Registration of UAV Images Using Improved Structural Shape Similarity Based on Mathematical Morphology and Phase Congruency

Jian Jin and Ming Hao

***Abstract*—Image registration is a basic step for remote sensing image processing, such as classification, change detection, and so on. Registration of unmanned aerial vehicle (UAV) images is a challenging task because of illumination change, obvious rotation, viewpoint change, and similar textures. The change in viewing angle between adjacent images causes significant difficulties in image matching, especially for residential areas and high-rise buildings. Therefore, the feature points on the buildings are removed by morphological processing when detecting feature points in this study. Considering the similarity of terrain structural properties between adjacent UAV images, the scene shape similarity feature (SSSF) descriptor is utilized to capture the structural similarity. However, the SSSF descriptor is influenced by the illumination changes of UAV images. The phase congruency algorithm is adopted to replace the Canny operator to detect structural features when calculating the SSSF descriptors because of its robustness for illumination change. Then, the normalized correlation coefficient of SSSF descriptors is used as the similarity metric to detect tie points. Finally, the point pairs with large residuals are eliminated by random sample consensus algorithm and the tie points are detected by iteratively removing the mismatched points. The proposed method was compared to some recent related methods using correct matching ratio, root-mean-square error, and time. Five experiments including habitation, bare land, mixed terrain, farmland, and buildings are carried out to evaluate the effectiveness of the proposed method. Registration results demonstrate that the proposed method is robust for illumination and viewpoint change of UAV images and usually generates more accurate registration results than some popular registration methods.**

***Index Terms*—Mathematical morphology, phase congruency, remote sensing, unmanned aerial vehicle (UAV) image registration.**

I. Introduction

UNMANNED aerial vehicle (UAV) photogrammetry has been widely used due to its advantages of simple operation, low cost and high efficiency [1]. Hence, it has been widely used in the fields of large-scale topographic map mapping [2], natural disaster emergency support [3], and urban surveying and mapping [4]. As the fundamental work of the above applications, image registration aligns images obtained at different time phases, from different viewpoints, and by different sensors [5].

In order to deal with the problems appeared in the process of image registration, a large number of matching methods have been proposed, and these methods can be mostly classified into two categories: grayscale-based and feature-based methods.

The grayscale-based methods match images by computing the maximum similarity of intensity values based on the image grayscale space. The typical similarity measures include approaches using mutual information [6], cross correlation [7], and phase correlation [8]. Through the research and analysis of the existing image matching algorithm, Pratt [9] explained the various similarity measure functions in the image matching algorithm. Evangelidis and Psarakis [10] proposed a new similarity metric, the application of an enhanced correlation coefficient (ECC), which is invariant to photometric distortions. Zhou *et al.* [11] utilized the scale-invariant feature transform (SIFT) based combined with a gradient-based method to obtain coarse estimation and refine results. Uss *et al.* [12] proposed a novel accuracy bound derived from the Cramer–Rao lower bound for analyzing image registration. The grayscale-based methods are easy to be carried out and can get great matching accuracy. However, these methods are not suitable for real-time image registration because of its computation complexity. Moreover, illumination change between multitemporal images causes serious difficulty for image registration.

The feature-based methods are feature matching between multitemporal images. Such features commonly include point feature, structural feature, area feature, and so on. Kahaki *et al.* [13] proposed a contour-based corner detection method based on mean projection transform and parabolic fit approximation, which shows great repeatability, localization, and accuracy of repeatability for the detected points. SIFT proposed by Lowe [14] is widely used for matching images due to its immutability for image scaling or rotation. Then, a family of matching methods based on SIFT were proposed such as CSIFT [15], PCA-SIFT [16], and speed-up robust features (SURF) [17]. The ASIFT algorithm [18] analyzes the images with a different viewpoint, which is robust for viewpoint change but time-consuming. Li and Ye [19] proposed a robust sample consensus judging

Manuscript received December 10, 2019; revised February 11, 2020 and February 26, 2020; accepted March 20, 2020. Date of publication April 6, 2020; date of current version April 27, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41701504, in part by the Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation under Grant 2018NGCMZD01, in part by the Open Research Fund of Jiangsu Key Laboratory of Resources and Environmental Information Engineering, CUMT under Grant JS201704, and in part by a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions. *(Corresponding author: Ming Hao.)*

The authors are with the Jiangsu Key Laboratory of Resource and Environment Information Engineering, China University of Mining and Technology, Xuzhou 221116, China (e-mail: 1715882901@qq.com; cumthaoming@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.2982929

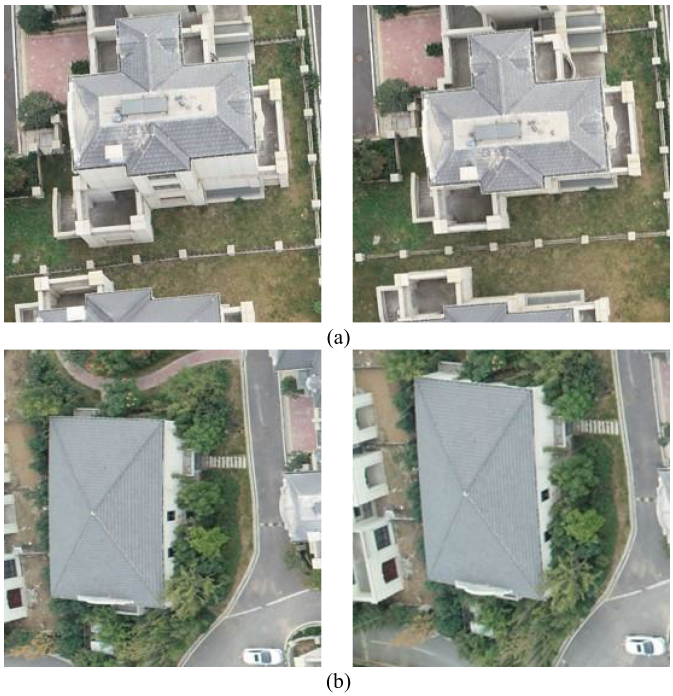


Fig. 1. Differences of buildings caused by the change of viewpoint. (a) Differences of high-rise buildings. (b) Deformation of buildings caused by the change of view angle.

