching algorithm based on histogram matching and SIFT algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 04, p. 1754006, 2017.
16. H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *European Conference on Computer Vision*, 2006, vol. 3951, pp. 407-417.
17. E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision*, 2011, pp. 2564-2571.
18. S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *International Conference on Computer Vision*, 2011, pp. 2548-2555.
19. X. Shen, C. Wang, X. Li, Z. Yu, and Z. He, "RF-Net: An End-To-End Image Matching Network Based on Receptive Field," presented at the *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
20. B. de Vos, D., F. Berendsen, F., M. Viergever, A., H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis*, vol. 52, pp. 128-143, 2019.
21. G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788-1800, 2019. IEEE
22. Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "LF-Net: Learning local features from images," presented at the *Advances in Neural Information Processing Systems*, 2018.
23. V. Balntas, E. Johns, L. Tang, and K. Mikolajczyk, "PN-Net: Conjoined Triple Deep Network for Learning Local Image Descriptors," *arXiv preprint arXiv:1601.05030* 2016.
24. X. Han, T. Leung, Y. Jia, R. Sukthankar, and 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*, 2015, pp. 3279-3286.
25. Y. Yuan, "Research on feature point matching method based on BP neural network," MA dissertation, Xi'an University of Science and Technology, 2013.
26. D. Ge, X. Yao, and Q. Zhang, "Development of Machine Vision System Based on BP Neural Network Self-learning," in *International Conference on Computer Science and Information Technology*, 2008, pp. 632-636.

27. Ł. Laskowski, J. Jelonkiewicz, and Y. Hayashi, "Extensions of Hopfield Neural Networks for Solving of Stereo-Matching Problem," presented at the International Conference on Artificial Intelligence and Soft Computing, 2015.
28. W. Mahdi, S. A. Medjahed, and M. Ouali, "Performance Analysis of Simulated Annealing Cooling Schedules in the Context of Dense Image Matching," *Computacion y Sistemas*, vol. 21, no. 3, pp. 493-501, 2017.
29. Z. Wang, H. Pen, T. Yang, and Q. Wang, "Structure-Priority Image Restoration Through Genetic Algorithm Optimization," *IEEE Access*, vol. 8, pp. 90698 - 90708, 2020.
30. C. Ostertag and M. Beurton-Aimar, "Matching ostraca fragments using a siamese neural network," *Pattern Recognition Letters*, vol. 131, pp. 336-340, 2020.
31. Y. Liu, X. Gong, J. Chen, S. Chen, and Y. Yang, "Rotation-Invariant Siamese Network for Low-Altitude Remote-Sensing Image Registration," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5746-5758, 2020.
32. W. Liu, C. Wang, X. Bian, S. Chen, and S. Yu, "Learning to Match Ground Camera Image and UAV 3-D Model-Rendered Image Based on Siamese Network With Attention Mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1608 - 1612, 2019.
33. L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. Torr, "Fully-Convolutional Siamese Networks for Object Tracking," presented at the European Conference on Computer Vision, 2016.
34. L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1910-1917.
35. X. Gao, F. Sattar, and R. Venkateswarlu, "Multiscale Corner Detection of Gray Level Images Based on Log-Gabor Wavelet Transform," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 17, no. 7, pp. 868-875, 2007.

