A Two-Step Method for Remote Sensing Images Registration Based on Local and Global Constraints

Yue Wu, *Member, IEEE*, Zhenglei Xiao, Shaodi Liu, Qiguang Miao, *Senior Member, IEEE*, Wenping Ma, *Senior Member, IEEE*, Maoguo Gong, *Senior Member, IEEE*, Fei Xie, and Yang Zhang

***Abstract*—In this article, we propose an effective method for remote sensing image registration. Point features are robust to remote sensing images with low quality, small overlapping area, and local deformation. Therefore, we extract point features from remote sensing images and convert the problem of remote sensing image registration into the problem of feature point matching. A correspondence set constructed solely on the similar of features often contains many false correspondences or outliers, so our key idea is to remove the mismatches in the initial correspondence set and obtain a stable correspondence through a two-step strategy. First, we use two constraints to construct the optimization model which can solve in linear time. The first constraint is that the topology of the points and their neighbors can be maintained after the spatial transformation. Another constraint is that the feature distance of the correct matches are similar to the neighbors. Then, we design a strategy to increase the number of inliers and raise the precision by a global constraint calculated from the solution in the previous step. Experiments on a variety of remote sensing image datasets demonstrate that our method is more robust and accurate than state-of-the-art methods.**

***Index Terms*—Feature descriptor, global information, image registration, locality preserving, scale-invariant feature transform (SIFT).**

I. Introduction

IMAGE registration is a fundamental and challenging step in image processing [1]. Many image processing tasks require image registration in advance, such as image fusion [2], [3], change detection [4], [5], and hyperspectral image processing [6], [7]. Remote sensing image registration aims at aligning two or more images that contain overlapping area [7]. These images often have the problem of low quality, occlusion, and local distortion because they are obtained under different conditions, such as by different sensors, from different viewpoints, or at different times, so the task of remote sensing image registration is extremely difficult. Feature-based methods can register images of completely different nature and handle complex image distortions [9], so we formulate the problem of remote sensing image registration to feature matching.

The goal of feature matching is to establish reliable feature correspondences between two sets of features [10], [11]. It is crucial problem for many vision-based task [12]–[15]. The selection of feature points is crucial to the final result of feature matching. So during the past few decades, many studies focused on designing more robust local descriptors and exploring improvements in descriptor matching using alternate distance metrics [16], [18]. The scale-invariant feature transform (SIFT) feature [19] is one of the most widely used point features due to its invariant to image scaling, rotation, affine transformation, and illumination change. Even now, SIFT also appears in many tasks [20]–[22]. Ye *et al.* [18] define a similarity metric named HOPCncc, which uses the normalized correlation coefficient (NCC) of the HOPC descriptors for multimodal registration. It is robust against complex nonlinear radiometric differences. Ye *et al.* [24] combine a feature detector named MMPC-Lap and a feature descriptor named local histogram of orientated phase congruency (LHOPC). MMPC-Lap is constructed by using the minimum moment of phase congruency for feature detection with an automatic scale location technique. LHOPC derives the feature descriptor for a key point by utilizing an extended phase congruency feature with an advanced descriptor configuration. It is sufficiently robust to both geometric and radiometric changes. Although many well-designed algorithms have proposed to extract the local descriptors for image registration, there still exist some outliers in the result because of similar scenes and local deformation, which requires new approaches to solve the matching problem well. Now the mismatch elimination algorithms are frequently used [25], [26]. These methods build an initial set according to the similarity of descriptors first and then delete mismatches from the initial set to find the stable correspondences. This article intends to this method.

Existing mismatch elimination algorithms always rely on geometric constraints to remove mismatches [27], [28]. These

Manuscript received June 17, 2020; revised September 17, 2020, November 15, 2020, February 3, 2021, and March 31, 2021; accepted May 3, 2021. Date of publication May 11, 2021; date of current version May 28, 2021. This work was supported in part by the Innovation Fund of Shanghai Aerospace Science and Technology under Grant SAST2019-090, in part by the National Natural Science Foundation of China under Grants 62036006 and 61973249, and in part by the Key Research and Development programs of Shaanxi Province under Grant 2021ZDLGY02-06. *(Corresponding author: Maoguo Gong.)*

Yue Wu, Zhenglei Xiao, and Qiguang Miao are with the School of Computer Science and Technology, Xidian University, Xi'an 710071, China (e-mail: ywu@xidian.edu.cn; zlXiao@stu.xidian.edu.cn; qgmiao@mail.xidian.edu.cn).

Shaodi Liu and Wenping Ma are with the Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, School of Artficial Intelligence, Xidian University, Xi'an 710071, China (e-mail: sd_liu666@163.com; wpma@mail.xidian.edu.cn).

Maoguo Gong is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China (e-mail: gong@ieee.org).

Fei Xie is with the Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China (e-mail: fxie@xidian.edu.cn).

Yang Zhang is with the Shanghai Aerospace Electronic Technology Institute, Shanghai 201800, China (e-mail: ustczy@163.com).

Digital Object Identifier 10.1109/JSTARS.2021.3079103

constraints can be roughly divided into local constraints and global constraints. Despite these algorithms being continuously proposed during the past few decades, there are still many challenges in handling remote sensing image registration [29]–[32]. First, due to changes in ground fluctuations or changes in imaging viewpoints, local distortion often occurs in remote sensing images, which reduces the accuracy of the algorithms that rely on local constraints. Even though many algorithms can improve the accuracy by controlling the threshold, there are also algorithms that set a strict threshold, which will not only reduce the number of correct matches but also cannot improve the accuracy. Then, global constraints need to use the information of all matches; however, there are often a large number of false matches in the initial correspondence set, which is more obvious in remote sensing images. Algorithms that rely on global constraints are susceptible to outliers; even robust algorithms can hardly achieve good results.

To address the above challenges, we design an efficient and robust strategy for remote sensing image registration. Obviously, for an image pair of the same scene or object, the local topological relationship formed by the correct feature points changes little. In addition, for a correct match, the area where these two points are located in the same physical area. So the feature distances of matches located in the same physical area should be similar. Based on the above observations, we design two metrics to describe the consistency of the neighbor topology and the similarity of the feature distances of matches located in the same area. We combined these two constraints to construct a mathematical model and obtained its solution in linear time. So far, our first step is an improvement on locality-preserving matching algorithm (LPM), which relies on local information to eliminate mismatches. The first step of our method is inspired by LPM, because LPM proposes a framework for the effective use of local information. In order to get more accurate results, we increase the feature distance on the basis of LPM to make full use of local information. However, methods that only rely on local information have similar defects, that is, that the accuracy of the results may not be satisfactory when there are local distortion or similar scenes. To further increase the number of correct matches and remove false matches, we design a strategy to extract global information that is suitable for all correct matches from the previous solution and use it to the initial matching set. Specifically, we solve the transform matrix between the two images according to the previous solution. Compared with calculating the transform matrix based on the initial matching set, this method can effectively avoid the influence of outliers. The transform matrix is correct enough to truly reflect the spatial changes of the two pictures. Most remote sensing image transformations are affine transformations, so we use the affine matrix as the global transformation matrix in the second step. So by using the affine matrix as a guide, we can get a better result than just using local constraints.

In summary, our contributions in this article are as follows.

1) We propose a method of using global information. Compared with the existing methods, our method is robust to outliers in initial matching set. At the same time, our method can be combined with other remote sensing image registration methods, which provides a possibility to improve the result of them.

2) We combine local topology information and local feature information to construct a mathematical model that can be solved in linear time. The solution of the model can well reflect the spatial transformation, and is more accurate.

3) We innovatively use a simple method to effectively combine global and local information for feature matching. Compared with other methods, our local and global information combination is more effective. It has the advantages of local and global constraints at the same time, and to a certain extent makes up for the disadvantages of them.

The remainder of this article is organized as follows. In Section II, we introduce the background and related works. In Section III, we introduce our algorithm in detail, and we make full use of local and global constraints for remote sensing image registration. Section IV provides the evaluations of our method in comparison with several state-of-the-art methods on different remote sensing image datasets. Section V gives the conclusion.

II. Related Work

In this article, we propose a feature-based remote sensing image registration method. Specifically, we formulate the registration problem as a matching problem. Here we briefly review the work for feature matching. In general, the problem of feature matching can be solved by two kinds of methods. The first type is the mismatch elimination algorithm. This type of method establishes accurate correspondences by filtering out unstable matches in the initial matching set [33]. Another type directly estimates the point-to-point matching relationship of two points without an initial matching set.