algorithm defined by the triangle-area representation combined with hypothesis and testing. Recently, shape features have been applied to define similarity descriptors for image registration and achieved great matching accuracy [20]. Ye *et al.* [21] proposed a shape descriptor named dense local self-similarity (DLSS) to describe the shape similarity between images. Spatial orientation feature matching, a new similarity metric, is defined by triplewise signal eigenvector correlation. This method supports various types of transformation in the original image, such as scale, translation, rotation, intensity noises and occlusion [22]. Xie *et al.* [23] used the extended phase correlation based on Log-Gabor filtering to tackle nonlinear radiometric and large-scale differences between images. Hao *et al.* [24] utilized the shape context algorithm to build a scene shape similarity feature (SSSF) descriptor, which integrates the similar structural information between images greatly. This method captures the structural properties between images accurately and efficiently. Most of the methods mentioned above are not robust for matching the images with a large viewpoint and illumination change.

As shown in Fig. 1, when performing aerial photography work in residential areas and high-rise buildings, the change in viewing angle between adjacent images causes significant difficulties in image matching. Hence, the viewpoint and illumination differences are still problems for UAV image registration.

In this study, in order to solve the problem of image registration stability of UAV images, the feature points extracted by SURF on the buildings are removed by morphological processing [25], [26] when detecting feature points. As for the illumination change of UAV images, the phase congruency algorithm [27] replaces the Canny operator to detect structural features when calculating the SSSF descriptors. Then, the normalized correlation coefficient (NCC) of SSSF descriptors is used as the similarity metric to detect tie points. The mapping relationship H is estimated at the preregistration stage. Subsequently, the template matching strategy and the mapping relationship H are used to detect matching points at the fine registration stage. Finally, the point pairs with large residuals are eliminated by random sample consensus (RANSAC) [28] algorithm and the tie points are detected by iteratively removing the mismatched points.

The rest of this article is organized as follows. Section II gives a detailed description of the methodology. Section III analyzes the matching accuracy of the proposed method and compared the proposed method with other popular registration methods. Finally, Section IV concludes this article.

II. Methods

Fig. 1 shows that when performing aerial photography work in residential areas and high-rise buildings, the viewpoint change causes significant differences in the geometric structure of buildings. Therefore, the feature points on the buildings are not reliable for image registration. When performing aerial photography work in residential areas and high-rise buildings, the change in viewing angle between adjacent images causes significant difficulties in image matching. Therefore, the morphological processing is used to remove the feature points on the buildings, and the SSSF descriptors of the remaining points are then calculated to detect tie points. The phase consistency algorithm has been used in the registration of remote sensing images due to its good illumination invariance. As a result, instead of the Canny operator, the phase consistency is used to extract the edge features of the template window when calculating the SSSF descriptor.

A. Building Extraction Using Morphology

In mathematical morphology, binary images are treated as a collection. Structural elements are the basic research objects, allowing structural elements to slide on binary images, and the purpose of image segmentation is achieved through the intersection and parallel operations. In this study, morphological processing is used to extract the building area from the UAV images, and feature points located on buildings detected by the SURF algorithm are removed.

1) Erosion and Dilation: Erosion and dilation [29], [30] are the most basic forms of computation in binary morphology. In binary morphology, the "probe" (the structural element that determines the shape) is used to detect the image, and find the area where the probe can be stored by performing a sounding operation on each set of regions in the image. The structural element is significant for erosion and dilation. The essence of structural elements is a 3×3 matrix. Specifically, for a shape, when all the structural elements are in the shape, the center of the structure element forms the eroded shape [see Fig. 2(a)], whereas when only the center of the pointer is needed to lie in the shape, the child elements of the structure element form the

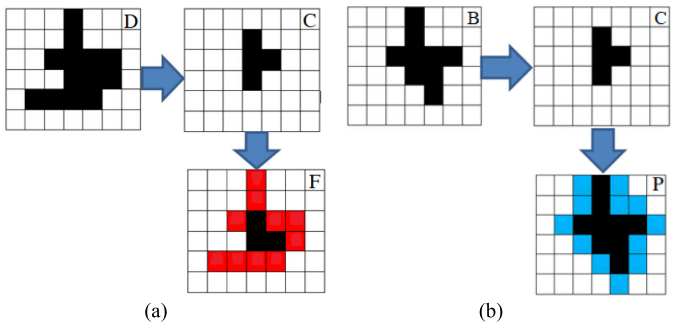


Fig. 2. Erosion and dilation schematic diagram. (a) Erosion. (b) Dilation.

structure of the dilation part [see Fig. 2(b)]

$$F = D\Theta C = \{(x, y)\,|C_{x,y} \subseteq D\} \tag{1}$$

$$P = B \oplus C = \left\{(x, y)\,|\,\left[C_{(x,y)} \cap B\right] \neq \emptyset\right\} \tag{2}$$

where D and B are the sensed graphics, C is the structural element, and coordinate (x, y) is the core of "probe" C.

Formula (1) and Fig. 2(a) are the erosion formula and erosion diagram, respectively; moreover, the "red grids" are the corroded pixels. The image F obtained by the etching refers to a set of central position points of the structural element C when the structural elements C are all placed in the target area D. Similar to this, the basic idea of dilation is that when the target area B has the central element of structural element C, the child elements of structural element C are regarded as the expanded elements. Therefore, the expanded area ("blue grids") is a collection of all the child points of structural element.