Figures

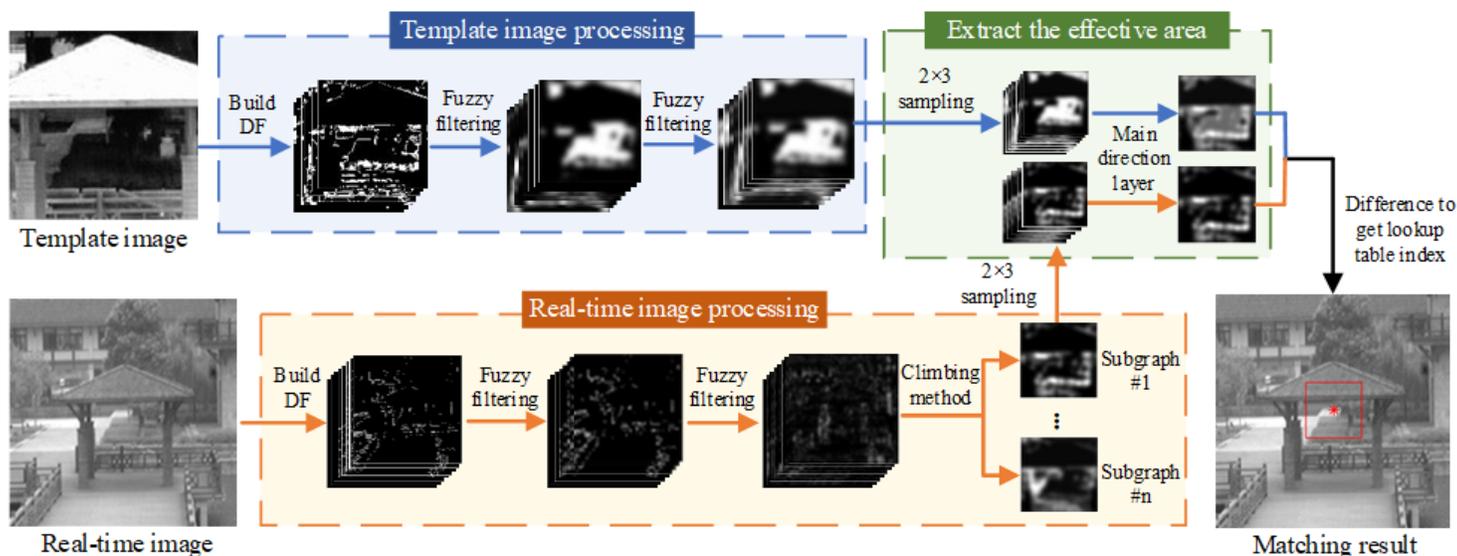


Figure 1

The flowchart of the proposed method.

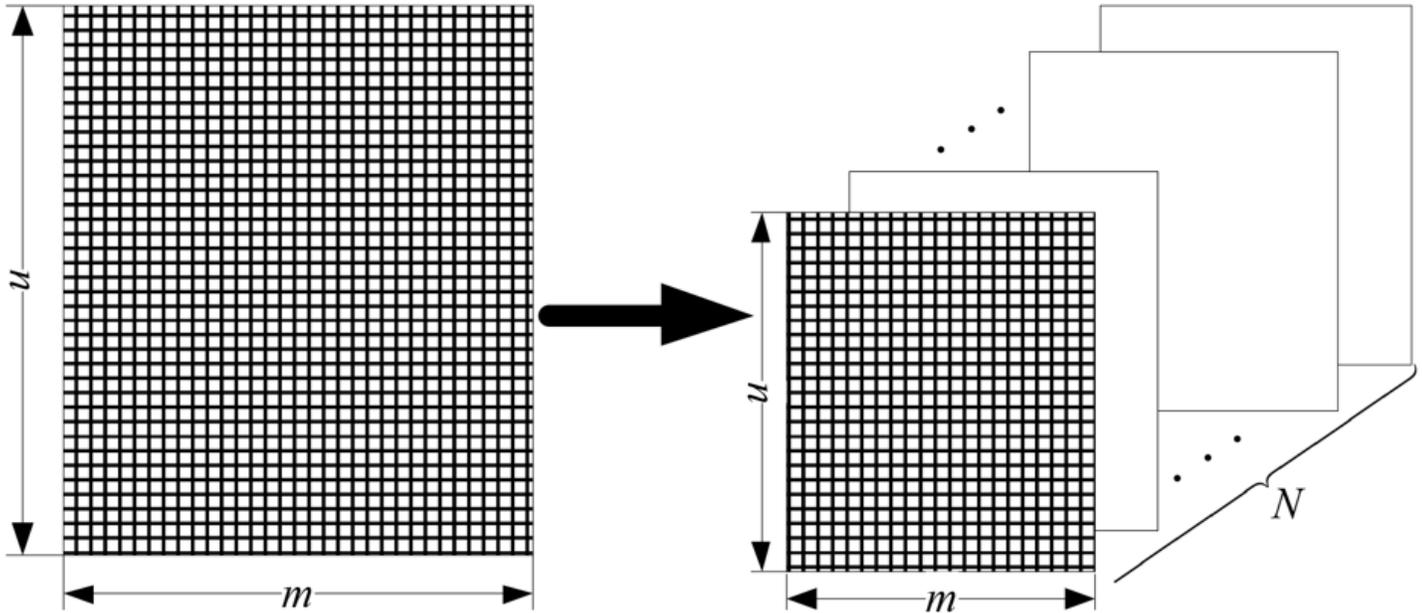


Figure 2

Schematic diagram of distribution field.



Figure 3

14×14 pixel image.