The quality of the feature is very important to the final result. So many methods focus on engineering robust local descriptors. David Lowe [34] proposed the classic SIFT algorithm. It is widely used in various situations due to it is invariant to rotation, scaling, and brightness changes. Many methods [18], [34] are based on it. For example, Lowe [19] compares the ratio between the nearest and the second nearest neighbors against a predefined threshold to get stable correspondences. Although there are many methods for constructing an initial matching set, it is difficult to avoid false matches in the results by using only local descriptors. So it is critical to solving the feature matching problem in a two-stage method. Mismatches are filtered by using additional constraints in the second stage.

Some methods use global information as constraints, which can be divided roughly into two main categories, say parameter estimation methods [11], [35], [36] and nonparametric interpolation methods [37]–[39]. The random sample consensus (RANSAC) algorithm [40] is a typical representative method of parameter estimation methods. It attempts to obtain the smallest possible outlier-free subset to estimate a provided parametric model by resampling. RANSAC and its similar methods [41]–[43] rely heavily on the accuracy of sampling results. Obviously, when there are a large number of outliers in

the initial matching set, the number of sampling times required is significantly increased, which greatly affects the efficiency of these methods. The non-parametric fitting methods learn a predefined nonparametric model based on prior knowledge or through regression. It maps feature points from one image to another and then eliminates false matches by checking whether each match is consistent with the estimated non-parametric function. The representative of such methods are [44] algorithm that eliminates outlier by estimation robust correspond function, vector field consensus (VFC) [45] algorithm based on vector field consistency, and [46] algorithm based on manifold regularization. These methods nearly use the information of all points, so the precision of these methods will decrease sharply when there are many outliers or there are independent motion structures in the point sets.

Some methods use local information as constraints [9], [47], [48]. These methods use constraints that are similar to graph matching to perform a robust estimation of correct matching based on the assumptions of local structural consistency or piecewise consistency. For example, the LPM [49] assumes that local geometric structures in the vicinity of inliers are invariant under geometric transformations which is similar to the idea of the first step in our method. Based on the assumption, it constructs a model with the neighborhood relationship and neighborhood topology of feature points as constraints, which effectively improves the calculation speed and robustness of the algorithm. Jiang *et al.*'s work [50] is an improvement in LPM. It considers the order of neighbors of each inlier and introduces a stricter measuring criterion for local constraint. Bian *et al.* [51] introduce a statistical strategy, which removes outliers based on the number of neighboring matches. The reason why it is feasible is the consistency of sports field of correct matches, which converts the quantity of matches into the quality of matches. These methods fully mined the information of local constraints, so they are efficient, but the accuracy of their results will decrease when there are local distortion or similar scenes. Our method combines local and global information. Its goal is to get a better result than the above methods that use only global or local information.

Except for the mismatch elimination algorithm, another feature-based matching method is to directly estimate the correspondence between two point sets. These methods can be divided roughly into two main categories. The first one is to estimate the transformation matrix between point sets for feature matching [52]–[55]. The transformation matrix can map the target point sets to the template point sets so that the matching points in the two sets can coincide as much as possible after transformation. Another is based on the graph model [56]–[60]. These approaches treat the feature points as the vertices of the graph, the vertices can be connected to edges, and the point set as the graph. Their goal is to search for the largest subset with similar graph structures from two point sets. Both of these approaches can be viewed as optimization problems, and only the way to construct the optimization equation is different. However, when there are a large number of outliers or the data degradation is serious, the performance of these algorithms will be greatly reduced or even fail.

III. Proposed Method

This section describes our method for remote sensing image registration by establishing accurate correspondences between two feature sets extracted from two remote sensing images. There are many well-designed feature descriptors that can efficiently establish initial correspondences. The SIFT descriptor with a distance ratio method [19] that compares the ratio between the nearest and the second-nearest neighbors against a predefined threshold is used to gain an initial matching set. The emphasis of this article is to eliminate mismatches, so the details of building an initial matching set will not be discussed here. In the following, we will introduce a two-stage feature matching method based on local and global information. First, we remove outliers from initial set and construct approximately accurate correspondences by combining the feature descriptor and local structure information. Then, the transformation of these correspondences can be used as global information to obtain a more effective result.

A. Construct an Approximately Correct Solution

Suppose a set of N initial correspondences $S = \{(x_i, y_i, m_i, n_i)\}_{i=1}^{N}$ extracted from different images have been obtained, where x_i and y_i are vectors denoting the spatial positions of feature points and m_i and n_i are vectors denoting the descriptors of feature points. This article tries to remove the outliers in S and get establish correspondences in a two-stage manner.

For an image pair of the same scene or object, the topological relationship formed by the correct feature points in an area is not changed easily under spatial transformations. In addition, points that constitute a correct match are located in the same physical area. Their feature distance can indicate the difference between the two images in the same area. So the feature distances of matches in the same area are similar. Based on the above observations, we design two metrics to build an optimization model. Denoting I is the unknown inliers set, the optimization model is

$$I^* = \underset{I}{\operatorname{argmin}}\, C(I; S, \lambda) \tag{1}$$

with the cost function C defined as

$$\begin{aligned} C(I; S, \lambda) = \sum_{i \in I} \Bigg\{ & \sum_{j | x_j \in N_{x_i}} [d(x_i, x_j) - d(y_i, y_j)]^2 \\ & + \sum_{j | y_j \in N_{y_i}} [d(x_i, x_j) - d(y_i, y_j)]^2 \\ & + \beta \sum_{j | x_j \in N_{x_i}, y_j \in N_{y_i}} [fd(m_i, n_i) - fd(m_j, n_j)] \Bigg\} \\ & + \lambda(N - |\lambda|) \end{aligned} \tag{2}$$

where $fd(.)$ and $d(.)$ are two measurement functions. The $fd(.)$ is a certain distance metric such as Euclidean distance. It calculates the Euclidean distance between two features. To increase the robustness of the algorithm, we normalized the feature distances. Assuming that the maximum value of all

feature distances is $x_{\max}$, the minimum value is $x_{\min}$, and the normalized value of any value x is $(x - x_{\min}) / (x_{\max} - x_{\min})$. Therefore, the range of the characteristic distance becomes 0–1, and the range of the characteristic distance term can be manipulated by β. The definition of d is shown in (3).

$$d(x_i, x_j) = \begin{cases} 0, x_j \in N_{x_i} \\ 1, x_j \notin N_{x_i} \end{cases}. \tag{3}$$

The d is used to measure whether a pair of points are neighbors. The N_x denotes the point x and its K neighbors. If there are two matches (x_i, y_i) and (x_j, y_j), x_j is the neighborhood point of x_i. And (x_i, y_i) is a correct match, due to the topological relationship formed by correct matches should be maintained, y_j should be the neighborhood of y_i, that is, $d(y_j, y_i)$ is 0, on the contrary, $d(y_j, y_i)$ is 1. So the first and second terms of (2) are to find the similarity of the topological structure of a matching. The third term $fd(.)$ compares the distance between the feature difference of a pair of matches and its neighbors. It is easy to find that the distance of the wrong match is large. β is used to balance the proportion of two distances. The last term is similar to the regularization term, which discourages the outliers. Its goal is to increase the number of correct matches in the result as much as possible and prevent the solution tending to make all matches wrong. The parameter $\lambda > 0$ and $\beta > 0$ controls the tradeoff between these three terms. Ideally, the optimal solution should be near zero, i.e., the first term of C should be zero, and the second term of C should be very small.