2) Building Area Extraction: The binary open operation is to perform the etching operation on the image first and then the dilation operation, and smooth the edge of the object without changing the contour of the object, and eliminate the small noise points connected thereto. The binary close operation is to first perform a dilation operation on the image, and then perform an etching operation on the image to eliminate the "holes" existing in the area. Inspired by this, open and close operations based on erosion and dilation can be used to help us get a complete and nonredundant range of buildings from UAV images. Fig. 3 shows the specific steps given as follows.

1) Converting the UAV image into a grayscale image and counting the grayscale value range of the building area.
2) Separating the building by multi-threshold segmentation according to the grayscale value of the building.
3) Removing noise points existing on the segmented image.
4) Use the morphological algorithm to remove the edge "burr" of the image and make up the "black spots" in the building.
5) Get the scope of the building. After the feature points are detected by SURF, the feature points in the range will be eliminated.

Some prior knowledge is needed before the use of morphological processing. The grayscale histogram needs to be calculated first. Therefore, the grayscale value range of the building is clear. If the size of the shape is fewer than the structural element, it will be regarded as the noise and will be removed. Therefore, the designed structure is suitable for buildings with different sizes. The remaining steps are as shown above.

B. Improvement of SSSF Descriptor

The SSSF descriptor is influenced by the illumination changes of UAV images. The phase congruency algorithm is adopted to replace the Canny operator to detect structural features when calculating the SSSF descriptors because of its robustness for illumination change.

1) SSSF Descriptor: Compared with satellite remote sensing images, UAV images have the advantages of high resolution, rich texture information, and obvious terrain features. In order to make full use of the strong similarity between contours in adjacent images, the SSSF descriptor [24], [31] is calculated to describe the terrain structural information between the master and slave images. The SSSF descriptor is a feature-based image registration method computed by scene shape information and shape context algorithm. This method first extends a template window that centers on a feature point, and then the edges of the template window are extracted by the Canny operator. Then, the shape context at the feature points is calculated for defining the SSSF descriptor. Fig. 4 shows the similarity of the descriptor of the tie points in the same template window. Therefore, the NCC of the SSSF descriptor between master and slave images (abbreviated as $\mathrm{SSSF}_{\mathrm{ncc}}$) is used as the similarity measure for multisource image registration

$$\mathrm{SSSF}_{\mathrm{ncc}} = \frac{\sum_{k=1}^{n}\left(h_p(k) - \overline{h_p(k)}\right)\left(h_q(k) - \overline{h_q(k)}\right)}{\sqrt{\sum_{k=1}^{n}\left(h_p(k) - \overline{h_p(k)}\right)^2 \sum_{k=1}^{n}\left(h_q(k) - \overline{h_q(k)}\right)^2}} \tag{3}$$

where p and q are the feature points of master and slave images, respectively, $h_p(k)$ and $h_q(k)$ are the shape context information of the points p and q within a template window, respectively, and $\overline{h_p(k)}$ and $\overline{h_q(k)}$ represent the means of SSSF descriptor within the template window.

Furthermore, a template registration method is applied to extract the tie points between the master and slave images. The advantages of SSSF descriptor is that it makes good use of similar terrain structural features between master and slave images, which is suitable for UAV image registration including large similar terrain edges feature very well. Therefore, this study uses $\mathrm{SSSF}_{\mathrm{ncc}}$ as the similarity measure to match tie points between master and slave images.

2) Improved Structural Information Using Phase Congruency: As mentioned above, the Canny operator is used to extract the edge features of the template window when calculating the SSSF descriptor. However, there is often an exposure difference between adjacent UAV images and the Canny operator is sensitive to illumination changes. In this study, the phase congruency is introduced to produce the terrain structural information of the image.

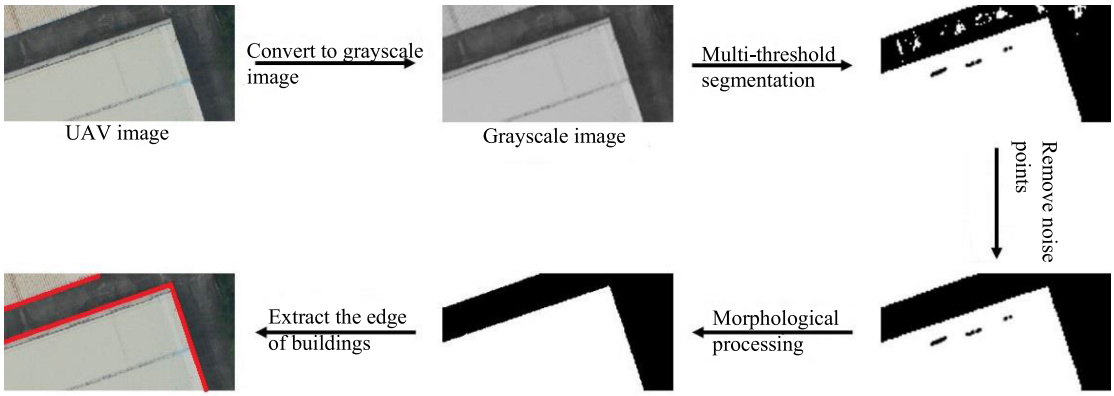


Fig. 3. Mathematical morphology for building extraction.