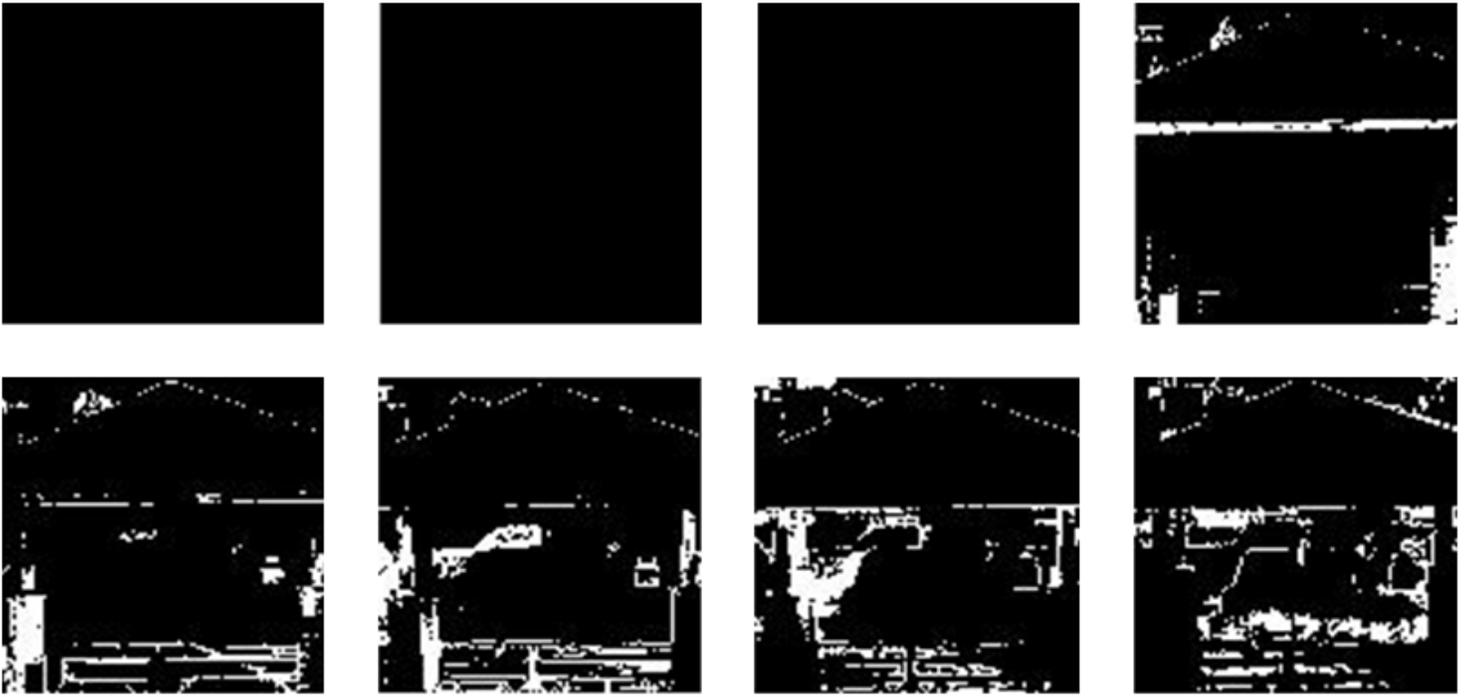


Figure 4

The 8 feature layers of the target.

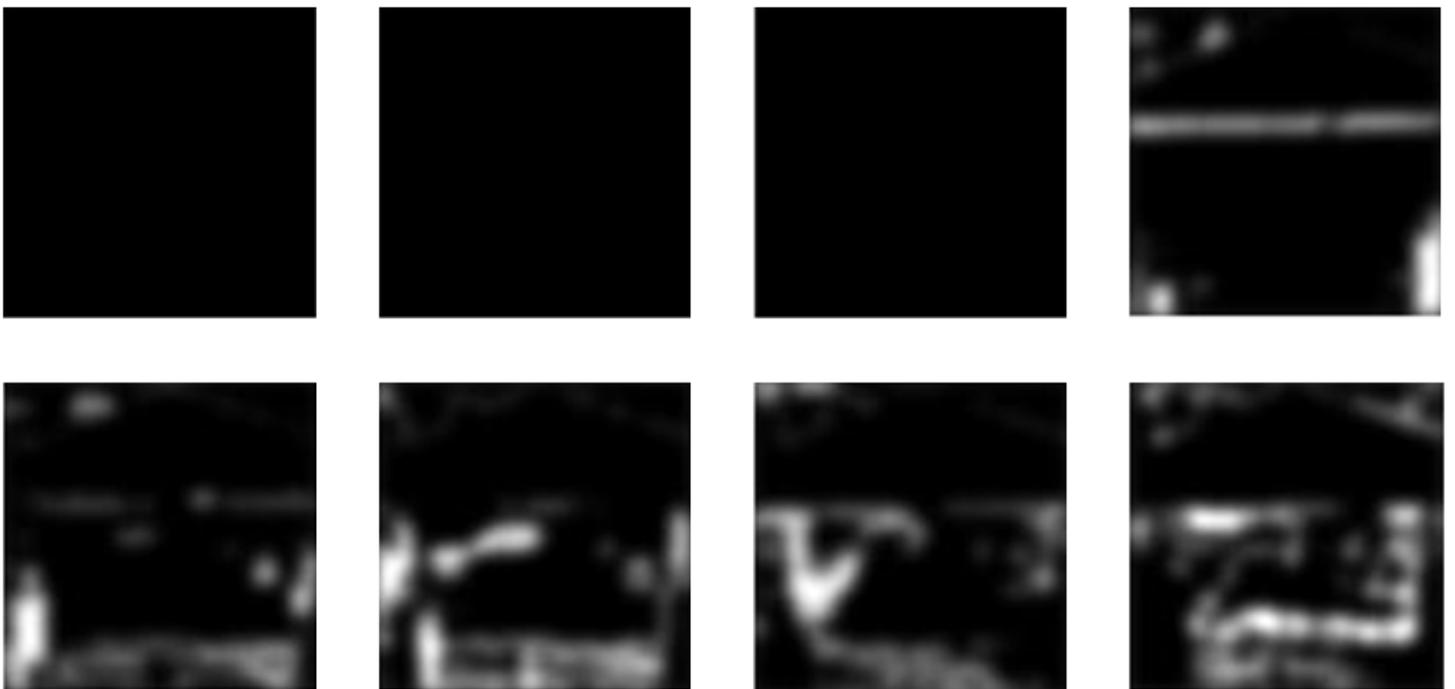


Figure 5

The feature layers of the image after the first convolution.