To calculate conveniently, we use an $N \times 1$ binary vector p to indicate the correctness of these N pairs of matches. Specifically, $p_i \in \{0,1\}$ denotes the correctness of the ith match (x_i, y_i, m_i, n_i), where $p_i = 1$ indicates the ith match is correct match and $p_i = 0$ indicates the ith match is error match. In this case, the cost function in (2) is converted to

$$\begin{aligned} C(p; S, \lambda) = \sum_{i=1}^{N} p_i \Bigg\{ & \sum_{j|x_j \in N_{x_i}} d(y_i, y_j) + \sum_{j|y_j \in N_{y_i}} d(x_i, x_j) \\ & + \frac{\beta}{K} \sum_{j|x_j \in N_{x_i}, y_j \in N_{y_i}} [fd(m_i, n_i) - fd(m_j, n_j)] \Bigg\} \\ & + \lambda \left(N - \sum_{i=1}^{N} p_i \right). \end{aligned} \tag{4}$$

It can be solved by optimizing p to minimize the cost (4) to eliminate false matches and construct stable feature matches, and K is the number of points in the neighborhood. Because $\sum_{j|x_j \in N_{x_i}} d(y_i, y_j)$ and $\sum_{j|y_j \in N_{y_i}} d(x_i, x_j)$ are numerically equal and physical consistency, it can be be further simplified. To optimize the objective (4), we merge the terms related to p_i and obtain

$$C(p; S, \lambda) = \sum_{i=1}^{N} p_i (c_i - \lambda) + \lambda N \tag{5}$$

where

$$\begin{aligned} c_i = & 2 \sum_{j|x_j \in N_{x_i}} d(y_i, y_j) \\ & + \frac{\beta}{K} \sum_{j|x_j \in N_{x_i}, y_j \in N_{y_i}} [fd(m_i, n_i) - fd(m_j, n_j)]. \end{aligned} \tag{6}$$

For each correspondence i in initial correspondence set, c_i is a fixed value, due to the neighborhood relationship and the feature distance between the feature points are fixed when the initial matching set does not change, and hence we can calculate all $\{c_i\}_{i=1}^{N}$ in advance. That is to say, in (5), only the value of p_i is uncertain. To optimize (5), any match with a cost smaller than λ will decrease the value of cost function, while any match with a cost bigger than λ will increase the value of cost function. Therefore, the optimal solution of p that minimizes (5) is determined by the following criterion:

$$p_i = \begin{cases} 1, c_i \leq \lambda \\ 0, c_i > \lambda \end{cases}, i = 1, \ldots, N. \tag{7}$$

And hence, the optimal match set I^* is determined by

$$I^* = \{i | p_i = 1, i = 1, \ldots, N\}. \tag{8}$$

From (7), we see that parameter λ is also a threshold used to judge whether a correspondence is correct. The setting of p_i is arbitrary when $c_i = \lambda$.

This strategy works well due to the following reasons. On the one hand, for an outlier (x_i, y_i, m_i, n_i), its local neighborhood structures are different between two images, and the feature distance between it and its neighbors is also large, which leads to a large cost c_i, and hence it will be easily identified as an outlier. On the other hand, for an inlier (x_j, y_j, m_j, n_j), even if there are some outliers in their neighborhood N_{x_j} or N_{y_j}, the proportion of inliers is still greater than that of outliers. m_j and n_j represent the same physical area. The feature distance of their neighbors will not be too different from theirs. Hence, its cost c_j will not be large.

In our evaluation, we adopt a simple strategy that searches K ($K = 4$ in default) nearest neighbors under the euclidean distance for constructing neighborhood N_x to each point x. Since the N_x is constructed based on the initial matching set, not all points in N_x are inliers. If the neighborhood N_x is constructed by all inliers, the result will be more accurate. As the result is not influenced by the outliers, the difference between the inlier and outlier will be enlarged, which makes the result easier to distinguish, especially when there are a large number of outliers in S. However, the true inlier set I cannot be gained in advance because our aim is to find the inlier set. To solve this problem, we use an approximation set to replace it. We previously obtained a solution I_0, which is mainly composed of inliers. So we use I_0 to construct a neighborhood for each match in S and solve the optimal I^* as

$$I^* = arg \min_{I} C(I; I_0, S, \lambda). \tag{9}$$

TABLE I
COMPARISON OF THE RESULT OF ONE STAGE AND TWO STATE

Method	The first stage	**The second stage**
Average Precision	92.09%	**95.54**%
Average Recall	97.60%	**98.81**%

B. Gain More Efficient Result Based on Transformation Matrix

Since we not only use local topology structure but also add a constraint about feature distance, we can get correspondence I_1 that is closer to the optimal solution. Then, the global information can be obtained by calculating the transformation matrix T through the optimal I_1. In remote sensing image registration tasks, the affine transformation can solve most of the registration problems. So we use the affine transformation as the global transformation T, and estimate its parameters by the least square method. So this transform matrix T can theoretically fit most correct matches, and we can get a more stable correspondence set. The goal of image registration is to obtain an accurate transformation matrix so that the matching points of the perceived image can perfectly coincide with the matching points of the reference image after transformation. But in this process, there is a high probability that the matching points do not coincide due to local deformation, but they are not far away. So the mismatches can be eliminated by the following formula:

$$s_i = fc(x_i, Ty_i) \tag{10}$$

$$p_i = \begin{cases} 1, s_i \le \lambda_2 \\ 0, s_i > \lambda_2 \end{cases}, i = 1, \ldots, N. \tag{11}$$

where

$$\lambda_2 = hf(I). \tag{12}$$

The metric function fc is used to calculate the Euclidean distance of two coordinates. The s_i is the distance of the matching points of the perceived image and the matching points of the reference image after transformation. λ_2 is a new threshold value, which is positively correlated with the pixels of the image, and $f(I)$ is the pixel of picture I (the diagonal of the picture), h is the scale factor, then we can get the final correspondence I_2 using (11) and (12).

To verify whether the second step can increase the number of correct matches and improve the accuracy, we conducted experiments to compare the results of only the first step and the results of the two steps while keeping the parameters consistent. Eighteen representative images are selected as experimental data. The experimental result is shown in Table I. The two-step strategy can effectively improve the accuracy and recall of the algorithm.

The procedure of the algorithm is shown in Algorithm 1.

Algorithm 1: Procedure of the Mismatch Elimination by Local and Global Constraints.

Input: The putative feature correspondences $S = \{(x_i, y_i, m_i, n_i)\}_{i=1}^{N}$ and Parameters K, β, λ, and h.
Output: The correspondence I_2

1: Construct neighborhood $\{N_{x_i}, N_{y_i}\}_{i=1}^{N}$ based on S;
2: Calculate cost $\{c_i\}_{i=1}^{N}$ using (6);
3: Determine the approximate correspondence I_0 using (7) and (8);
4: Construct neighborhood $\{N_{x_i}, N_{y_i}\}_{i=1}^{N}$ based on I_0;
5: Calculate cost $\{c_i\}_{i=1}^{N}$ using (6);
6: Determine the approximate correspondence I_1 using (7) and (8);
7: Calculate the transformation matrix T between two images obtained by the I_1;
8: Calculate cost $\{s_i\}_{i=1}^{N}$ using (10) according to T;
9: Determine the final correspondence I_2 using (11) and (12);

C. Computational Complexity Analysis

Our method concludes two steps: the construction of the correspondences by local constraints and the construction of more correct matches based on the transformation. The computational complexity analysis is described as follows. We use the following formula to approximately approach spending time:

$$T_{\text{total}} = T_{\text{construction}} + T_{\text{reconstruction}} \tag{13}$$

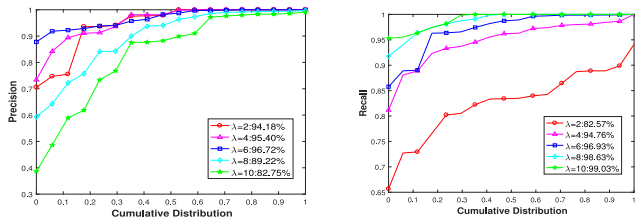
$T_{\text{construction}}$ is the time to construct the optimal correspondence by local constraints, and $T_{\text{reconstruction}}$ is the time to remove outliers and increase inliers based on the transformation. The time complexity to search the K nearest neighbors for each point in S is about $O((K+N)\log N)$ the time complexity to calculate the cost $\{c_i\}_{i=1}^{N}$ is $O(KN)$, and the time complexity to calculate the cost $\{s_i\}_{i=1}^{N}$ is $O(N)$. Therefore, the total time complexity of our method is about $O(KN + (K+N)\log N + N)$. Generally, the time complexity of our method can be simply written as $O(N \log N)$.