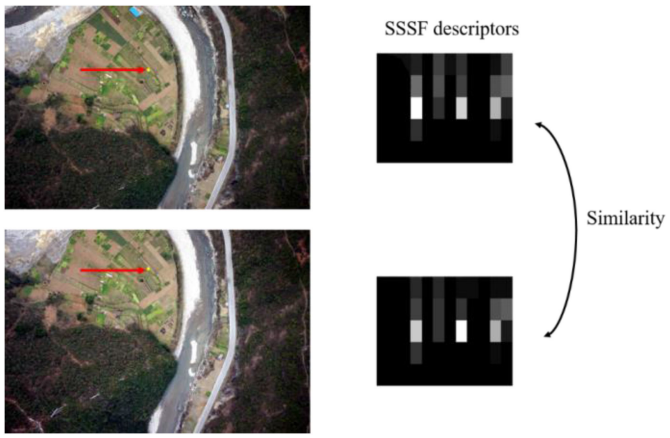


Fig. 4. Similarity of SSSF descriptor.

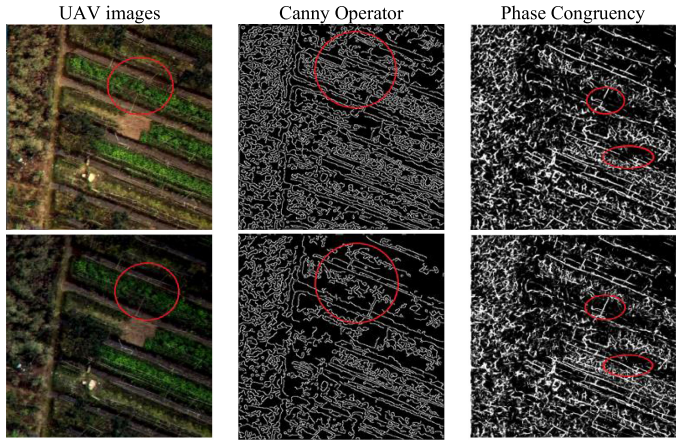


Fig. 5. Comparison between Canny and phase congruency for extracting edges feature.

Phase congruency is a dimensionless quantity, which reflects the phase feature information of image. It has good local illumination and contrast invariance. The two-dimensional phase congruency transformation function [32] is

$$\mathrm{PC}(x,y) = \frac{\sum_o \sum_n W_o(x,y)\,\lfloor A_{no}(x,y)\Delta\Phi_{no}(x,y) - T_o \rfloor}{\sum_o \sum_n A_{no}(x,y) + \varepsilon} \quad (4)$$

where (x, y) is the image coordinate, $W_o(x,y)$ is the weighting factor of frequency expansion, $A_{no}(x,y)$ and $\Phi_{no}(x,y)$ are the amplitude and phase of point (x, y) in the given filter n and o directions, $\lfloor\ \rfloor$ represents that if the value is positive, it is unchanging; otherwise, the value equals to 0, and ε is a constant that avoids division by zero.

The phase congruency is free from the influence of the local light and dark changes of the image, and includes the angle, line, and texture information [33]. Especially when the image edge contrast is low, the phase congruency gets more robust results. As shown in Fig. 5, when there is an illumination difference between adjacent UAV images, the phase congruency can still obtain relatively complete structural features, whereas the structural features obtained by the Canny operator are quite different. Therefore, the phase congruency is applied to extract the edges feature due to its stability for illumination difference.

3) Tie Points Matching: In order to detect more robust feature points, when using the SURF algorithm to extract feature points, the SURF feature points on the building are removed in the master images. In the preregistration stage, the SURF algorithm is applied to achieve the mapping relation between the master and slave images by a few feature points. In fine matching stage, the match points are found by the template matching strategy. Finally, the tie points between the master and slave images are achieved after the mismatches are eliminated by RANSAC algorithm. The detailed matching process of this study is as follows.

1) *Feature points detection:* Due to the influence of the imaging mode (central projection) of the UAV images, the edge distortion of the UAV image is the largest. Moreover, the change in viewpoint can lead to a significant deformation of buildings in adjacent UAV images. Therefore, in order to obtain the feature points with high stability, the feature points on the edge of the image and the building are removed when extracting the feature points. The main process includes the following two steps.

Fig. 6. Removal of feature points located in buildings.

Fig. 7. Distribution of feature points.

Step 1: Extracting feature points and removing image structural feature points by using SURF operator with good stability and efficiency.

Step 2: Morphological treatment extracts the scope of the building and removes feature points that fall within buildings (as shown in Fig. 6).

2) *Preregistration stage:* In this stage, as shown in Fig. 7, the mapping relationship *H* between the master and slave images is obtained by the SURF algorithm using few uniformly distributed feature points, preparing for the template fine matching phase

$$\begin{bmatrix} x \\ y \\ l \end{bmatrix} = H \cdot \begin{bmatrix} x' \\ y' \\ l \end{bmatrix} \tag{5}$$

where (x, y) and (x', y') are a pair of matching points and *H* is the transformation relationship between the two images.

3) *Fine registration stage:* For one feature point (x, y) in the master image, the SSSF descriptor is first calculated under the template widows with different size (ranging from 20×20 to 100×100 pixels). Then, the correspondence point (x', y') in the slave image is calculated by the transformation relationship *H* obtained from the preregistration stage. A search window of size 10×10 pixels centering on (x', y') is extended from the slave image for matching feature points. The NCC is used as a similarity measure between descriptors to search matches within the search window, and the point with the highest similarity in the search window is regarded as the final tie point.

4) *Elimination of mismatched tie points:* The above resultant tie points are not completely correct because of the uncertainty factors such as occlusion. The RANSAC algorithm is used to eliminate the matched point pairs with a large residual error. Subsequently, the transform relationship between matches is described by projective transform, which is set up using the least-squares method with all tie points. Then, the residuals and the root-mean-square errors (RMSEs) of tie points are calculated. If the RMSE is larger than the given threshold (e.g., 1.2 pixels), the point pair with the largest residual is eliminated. The above steps are repeated until RMSE meets the requirements.