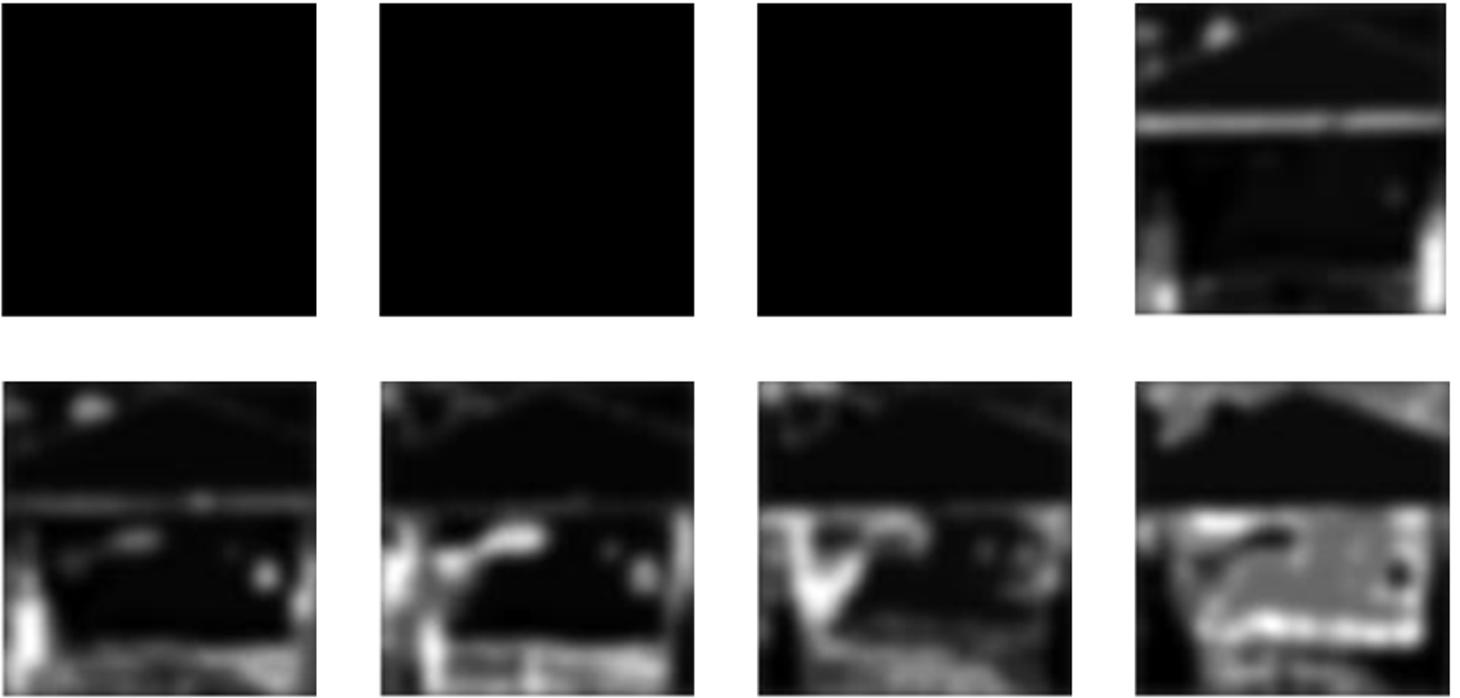


Figure 6

Each feature layer of the small target after the second convolution.

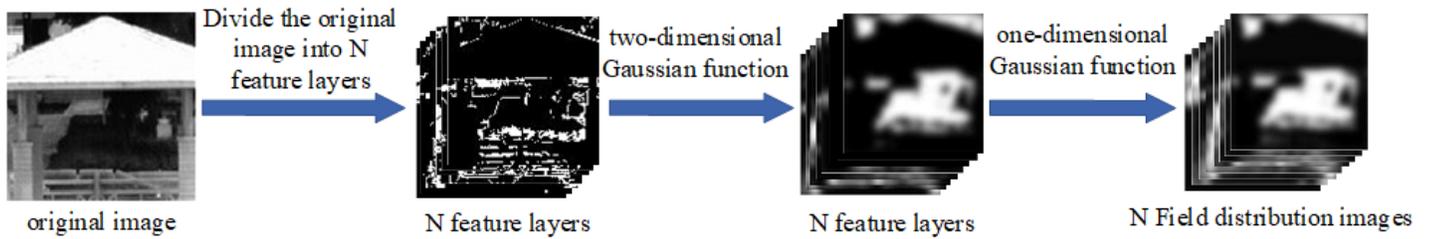


Figure 7

Field distribution images calculation process.