IV. EXPERIMENT AND RESULTS

A. Evaluation Criterion

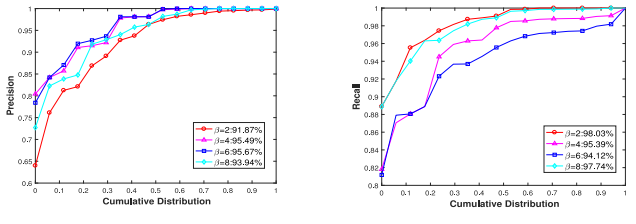
The number of correct correspondences is an important criterion to evaluate the effectiveness of the proposed method. Our method of marking the correct match is as follows. First, select a part of the matches, manually select the correct matches, and estimate the transformation parameters between the images. Then, other matches suitable for the transformation relationship within a certain error range are regarded as the correct matches. In order to improve the accuracy of the labeling, we manually modified the labels on the area with large changes. In addition, there are many effective criteria. For our experimental results, the evaluation is carried out using the following criteria.

1) Number of correct matches (NCM)
   The number of correct correspondences is used as the criterion to evaluate the robustness of the proposed method.
2) Precision
   Precision is defined as the ratio of correct matches divided by the sum of correct matches and false matches. It can be



(a) (b)

Fig. 1. Accuracy and recall rate of different λ when $\beta = 4$.



(a) (b)

Fig. 2. Accuracy and recall rate of different β when λ=6.

expressed as follows:

$$\text{Precision} = \frac{\text{correct matches}}{\text{correct matches+false matches}}. \quad (14)$$

3) Recall
Recall is the number of correctly matched points with respect to the number of corresponding points between two images of the same scene. Correspondences are the total matches that satisfy the transformation matrix within the range of allowable threshold.

$$\text{Recall}=\frac{\text{correct matches}}{\text{correspondences}}. \quad (15)$$

4) Root-mean-square error (RMSE)
The accuracy is evaluated by the root-mean-square error (RMSE) criterion. The RMSE can be expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - X_i')^2 + (Y_i - Y_i')^2} \quad (16)$$

where (X_i', Y_i') means the transformed coordinates of (X_i, Y_i). N means the total number of correctly matched pairs.

B. Parameters and Test Datasets

1) Parameters Settings: There are four parameters in our method: K, β, λ, and h. The K is the number of nearest neighbors for neighborhood construction, the β controls the proportion of the feature distance and the local neighborhood information, the λ is the threshold for judging whether a match is a correct correspondence, and the h sets the maximum distance of the correct matching points of the perceived image and the matching points of the reference image after transformation. In our evaluation, we set the default values as $K = 4$, $\beta = 4$, and $\lambda = 6$. The relationship between β and λ is difficult to determine. To analyze the sensitivity of the two parameters, we will fix one value and change another value to experiment. When $\beta = 4$, $\lambda = 6$, the result of first step is the best. The experimental results are shown in Figs. 1 and 2. The X-axis of these pictures means the cumulative distribution. It is the probability that the value of test data less than the value of current ordinate. So Fig. 1 is the result of fixed β changing λ. The accuracy rate reaches the extreme value when $\lambda = 6$. The recall rate increases as λ increases, so to ensure precision and get enough recall at the same time, we set λ to 6. Fig. 2 is the result of fixed λ changing β. The accuracy and recall are the maxima when $\lambda = 6$ and $\beta = 4$, respectively. It is worth noting that the input has an influence on the selection of the algorithm's threshold. Our input is different from the original LPM input. In the experimental environment of this article, the threshold of LPM is no longer 6, but 4. The experiment result shows that which is β has little effect on the final result, so we finally set β to 4. The way to determine h is simple. The larger the h, the lower the accuracy, and the higher the recall of the result. We compared the experimental results of different h on 18 pictures. The different results are shown in Fig. 3. The recall increases with the increase of h when $h < 0.038$, and the rate of increase tends to slow down when $h > 0.029$, and the accuracy decreases with the increase of h when $h > 0.026$; when $h > 0.032$, the rate of decrease becomes larger. When the value of h is between 0.029 and 0.038, the recall is high and when the value of h is between 0.026 and 0.032, the precision is high. Finally, we chose $h = 0.032$.

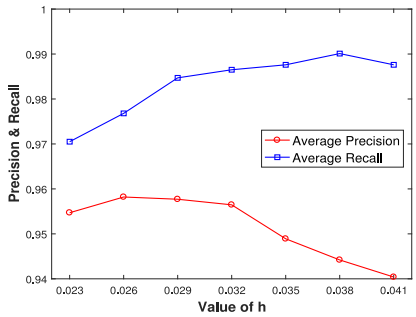


Fig. 3. Comparison of the average accuracy and average recall rate corresponding to different h values.

2) Test Datasets: To evaluate the effectiveness of our method, we selected a total of 102 pairs of images from four datasets for testing. We selected 10 pairs of representative images from them to show. These images are shown in Fig. 4.

1) *Image pairs 1 and 2:* These images are from CIAP.[1] In this dataset, 30 pairs of images are selected for testing. These images are all of size 700 × 700 and have already been orthorectified. The feature matching task for such image pairs typically arises in the image mosaic problem. The images are publicly available (from the Erdas example data), which are captured over eastern Illinois, IL, USA.
2) *Image pairs 3, 4, and 5:* These images are from UAV [61]. In this dataset, 22 pairs of images are selected for testing. These images only a few parts of scenes in each pair of these images are overlapped, which greatly increases

[1]Online. [Available]: http://download.intergraph.com/downloads/erdas-imagine-2013-2014-example-data

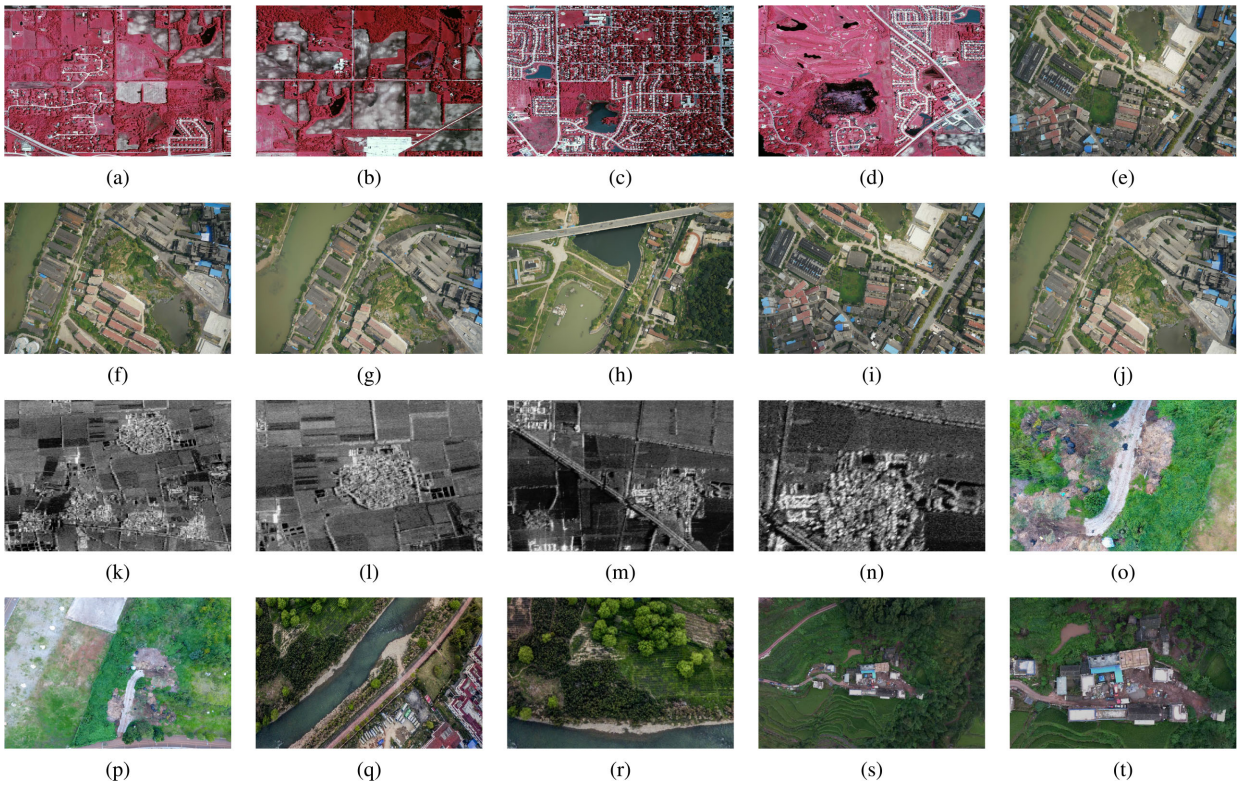


Fig. 4. Ten image pairs selected from the test data. (a) and (b) Pair 1. (c) and (d) Pair 2. (e) and (f) Pair 3. (g) and (h) Pair 4. (i) and (j) Pair 5. (k) and (l) Pair 6. (m) and (n) Pair 7. (o) and (p) Pair 8. (q) and (r) Pair 9. (s) and (t) Pair 10.

the difficulty of matching. Most of them are captured by a UAV over Yongzhou city and HeJiangdong villages of Hunan province, China, which are suitable for image Mosaic tasks.