III. Experiments Results and Evaluations

Four sets of UAV images have been selected to verify the effectiveness of the proposed method. The experiments mainly have two objectives: 1) analyze the effect of the template window on the proposed method, and 2) compare with other algorithms such as SURF, ECC, and DLSS. The details of test sets, implementation, and experimental analysis are as follows.

A. Details of Test Set

The image pair of datasets are shown in Fig. 8, where each pair was composed of a master and a slave image obtained from adjacent UAV images. These UAV images are a set of data from the same area taken under the same conditions. They have the same resolution of 0.2 m and a size of 5456×3632 pixels. The differences between each set are as follows.

1) *Habitation:* This image pair is located in a residential area. There is a large area of cultivated land in the center of images. Some buildings are located in the upper left of the master image. When extracting the feature points, the points on the buildings should be removed. Generally speaking, habitation has rich terrain feature, and no obvious local deformation exists besides more buildings [see Fig. 8(a)].

2) *Bare land:* As shown in Fig. 8(b), a large area of bare land exists in the master image, which may lead to the difficulty of extracting feature points and computing SSSF descriptor. Although the road on the left side of the images may provide some structural features, the rotation and deformation differences caused by the change in viewpoint are unavoidably appeared between master and slave image.

3) *Mixed terrain:* The characteristic of this set is the rich terrain structure such as flat cultivated land, river, road, mountain, and woodland. Also because of this, there is obvious local deformation in mountain and woodland.

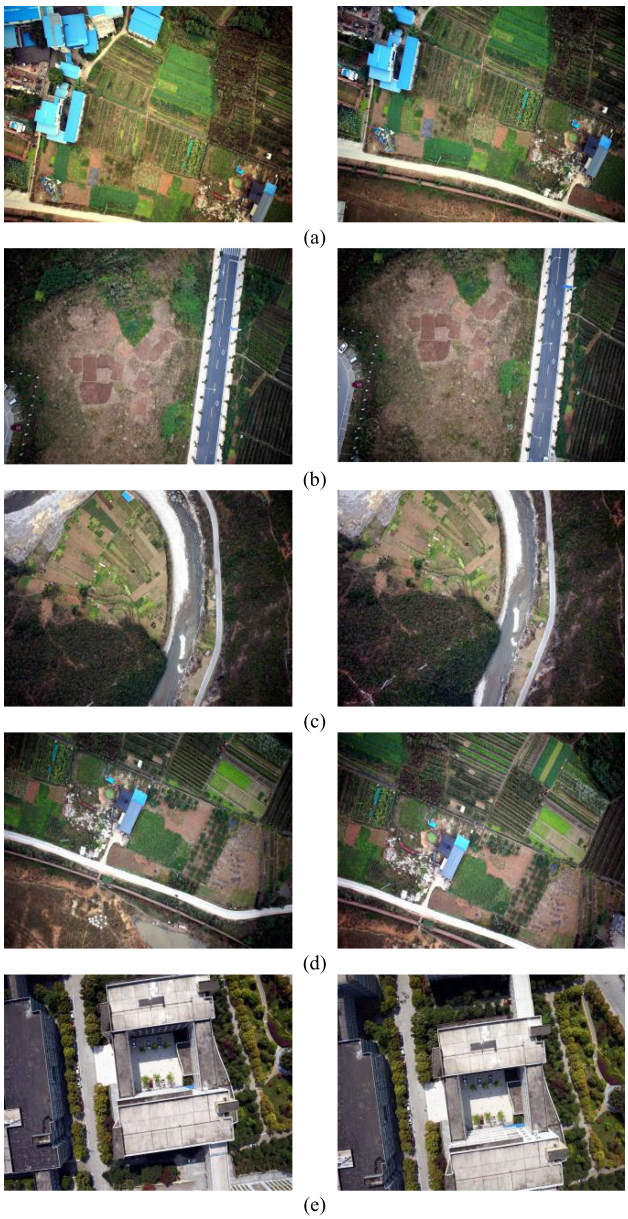


Fig. 8. Image pairs of experiments; the left one is master image and the right one is slave image. (a) Habitation. (b) Bare land. (c) Mixed terrain. (d) Farmland. (e) Buildings.

4) *Farmland:* Similar to dataset 1, the image pair also has a large area of cultivated land. The difference is that there are more terrain features in the master such as roads and lakes, whereas there are also multiple pieces of bare land, which may affect the accuracy of the proposed method.
5) *Buildings:* The characteristic of this set is the large area buildings, which leads to the few feature points. Some roads and plants will become the main features for the UAV image registration.

B. Registration Process and Assessment Criteria

In the experiments, the transform relationship H was calculated by SURF algorithm using about 15 evenly distributed feature points in preregistration stage. The SURF algorithm was first used to extract the feature points on the master image. In order to obtain more robust feature points, the points lied in the range of buildings and the surrounding of images were removed by morphological processing. Then, in order to analyze the influence of the proposed method with regard to the change of template window size, a template window centering on the feature points with different sizes (ranging from 20×20 to 100×100 pixels) was extended to calculate the SSSF descriptor. Before that, the phase congruency algorithm was first used to detect the structural features over the template window. In fine registration stage, the correspondence feature points in the slave images were first calculated by the transform relationship H obtained in preregistration stage. Subsequently, a search window of 10×10 pixels was applied to find the points with the highest similarity as the matching points. Finally, the RANSAC algorithm was used to eliminate the points with large residuals.

RMSE, computation time, and correct matching ratio (CMR) are used to evaluate the effectiveness of the proposed method. The CMR is computed as $\mathrm{CMR} = C/T$, where C is the number of correctly matched point pairs and T is the total number of match point pair. After the process of registration, every feature point finds its corresponding point in the slave image. Therefore, the residual error of each point pair can be calculated by the projective transformation and the point pairs with large residuals (threshold is 1.2 pixels) were removed using RANSAC algorithm. Then, the point pairs with larger residuals will be iteratively removed until the RMSE is smaller than 1.0 pixel. After the mismatched point pairs were eliminated, the tie points between UAV images are detected. Since the matching point pairs are too much, only one-fifth of all tie points are shown in Figs. 9 and 11, where these points are evenly distributed. Fig. 9 shows the results of tie points registration, and Fig. 10 shows the red–green registration result.