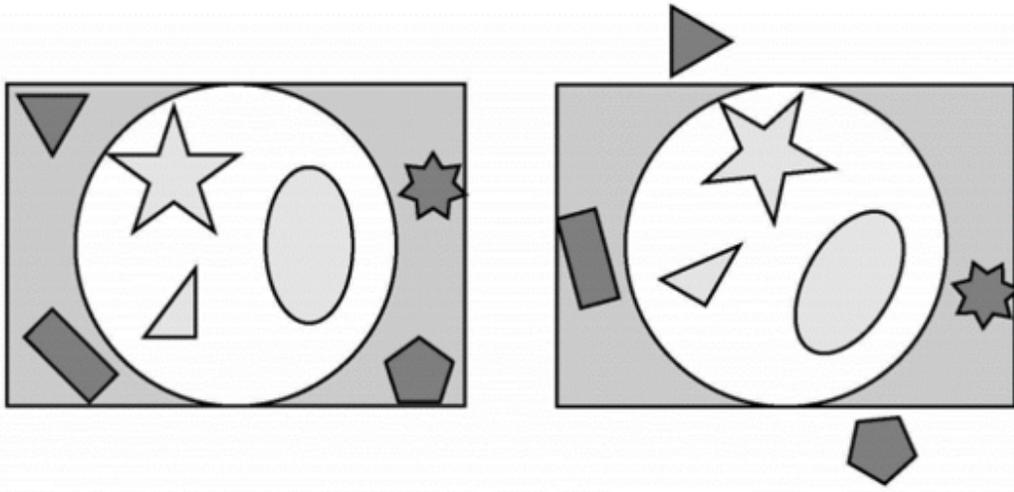


Figure 8

Schematic diagram of effective area selection.

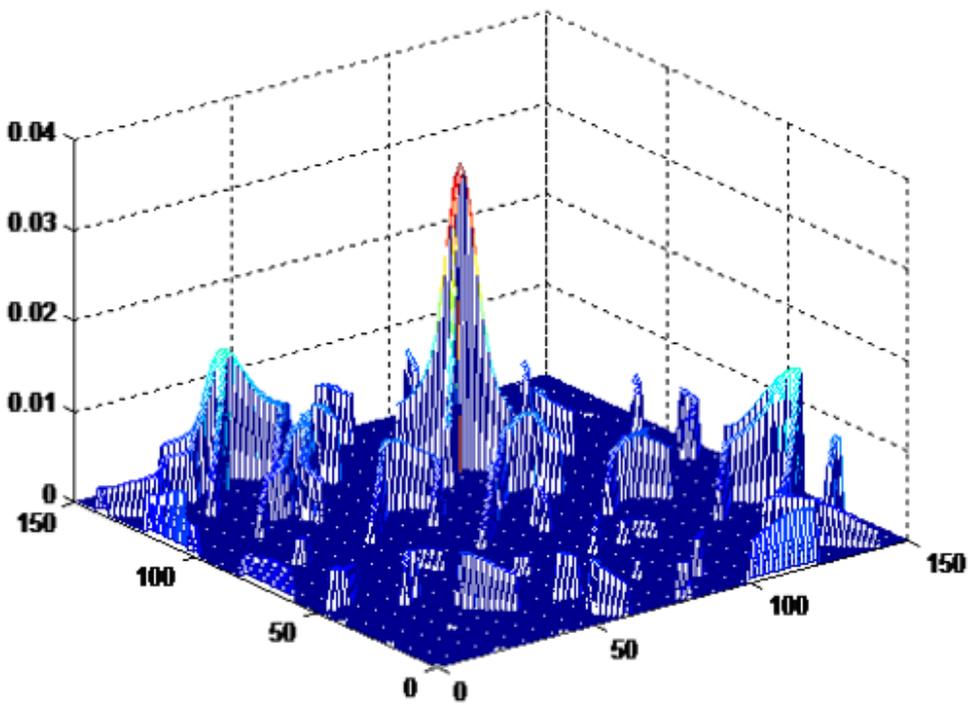
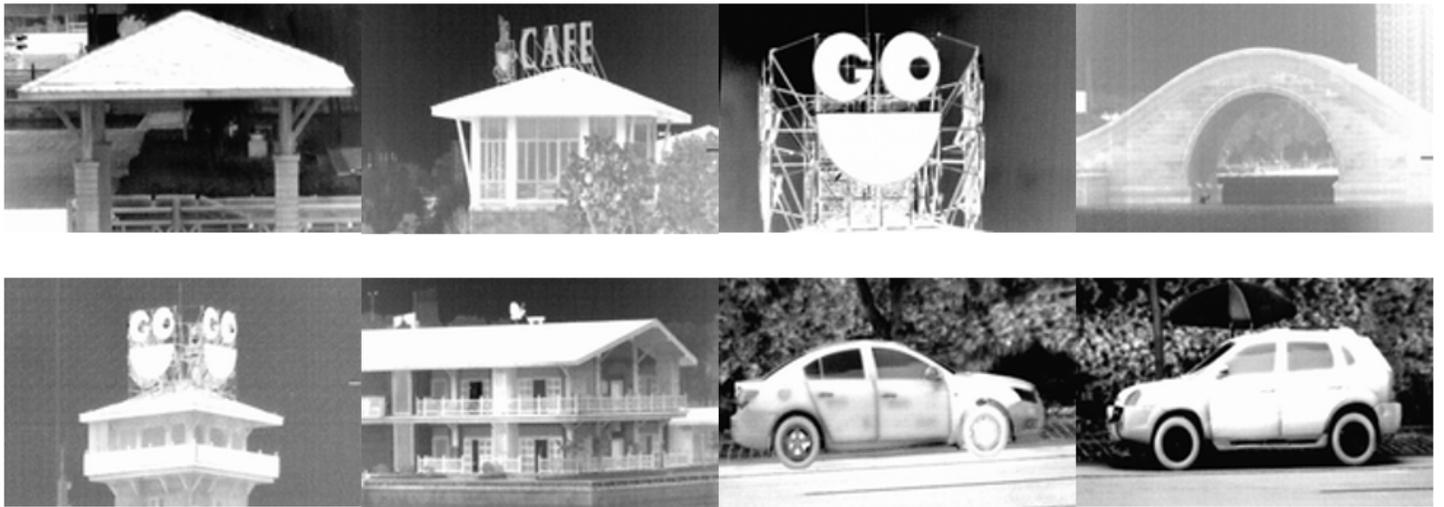