3) *Image pairs 6 and 7:* These images are from SAR. In this dataset, 10 pairs of images are selected for testing. These image pairs are corrupted with strong noise, which brings great challenges to the image matching task. Every image pair is obtained by synthetic-aperture radars (SARs) on a satellite. They mainly include farmland, cities, mountains, and rivers. Feature matching for such image pairs has been widely used in the military field, disaster monitoring, and resource detection. These image pairs can be modeled with similarity or rigid transformation in most cases.
4) *Image pairs 8, 9, and 10:* These images are from SUIRD. Forty pairs of images are selected for testing. This dataset is a public dataset for image registration/matching research. It includes 50 pairs of images and their groundtruth. These image pairs include horizontal, vertical and mixed perspective variations, vertical and their mixture which produces problems of low overlap, image distortion, and severe outliers. Since our algorithm requires feature descriptors, we reannotate the data set.

C. Experimental Results

In this section, we test the performance of our proposed method on different kinds of datasets and compare it with other feature matching methods. Five state-of-the-art methods including RANSAC [40], MR-RPM [46], LPM [49], mTop-KPR [50], and GLPM [9] are chosen for comparison. To be fair, we implement these algorithms based on publicly available codes and try our best to tune their parameters to make them best. To verify whether our algorithm can find more correct matches and remove false matches, the method only with first step is as a comparative method.

1) Qualitative Results: First of all, some representative pictures are extracted from the datasets for experiments, where both CIAP and SAR contain two pairs of images, and both UAV and SUIRD contain three pairs of images. The selected image pairs include most challenges about remote sensing images, such as ground relief variations, severe noise, and small overlap area. The initial inlier percentages of these ten images are 0.98, 0.93, 0.88, 0.34, 0.75, 0.73, 0.74, 0.55, 0.95, and 0.78. The match results of the ten image pairs are shown in Fig. 5, and the NCM, the Precision, the Recall, and RMSE for each pair are listed in Table II. As shown in Fig. 5, our algorithm can effectively remove the mismatches in the initial correspondence set and achieve an effective result even when the overlapping area is small and the image quality is low. By using our proposed method to filter out the mismatches, the precision and recall of these image pairs are being $(100\%, 99\%)$, $(100\%, 99\%)$, $(100\%, 97\%)$, $(91\%, 100\%)$, $(98\%, 98\%)$, $(100\%, 100\%)$, $(95\%, 100\%)$, $(96\%, 98\%)$, $(99\%, 99\%)$, and $(100\%, 100\%)$. The accuracy and recall of our method have always been the highest. In pairs 4, 6, and 7, the NCM of some algorithms is the same as ours, but our recall is very high, which shows the effectiveness of our method that uses the transformation matrix to increase the

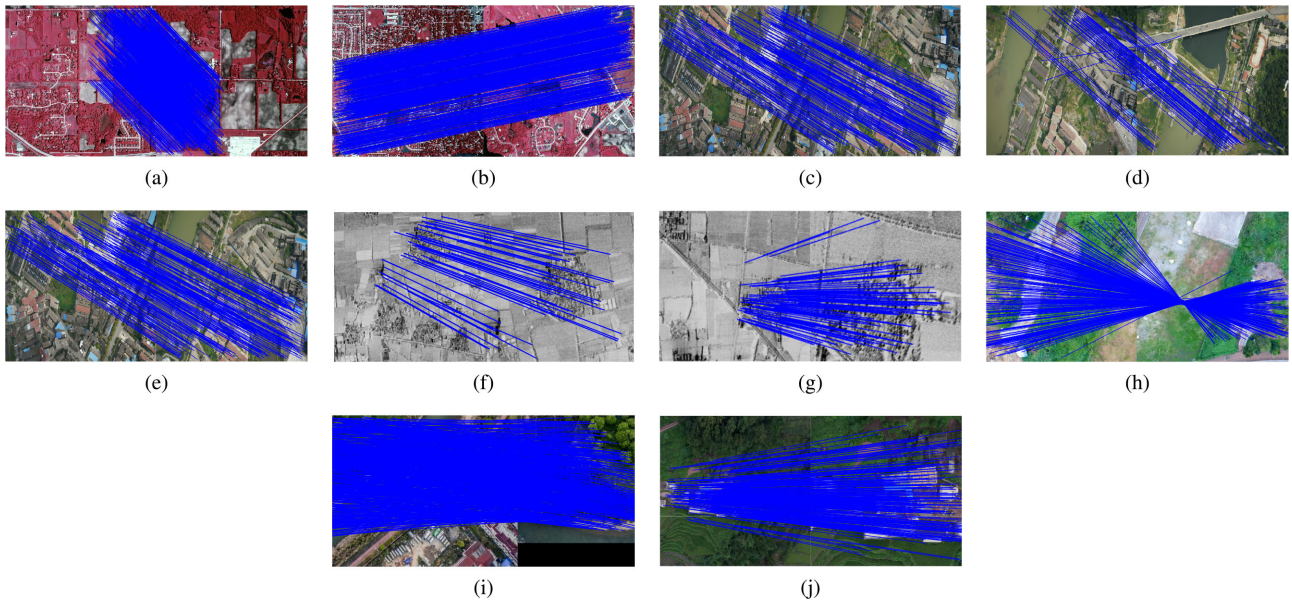


Fig. 5. Feature matching results of the proposed method on ten representative image pairs. (a) Pair 1. (b) Pair 2. (c) Pair 3. (d) Pair 4. (e) Pair 5. (f) Pair 6. (g) Pair 7. (h) Pair 8. (i) Pair 9. (j) Pair 10.

number of correct matches. It is worth noting that although the accuracy and recall of LPM are not as good as ours in pair 1, its RMSE is smaller than ours, because LPM eliminates mismatches through the topology structure of neighbors, which is very suitable for the situation with a few outlier points. The effect of LPM is not as good as ours for the picture with lots of outlier points. One reason is that our algorithm takes into account the differences in the feature distance in a region, which effectively removes a lot of mismatches. Another reason is that we use the global constraints to get more correct matches and remove error matches. The results of the mTopkRP and GLPM are similar to LPM, and the reason may be that they also rely only on local constraints. In addition, since we used the SIFT with distance ratio to get the initial matching set, there are not many outliers in our initial matching, which may cause GLPM to not play its advantages. Our algorithm belongs to the method of estimating transformation matrix like RANSAC, but our algorithm performs better than RANSAC. The reason is that RANSAC is a random method, which depends on the accuracy of the points obtained by sampling. When there are many error points, the number of iterations it requires will increase greatly, which takes a lot of time. Our algorithm achieves satisfactory results in all the images, although it is not outstanding in the SAR image; the registration effect is worse due to the low quality and high noise in the SAR images.

To show the advantages of our algorithm more clearly, we have visualized part of the experimental results. The comparison algorithms are LPM, RANSAC, and mTopkRP, respectively. The experimental results are shown in Fig. 6. The four columns of Fig. 1 are the results of LPM, mTopkRP, RANSAC, and our method on five pairs of images. Each row represents the experimental results of the above four methods on one pair of images. For example, the four pictures in the first row are the matching results of LPM, mTopkRP, RANSAC, and our algorithm on the first data. In the first row, there are some mismatches between the result of LPM and the result of RANSAC. Compared with the result of our algorithm, the result of mTopkRP has fewer correct matches. This shows that our method has the highest precision, and the recall is close to LPM and RANSAC. The visualized results are consistent with the results of numerical analysis. The results from the second row to the fifth row show that in addition to our method, the results of the other three methods all have incorrect matches. Although it is difficult to compare the recall through the visualized results, it can be concluded from the previous numerical analysis that the recall of our results is close to or even higher than that of LPM and RANSAC. So it can prove that our algorithm can achieve the highest precision when the recall is close to other algorithms, and our constraint combined local and global information is effective. Our method uses the solution obtained by strict local constraint to construct the global information, thus avoiding the mismatching caused by the local similarity of the images. At the same time, to compensate for the loss of correct matches due to strict local constraints, we use global constraints to search for potential correct matches to increase the number of correct matches of the final result. As the experiment shows, our method achieve effective results.