C. Accuracy Analysis

The matching results of the four datasets are shown in Fig. 9. From a global perspective, the tie points between the master and slave images are matched very well regardless of the obvious deformation and viewpoint change. Therefore, in order to quantitatively analyze the registration accuracy, 40 feature points were manually selected to calculate RMSEs of the four datasets. The accuracy of the proposed method is shown in Table I.

As shown in Table I, the size of the template window has slight influence on the RMSEs of the experiments and all datasets achieve good registration accuracy. Especially the best accuracies are obtained at 60 or 70 template window size. These experiments indicate that the proposed method is robust for the illumination change, obvious rotation, viewpoint change between UAV images.

The effect of template window size on the registration accuracy was analyzed. Obviously, the smaller the template window, the more difficult it is to extract accurate terrain structural features. The calculation of SSSF descriptors are directly affected by the less structural features. On the other hand, due to the center projection of UAV images, the deformation and local distortion of the edge of the template window are more obvious

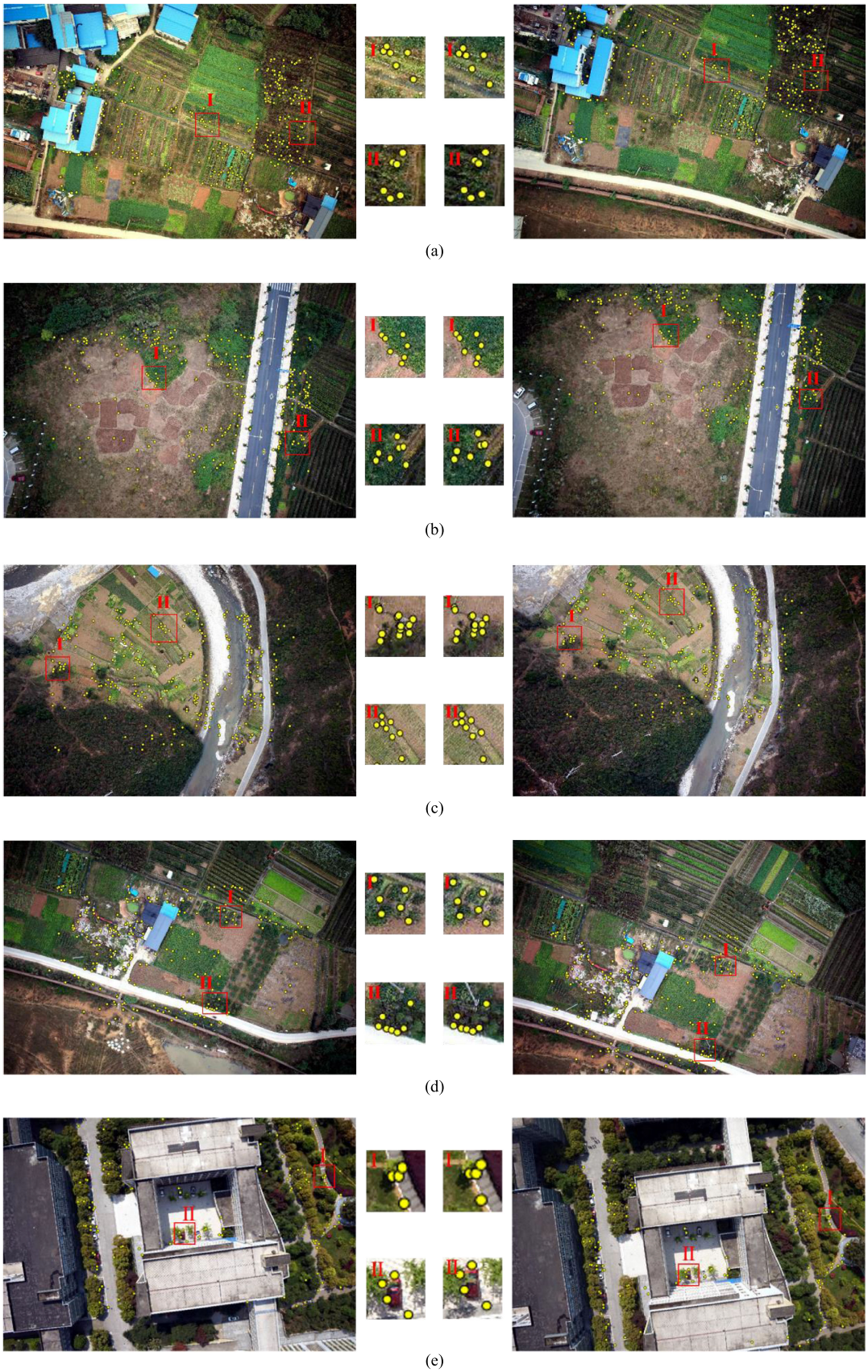


Fig. 9. Distribution of tie points. (a) Habitation. (b) Bare land. (c) Mixed terrain. (d) Farmland. (e) Buildings.