Figure 9

Example of surface related to climbing method.



(a) 108×168 pixel infrared template images



(b) 256×256 pixel real-time visible images

Figure 10

Infrared and visible light experimental images.

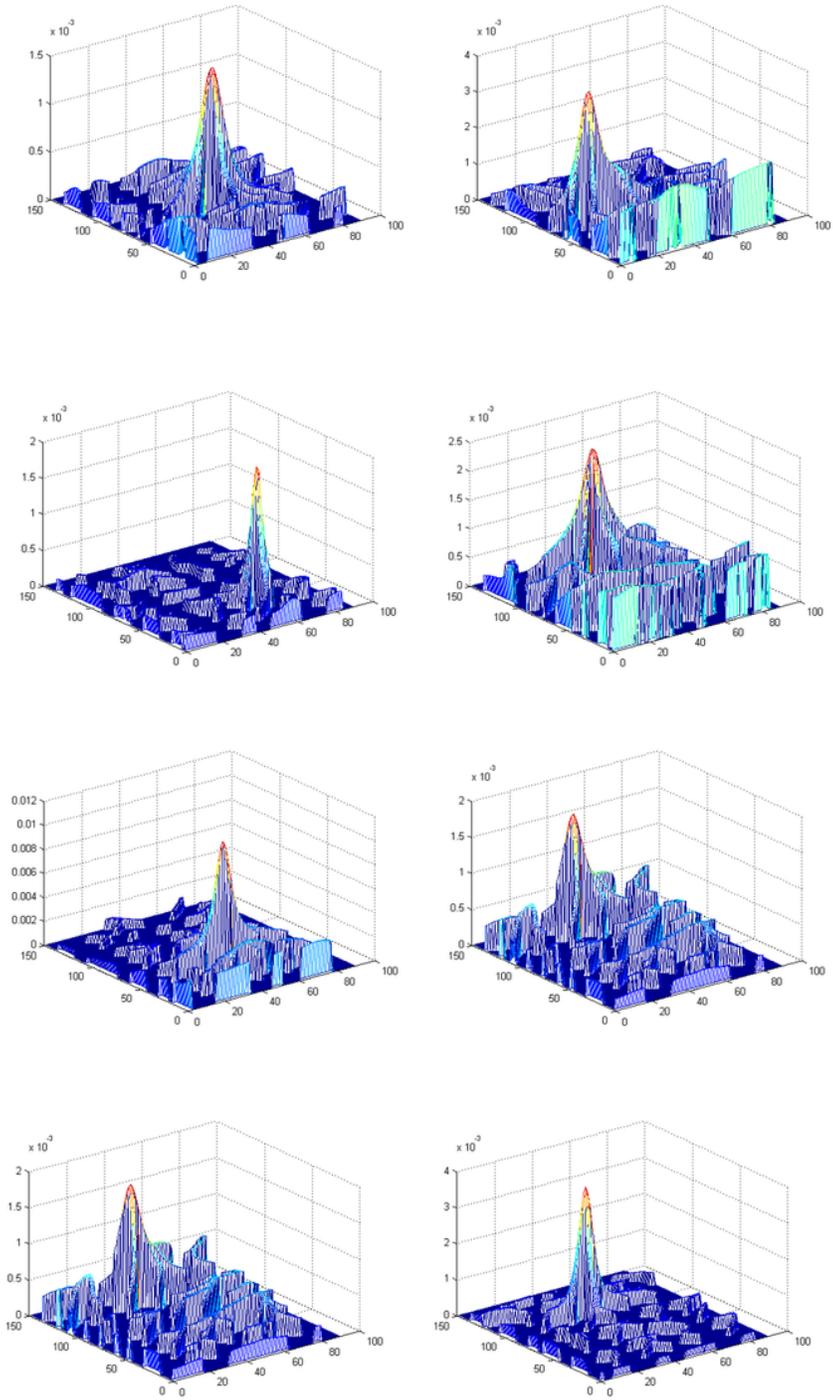


Figure 11

Matching correlation surfaces under translational change.