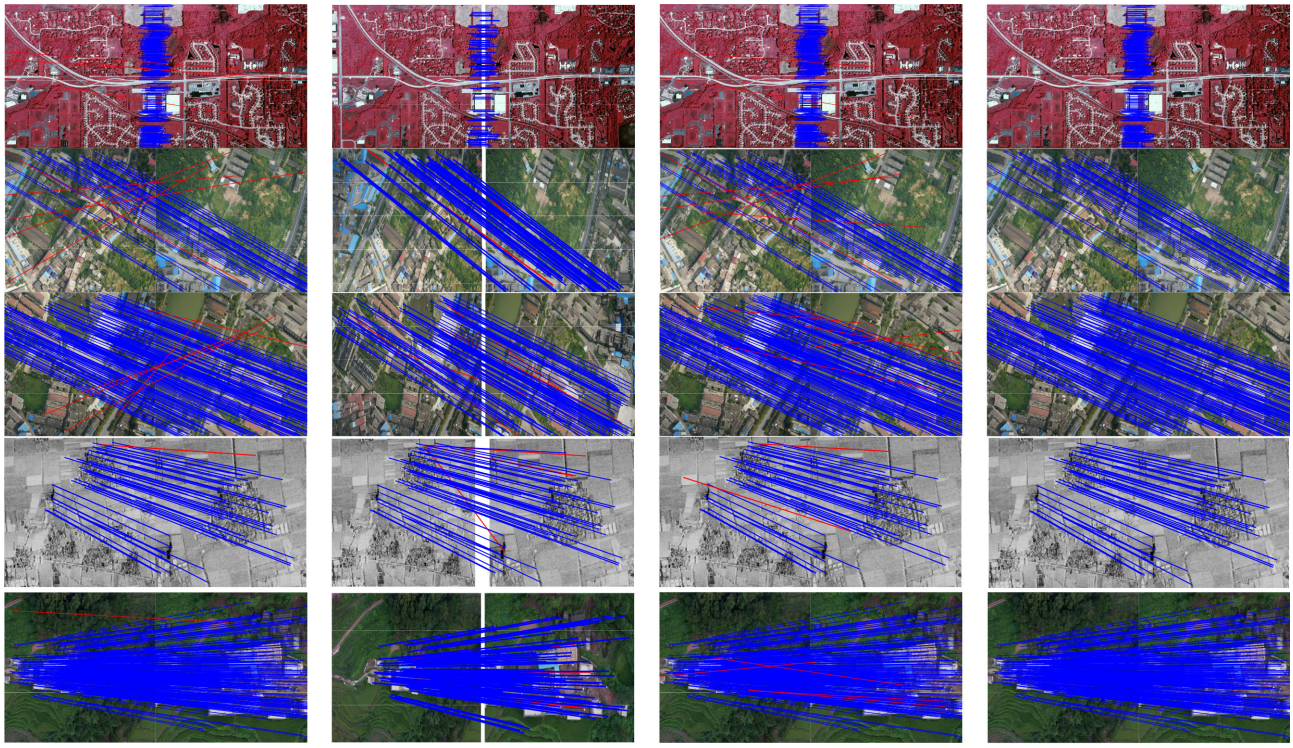
2) Quantitative Results: To provide a comprehensive quantitative evaluation of our method, we next conduct experiments on all image pairs in the datasets. The initial inlier percentage, precision, and recall statistics of the five algorithms are reported in Fig. 7. The results on the datasets are similar to those before, and our method achieves the highest accuracy when the recall is similar. In addition, when there are a few outliers in the initial matching set, all algorithms can achieve good results. However, when there are many outliers, the results of other methods except mine will deteriorate sharply. The accuracy of RANSAC and MR-RPM will decrease to below 0.5, and the LPM will also be below 0.7, which is like GLPM and mTopKPR. The curves

TABLE II
COMPARISON OF NUMBER OF CORRECT MATCHES(NCM), PRECISION, RECALL, AND RMSE OF DATA SELECTED FROM TEST DATASET

Data		The original	RANSAC	MR-RPM	LPM	mTopkRP	GLPM	Proposed
Data1	NCM	1118	1109	981	1114	1116	1115	**1117**
	Precision	0.98	1	0.99	1	1	0.99	**1**
	Recall	1	0.99	0.87	0.99	0.99	0.99	**0.99**
	RMSE	78.88	0.24	17.38	**0.15**	0,22	0.45	1.11
Data2	NCM	761	752	753	755	756	756	**760**
	Precision	0.93	0.99	0.99	0.99	0.99	0.99	**1**
	Recall	1	0.98	0.98	0.99	0.99	0.99	**0.99**
	RMSE	60.26	2.06	6.38	6.62	5.37	3.89	**3.26**
Data3	NCM	195	188	185	179	188	185	**191**
	Precision	0.88	0.99	0.92	0.98	0.98	0.98	**1**
	Recall	1	0.96	0.94	0.91	0.96	0.94	**0.97**
	RMSE	94.32	26.66	19.45	23.83	20.35	25.66	**13.29**
Data4	NCM	62	62	62	61	61	61	**62**
	Precision	0.34	0.86	0.74	0.87	0.88	0.88	**0.91**
	Recall	1	1	1	0.98	0.98	0.98	**1**
	RMSE	222.17	30.05	221.39	57.25	27.57	29.39	**25.55**
Data5	NCM	189	188	179	190	190	191	**195**
	Precision	0.75	1	0.96	0.97	0.98	0.98	**0.98**
	Recall	0.97	0.96	0.91	0.97	0.98	0.98	**0.98**
	RMSE	171.85	6.66	62.80	75.86	70.63	60.28	**6.62**
Data6	NCM	58	49	47	58	57	57	**58**
	Precision	0.73	0.96	0.97	0.99	0.95	0.97	**1**
	Recall	1	0.84	0.81	1	0.98	0.98	**1**
	RMSE	40.40	13.04	20.22	1.73	37.91	30.88	**0.67**
Data7	NCM	89	81	85	82	86	85	**89**
	Precision	0.74	0.84	0.95	0.93	0.91	0.94	**0.96**
	Recall	1	0.91	0.96	0.92	0.96	0.96	**1**
	RMSE	38.25	15.27	9.05	12.42	9.87	6.99	**0.47**
Data8	NCM	315	292	304	301	292	291	**310**
	Precision	0.55	0.92	0.75	0.91	0.90	0.92	**0.96**
	Recall	1	0.92	0.96	0.95	0.92	0.92	**0.98**
	RMSE	246.25	9.33	238.28	49.58	55.23	50.86	**9.33**
Data9	NCM	718	703	714	701	711	712	**715**
	Precision	0.95	0.99	0.98	0.99	0.99	0.99	**0.99**
	Recall	1	0.97	0.99	0.97	0.99	0.99	**0.99**
	RMSE	70.45	5.07	30.29	5.01	10.28	9.86	**4.10**
Data10	NCM	315	312	310	313	310	311	**315**
	Precision	0.78	0.99	0.90	0.99	0.99	0.99	**1**
	Recall	1	0.99	0.98	0.99	0.98	0.98	**1**
	RMSE	63.63	10.13	17.42	15.34	10.76	12.34	**9.87**

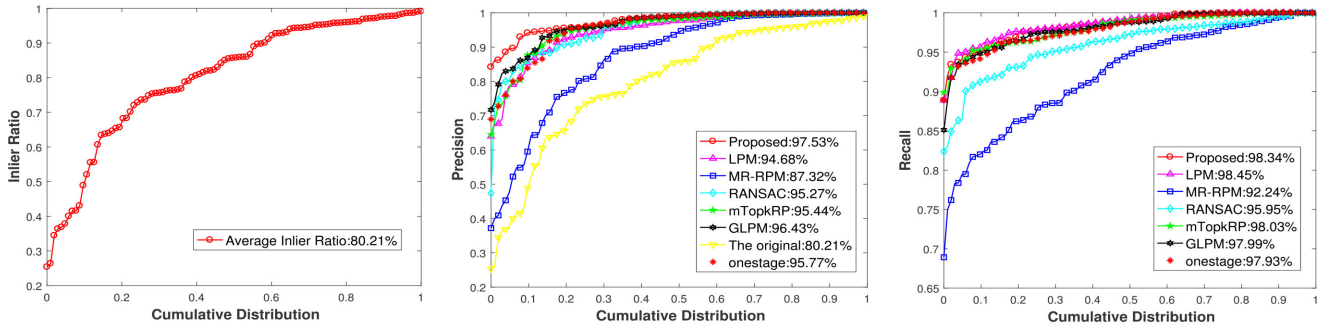
The results in bold are the best.