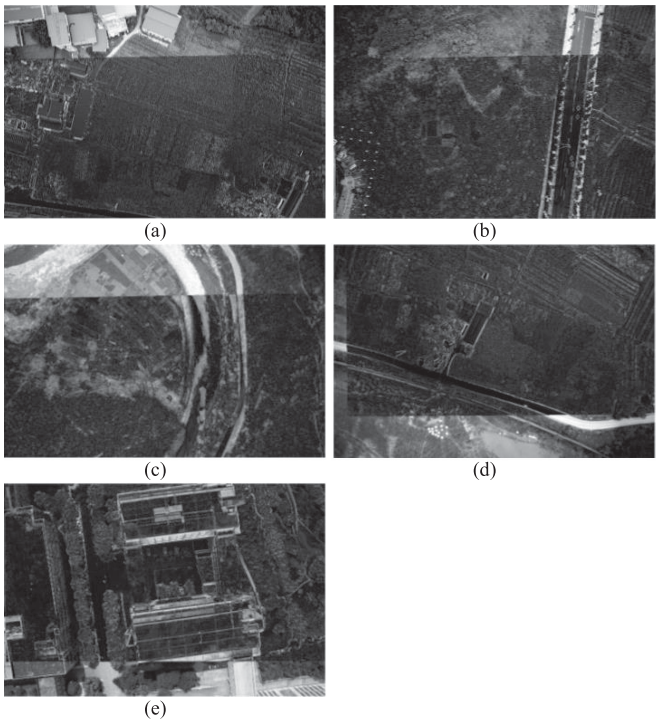


Fig. 10. Graph of difference between master and matched images. Grayscale is used to represent the difference between the *R*-band of master image and the *G*-band of registration image. (a) Habitation. (b) Bare land. (c) Mixed terrain. (d) Farmland. (e) Buildings.

across a larger template window compared to the image center (feature points), which leads to the drop of the similarity of structural features. Therefore, when the size of the template window continuously increase/decrease until the similarity of terrain features is the highest, the experiments get the best accuracies.

The accuracies of each dataset have some differences. Categories 1, 3, 4, and 5 get the best registration accuracy when the template size is 60×60 pixels, whereas the "greatest" template window of bare land is 70×70 pixels. Moreover, the matching accuracies of bare land are lower than the other sets. The reasons for these differences are as follows. The suitable template window of bare land has such difference with other tests that the bare land contains less terrain features due to the large range of bare land. Therefore, a larger template window is needed to collect enough structural features to calculate SSSF descriptor. However, larger template window also results in the matching accuracy of bare land lower than the other sets.

The matching accuracies of the bare land and mixed terrain are lower than the habitation and farmland. On the one hand, the multiple bare land indeed affects the matching accuracy of bare land. On the other hand, bare land achieves the lowest matching accuracy due to the deformation of terrain and less structural features between two images. In addition, although there are fewer feature points used for image registration, the matching accuracies of buildings are always well. However, they are lower than the other dataset. The deformation of buildings results in the decrease of SSSF descriptor similarity. Therefore, large area buildings literally affect the effectiveness of the proposed method. Nevertheless, the all RMSEs of bare land are lower than 1 pixel. In summary, the proposed method achieves good matching accuracy. The matching results of Figs. 9–11 and Table III are obtained based on the most suitable template window size. As shown in Fig. 9 and Table III, the proposed method has been successfully used to match UAV images. The CMR and the number of tie points are affected by terrain structural features and the CMRs of all tests are more than 90%, which also proves the stability of the proposed method.

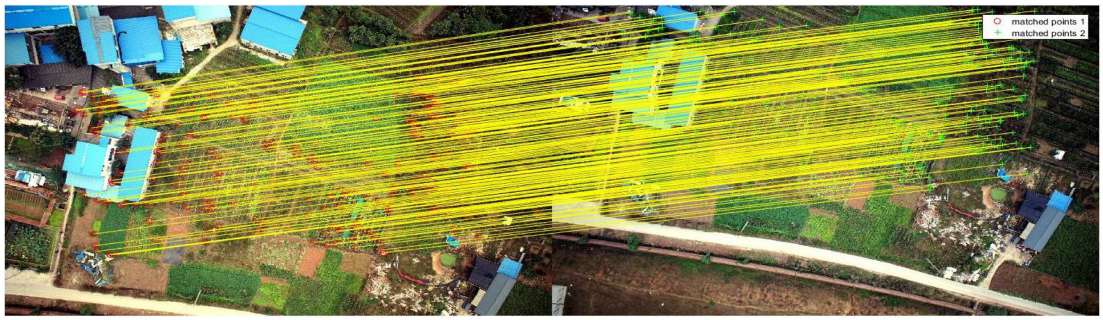
Moreover, the variable-controlling approach is used to test the effectiveness of morphological processing and phase congruency, respectively. The experimental results are shown in Table II. Experimental results show that each improvement is effective for the proposed method. Moreover, compare with SSSF descriptor, N1 and N2 all get better registration accuracy.

In summary, the experiments evaluate that the accuracies of the proposed method with different template window size are always stable. CMRs, RMSEs, and the number of tie points quantitatively certificate the proposed method's robustness for the viewpoint and illumination change between UAV images.