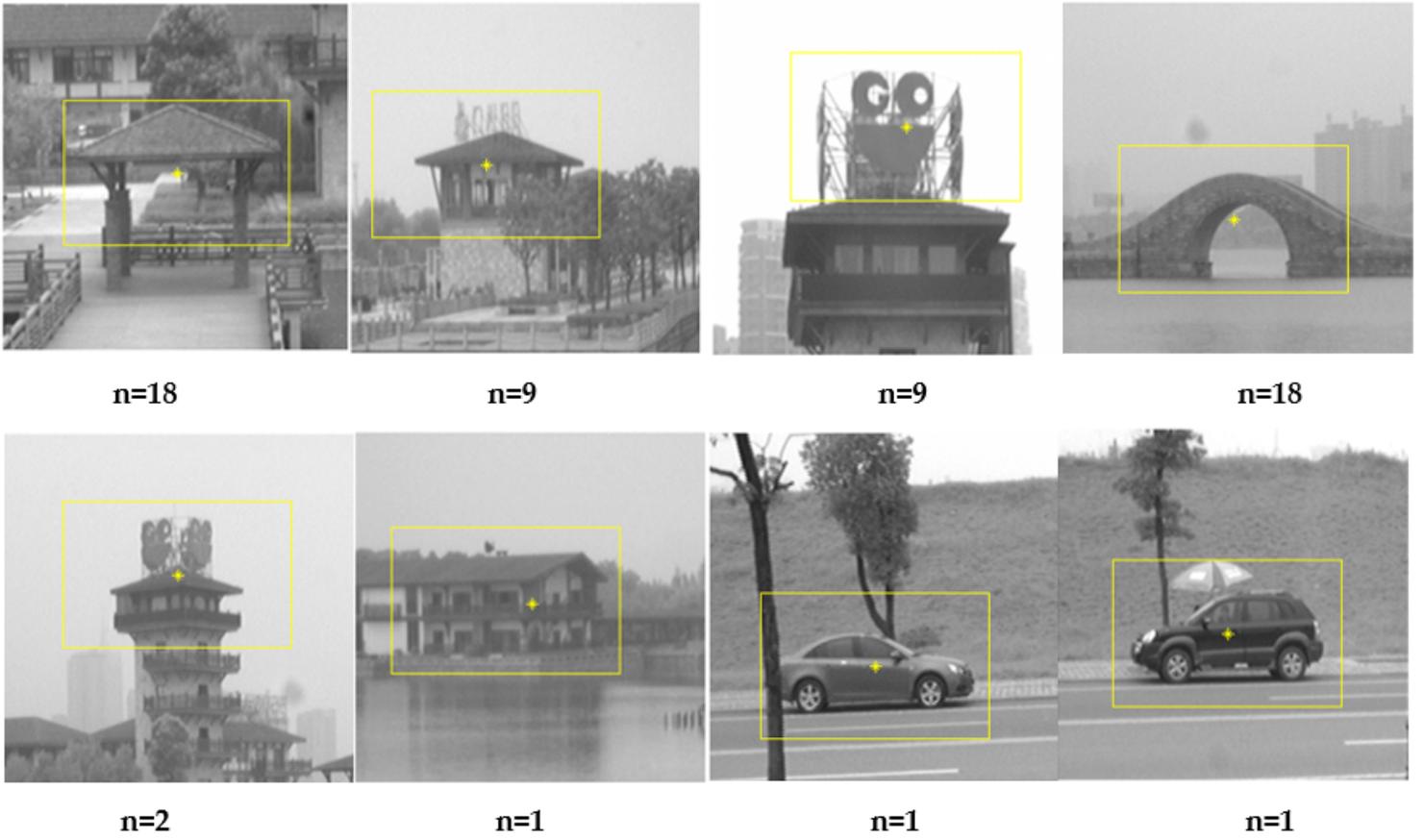


Figure 12

Matching results under translation transformation.