of accuracy and recall rate of RANSAC can be simply viewed as two stages with a large difference, which indicates that the number of iterations of RANSAC in the first stage is not enough to find the appropriate transformation matrix. The accuracy of MR-RPM is the lowest among the algorithms when there are a large number of error points, because it is based on global constraints, which is easy to be affected by error points. The accuracy and recall of our algorithm always remain above 0.8, which indicates that our strategy is robust. Our method can dig out the transformation matrix that is suitable for the most correct matches from the matches containing a large number of outlier points, and use it to find a more effective result. To



(a) (b) (c) (d)

Fig. 6. Visual results on five test data. Blue and red lines represent the inliers and outliers in the selected correspondence set by LPM, mTopkRP, RANSAC, and our method, respectively. (a) LPM. (b) mTopkRP. (c) RANSAC. (d) OUR.



(a) (b) (c)

Fig. 7. Quantitative comparisons of RANSAC, MR-RPM, LPM, mTopkRP, GLPM, only one stage, and the proposed method on the datasets. (Left to right) Initial inlier ratio, precision, and recall with respect to the cumulative distribution.

prove it, we compared the results of the algorithm with only the first step to the result of our method. The result is that through the second step, the precision and recall of our algorithm are improved. Running time is also one of the criteria for measuring algorithms. Through time complexity analysis before, our algorithm can be solved in linear time. Below we quantify the specific running time through experiments. The experimental data and the comparison algorithms are the same as the previous experiments. In order to highlight the time of our algorithm, we did not take the first stage separately for comparison. The experimental results are shown in Fig. 8. The experiment result shows that the RANSAC is a sampling-based algorithm, so it is the fastest. Because the run time of RANSAC is related to the number of iterations, its running time has little tendency to

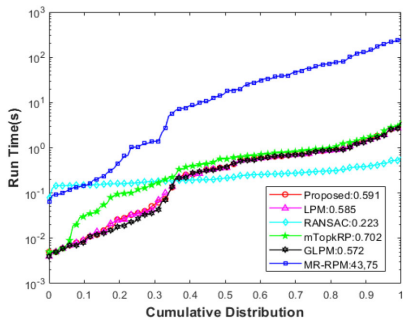
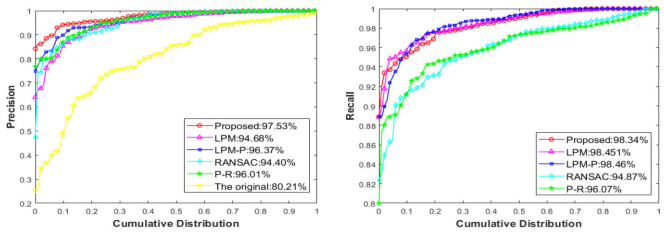


Fig. 8. Quantitative comparisons of RANSAC, MR-RPM, LPM, mTopkRP, GLPM, and the proposed method on the datasets; run time with respect to the cumulative distribution; the legend indicates the average running time of each algorithm.



(a) (b)

Fig. 9. Quantitative comparisons of RANSAC, our first step with RANSAC(P-R), LPM, LPM with global post-processing (LPM-P) and the proposed method on the datasets. (Left to right) Precision and recall with respect to the cumulative distribution; the legend indicates the average value of each algorithm.

change. Therefore, the running time is longer than other methods based on local information when the initial matching set is small. As the algorithm complexity shows, the running time of our algorithm is close to LPM, mTopKRP, and GLPM. The reason why the running time of our algorithm is slightly higher than that of LPM and GLPM is that the calculation of feature distance and global information is required in our algorithm. MR-RPM has the longest running time, mainly due to the slow convergence speed caused by noise.

3) Further Results: Our method combines global and local constraints, and they act on the first and second steps, respectively. They are not simple combinations. The first step and the second step are closely related, and replacing any one of them has an impact on the final result. The results of the first step provide a basis for the second step, so the results obtained by using local constraints in the first step must not only be accurate but also correctly reflect the transformation information. In theory, our second step can also improve the effects of other algorithms, which provides a possibility to improve the existing remote sensing image registration method. To verify them, we did further experiments. The experimental data has not changed. We replace the first and second steps of our method with other methods based on local information and global information, respectively, to compare with our method. Specifically, we replace our first step with LPM, and we call this method LPM-P. In addition, we replace our second step with RANSAC. We call this method P-R. The P represents the part we proposed. The experimental results are shown in Fig. 9. From Fig. 9, our first and second steps can work well. Compared with LPM and RANSAC, the accuracy and recall of LPM-P and P-R are improved. At the same time, the experimental results show that our first and second steps complement each other. When the recall rate is close to the highest, our results have the highest accuracy. Although the accuracy and recall rate of P-R are improved relative to RANSAC, they are not as good as our algorithm. The reason may be related to the randomness of the RANSAC. Although our first step reduced the number of false matches, the accuracy of the transformation matrix obtained by sampling is still not as good as the transformation matrix obtained by relying on all matches. Our method effectively combines local and global information. When the results of local constraints are not as good as other algorithms, the best precision and near-best recall can be achieved after combining global information.

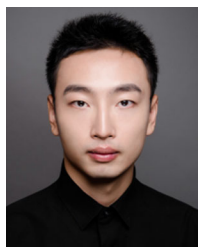
V. Conclusion

In this article, we have proposed an efficient mismatch removal method for remote sensing image registration, which is based on the stable local and global constraints between the two images of the same object or scene. The feature distance and the consistency of neighbors are introduced to measure the similarity of two local structures of an image pair. We use the above measurements to construct an optimization model and obtain its approximate solution in linear time. Then, the transformation matrix that is suitable for all the correct matches is calculated based on the approximate solution. Finally, we obtain a more effective result under the guidance of the transformation matrix. The qualitative and quantitative results on various remote sensing images have demonstrated the robustness and superior performance of our method for remote sensing image registration. The advantage of our article is that it fully combines local and global constraints, and the disadvantage is that the global transformation model is very important to the final result. Sometimes, affine transformation cannot handle some complex nonrigid transformation cases, so we have not found a perfect global transformation model to handle all registration tasks. It could be a problem for us to solve later.

References

[1] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1493, May 2007.
[2] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[3] M. Ehlers, "Multisensor image fusion techniques in remote sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 46, no. 1, pp. 19–30, 1991.
[4] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, Mar. 2005.
[5] P. R. Coppin and M. E. Bauer, "Digital change detection in forest ecosystems with remote sensing imagery," *Remote Sens. Rev.*, vol. 13, no. 3/4, pp. 207–234, 1996.
[6] H. Erives and G. J. Fitzgerald, "Automatic subpixel registration for a tunable hyperspectral imaging system," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 3, pp. 397–400, Jul. 2006.
[7] M. Vakalopoulou and K. Karantzalos, "Automatic descriptor-based co-registration of frame hyperspectral data," *Remote Sens.*, vol. 6, no. 4, pp. 3409–3426, 2014.
[8] Y. Ye, L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu, "Fast and robust matching for multimodal remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9059–9070, Nov. 2019.
[9] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.
[10] K. Yang, A. Pan, Y. Yang, S. Zhang, S. H. Ong, and H. Tang, "Remote sensing image registration using multiple image features," *Remote Sens.*, vol. 9, no. 6, p. 581, 2017.
[11] Y. Wu, W. Ma, Q. Su, S. Liu, and Y. Ge, "Remote sensing image registration based on local structural information and global constraint," *J. Appl. Remote Sens.*, vol. 13, no. 1, 2019, Art. no. 0 16518.
[12] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "Fusiongan: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, 2019.
[13] R. Xie, M. Xia, J. Yao, and L. Li, "Guided color consistency optimization for image mosaicking," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 43–59, 2018.
[14] J. Li, Q. Hu, M. Ai, and R. Zhong, "Robust feature matching via support-line voting and affine-invariant ratios," *ISPRS J. Photogrammetry Remote Sens.*, vol. 132, pp. 61–76, 2017.

[15] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, 2016.
[16] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to sift or surf." in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.
[17] Z. Fu, Q. Qin, B. Luo, H. Sun, and C. Wu, "HOMPC: A local feature descriptor based on the combination of magnitude and phase congruency information for multi-sensor remote sensing images," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1234.
[18] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A sift-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jun. 2014.
[19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[20] S. A. K. Tareen and Z. Saleem, "A comparative analysis of sift, surf, kaze, akaze, orb, and brisk," in *Proc. Int. Conf. Comput. Math. Eng. Technol.*, 2018, pp. 1–10.
[21] Y. Ke, R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. 506–513.
[22] A. Sedaghat and H. Ebadi, "Remote sensing image matching based on adaptive binning SIFT descriptor," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5283–5293, Oct. 2015.
[23] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.
[24] Y. Ye, J. Shan, S. Hao, L. Bruzzone, and Y. Qin, "A local phase based invariant feature for remote sensing image matching," *ISPRS J. Photogrammetry Remote Sens.*, vol. 142, pp. 205–221, 2018.
[25] R. Horaud, F. Forbes, M. Yguel, G. Dewaele, and J. Zhang, "Rigid and articulated point registration with expectation conditional maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 587–602, Mar. 2011.
[26] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, pp. 774–781.
[27] Q. Tang, J. Yang, W. Jia, X. He, Q. Zhang, and H. Liu, "A GMS-guided approach for 2D feature correspondence selection," *IEEE Access*, vol. 8, pp. 36 919–36 929, 2020.
[28] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6733–6741.
[29] X. Dai and S. Khorram, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2351–2362, Sep. 1999.
[30] S. Murala, R. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.
[31] Q. Li, G. Wang, J. Liu, and S. Chen, "Robust scale-invariant feature matching for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 287–291, Apr. 2009.
[32] A. A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, and I. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," *IEEE Trans. Image Process.*, vol. 12, no. 12, pp. 1495–1511, Dec. 2003.
[33] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2018.
[34] D. G. Lowe *et al.*, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.
[35] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.
[36] D. Barath, J. Matas, and J. Noskova, "MAGSAC: Marginalizing sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 197–10205.
[37] G. Wang, Z. Wang, Y. Chen, and W. Zhao, "A robust non-rigid point set registration method based on asymmetric Gaussian representation," *Comput. Vis. Image Understanding*, vol. 141, pp. 67–80, 2015.
[38] G. Wang, Q. Zhou, and Y. Chen, "Robust non-rigid point set registration using spatially constrained Gaussian fields," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1759–1769, Apr. 2017.
[39] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, 2015.
[40] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[41] P. H. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.
[42] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Joint Pattern Recognit. Symp.*, 2003, pp. 236–243.
[43] O. Chum and J. Matas, "Matching with prosac-progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 220–226.
[44] X. Li and Z. Hu, "Rejecting mismatches by correspondence function," *Int. J. Comput. Vis.*, vol. 89, no. 1, pp. 1–17, 2010.
[45] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.
[46] J. Ma, J. Jiang, C. Liu, and Y. Li, "Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration," *Inf. Sci.*, vol. 417, pp. 128–142, 2017.
[47] W.-Y. Lin *et al.*, "CODE: Coherence based decision boundaries for feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 34–47, Jan. 2018.
[48] W.-Y. D. Lin, M.-M. Cheng, J. Lu, H. Yang, M. N. Do, and P. Torr, "Bilateral functions for global motion modeling," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 341–356.
[49] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.
[50] J. Jiang, Q. Ma, T. Lu, Z. Wang, and J. Ma, "Feature matching based on top k rank similarity," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 2316–2320.
[51] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4181–4190.
[52] D. Chetverikov, D. Stepanov, and P. Krsek, "Robust euclidean alignment of 3d point sets: The trimmed iterative closest point algorithm," *Image Vis. Comput.*, vol. 23, no. 3, pp. 299–309, 2005.
[53] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understanding*, vol. 89, no. 2/3, pp. 114–141, 2003.
[54] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[55] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1482–1489.
[56] L. Babai, "Groups, graphs, algorithms: The graph isomorphism problem," in *Proc. ICM*, 2018, pp. 3303–3320.
[57] J. Yan, M. Cho, H. Zha, X. Yang, and S. M. Chu, "Multi-graph matching via affinity optimization with graduated consistency regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1228–1242, Jun. 2016.
[58] H. Liu and S. Yan, "Common visual pattern discovery via spatially coherent correspondences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 1609–1616.
[59] A. Zanfir and C. Sminchisescu, "Deep learning of graph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2684–2693.
[60] K. Adamczewski, Y. Suh, and K. Mu Lee, "Discrete tabu search for graph matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 109–117.
[61] Y. Xu, J. Ou, H. He, X. Zhang, and J. Mills, "Mosaicking of unmanned aerial vehicle imagery in the absence of camera poses," *Remote Sens.*, vol. 8, no. 3, p. 204, 2016.



Yue Wu received the B.Eng. and Ph.D. degrees from Xidian University, Xi'an, China, in 2011 and 2016, respectively.

Since 2016, he has been a Teacher with Xidian University. He is currently an Associate Professor with Xidian University. He has authored or co-authored more than 50 papers in refereed journals and proceedings. His research interests include computational intelligence and its applications.

Zhenglei Xiao received the B.E. degree from the Department of Software Engineering, Henan University of Technology, Zhengzhou, China, in 2019. He is currently working toward the master's degree with the School of Computer Science and Technology, Xidian University, Xi'an, China.

His research interests include image registration and point cloud registration.

Shaodi Liu was born in Gongyi, Henan, China, in 1995. He received the B.S. degree in intelligent science and technology from Xidian University, Xi'an, China, in 2017, where he is currently working toward the Ph.D. degree.

His research interests include image registration and point cloud registration.

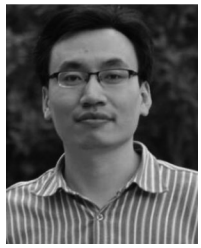
Qiguang Miao (Senior Member, IEEE) received the M.Eng. and Doctor degrees in computer science from Xidian University, Shaanxi, China.

He is currently a Professor with the School of Computer Science and Technology, Xidian University. His research interests include intelligent image processing and multiscale geometric representations for images.

Wenping Ma received the B.S. degree in computer science and technology and the Ph.D. degree in pattern recognition and intelligent systems from Xidian University, Xi'an, China, in 2003 and 2008, respectively.

Since 2006, she has been with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, where she is currently an Associate Professor. She has published more than 30 SCI papers in international academic journals, including the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *Information Sciences*, *Pattern Recognition*, *Applied Soft Computing*, *Knowledge-Based Systems*, *Physica A-Statistical Mechanics and its Applications*, and the *IEEE Geoscience and Remote Sensing Letters*. Her research interests include natural computing and intelligent image processing.

Dr. Ma is a member of the Chinese Institute of Electronics and the China Computer Federation.



Maoguo Gong (Senior Member, IEEE) received the B.S. degree (Hons.) in electronic engineering and the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003 and 2009, respectively.

Since 2006, he has been a Teacher with Xidian University. In 2008 and 2010, he was promoted as an Associate Professor and as a Full Professor, respectively. His research interests include computational intelligence with applications to optimization, learning, data mining, and image understanding.

Dr. Gong has received the support of Top-Notch Young Professionals for the prestigious National Program from the Central Organization Department of China, the Excellent Young Scientist Foundation Program from the National Natural Science Foundation of China, and the New Century Excellent Talent in University from the Ministry of Education of China. He is an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Fei Xie received B.S. degree in school of computer science and technology from Xianynag Normal Univesity, in 2007, master's degree in computer application technology from the University of Science and Technology of China in 2009 and Ph.D degree in the Computer application from Northwestern University, Xi'an, China, in 2019. He is an Researcher with School of AOAIR, Xidian University, Xi'an, China. His research interests include image processing and pattern recognition.

Yang Zhang was born in 1987. He received the Ph.D. degree in microelectronics and solid-state electronics from the University of Science and Technology of China, Hefei, China, in 2016.

Currently, he is with Shanghai Aerospace Electronic Technology Institute, Shanghai, China. He has authored or coauthored more than 10 high-level papers in related fields. His research interests include new system radar design and artificial intelligence research.