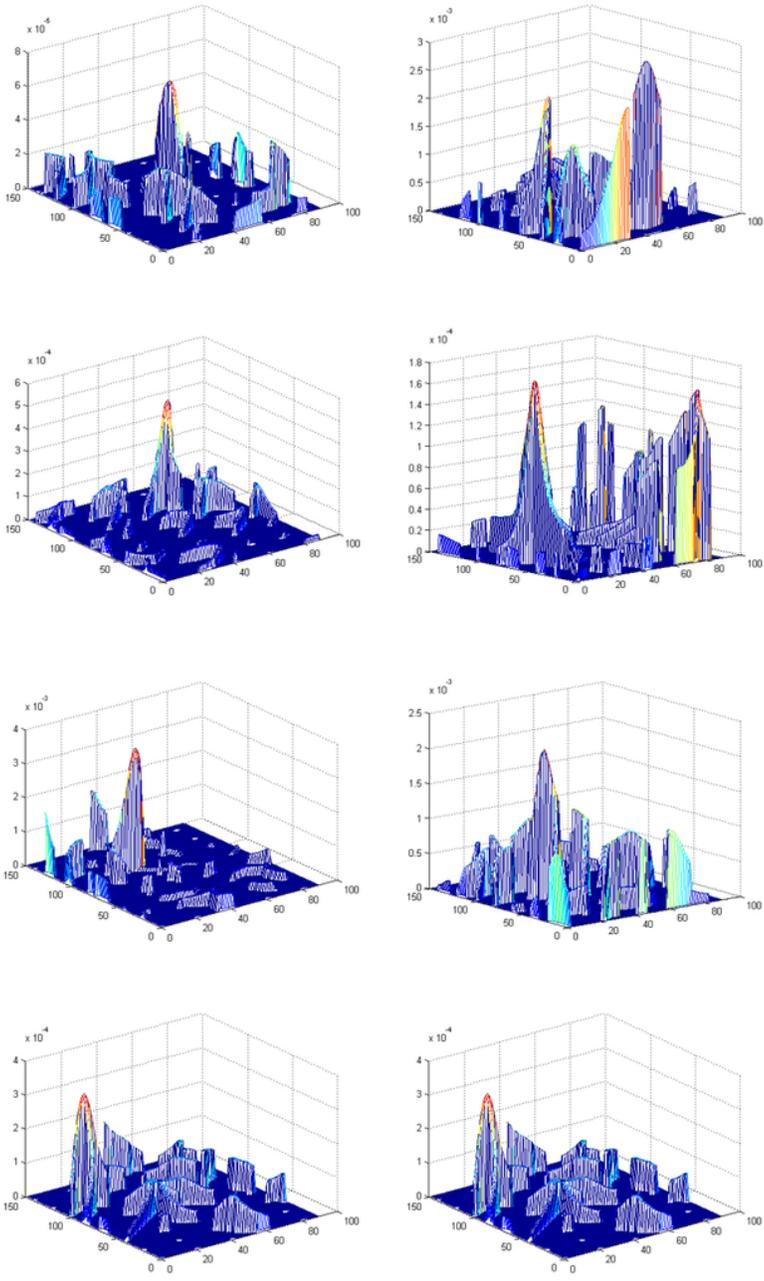


Figure 13

Matching related surfaces under rotational transformation.

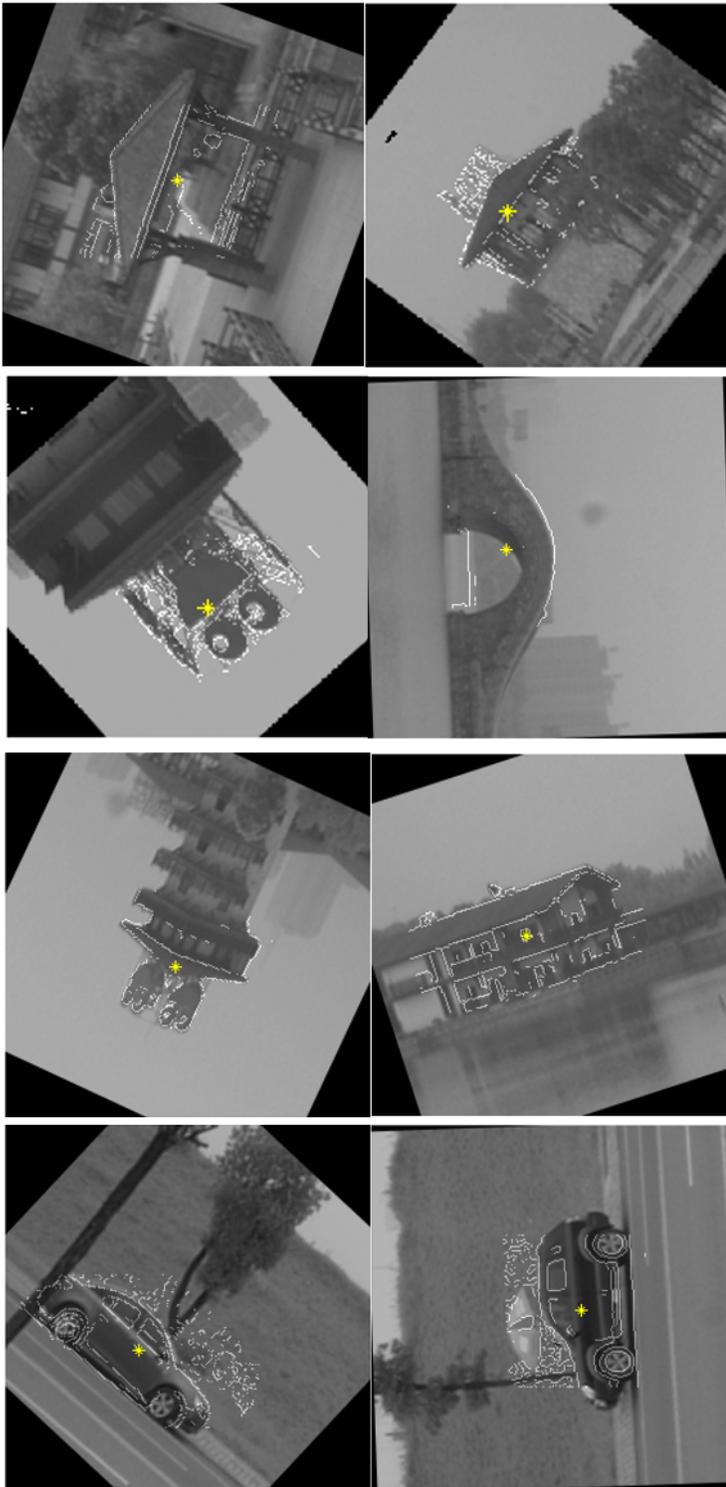


Figure 14

Matching results under rotation change.