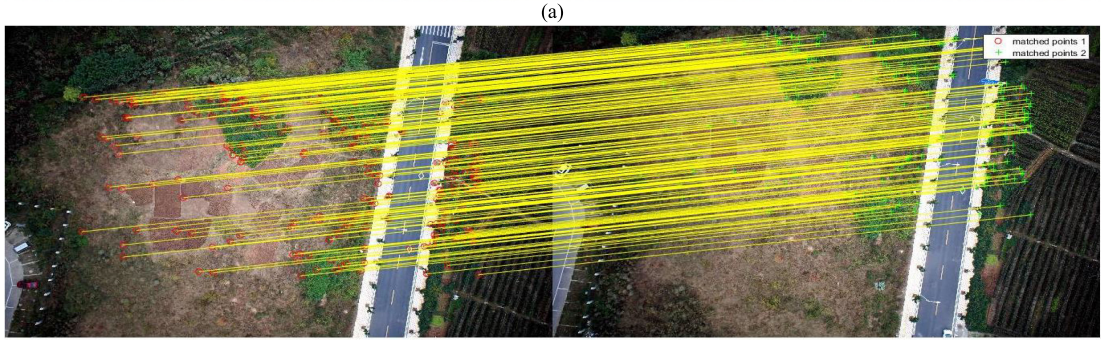
D. Comparison With Other Methods

The proposed method was compared with SURF, ECC, DLSS, and SSSF to verify its effectiveness for UAV image registration. ECC algorithm is a modified version of the correlation coefficient for image registration, which has the desirable characteristic of being invariant to photometric distortions. Considering that the resulting similarity metric is a nonlinear function, the forward additive approach and the inverse compositional methods are implemented for its maximization. ECC algorithm is robust to noise and photometric distortions. The second comparison algorithm DLSS is a shape descriptor defined by self-similarities over images. Then, the NCC of the DLSS descriptors is used as the similarity metric to detect tie points between images. DLSS algorithm is robust against nonlinear intensity differences because of its shape similarity. Since ECC and DLSS have some similarities with the proposed method, it was compared with ECC and DLSS to evaluate the performance of the proposed method. The SSSF descriptor has been described in Section II. The comparison results are shown in Table III.

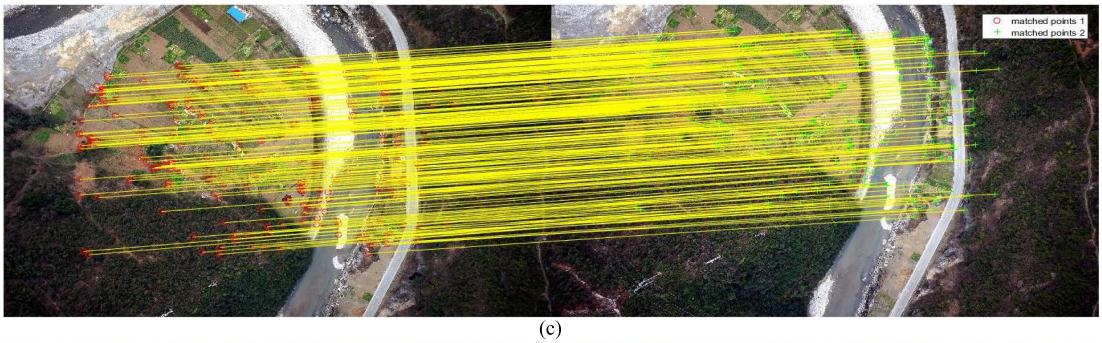
The comparison experiments were implemented to five image pairs. The same datasets had the same number of feature points for all methods used in this study. It is known that the spatial resolution of the UAV images is very high. Tens of thousands of feature points can be extracted by the SURF algorithm. It is redundant for image registration. Therefore, the detected feature points are diluted evenly. As the input of each algorithm, it is used for image registration. The difference is that the feature points of the proposed method are processed by the mathematical morphology. This reduces the influence of different feature point extraction methods on the comparison of experimental results. In order to ensure the fairness of the comparative experiment, the most appropriate parameters were set to the other four methods to analyze the matching performance. In this study, multiple sets of experiments show that the parameter σ of ECC was set to 8. Besides, the template window size of DLSS was set to



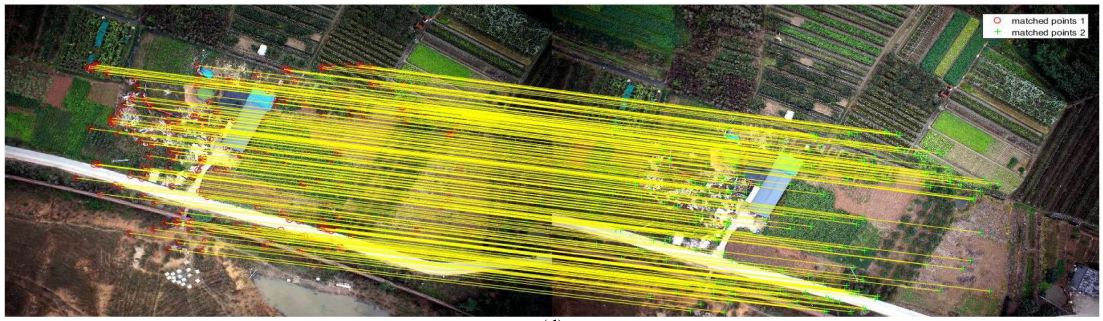
(a)



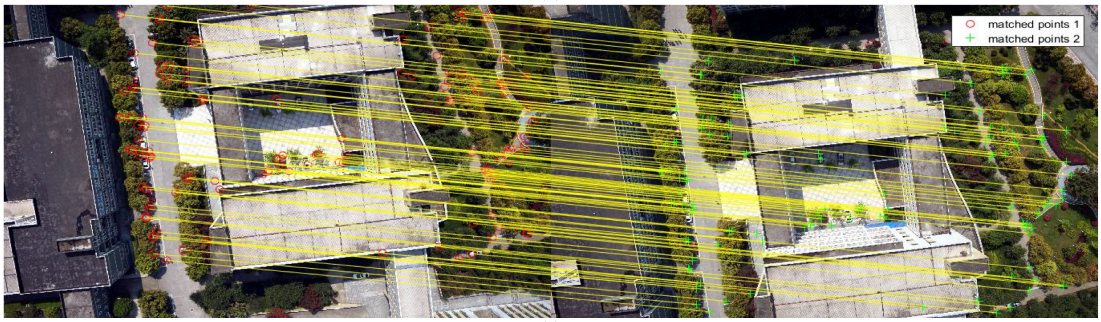
(b)



(c)



(d)



(e)

Fig. 11. Tie points registration results. (a) Farmland 1. (b) Bare land. (c) Mixed terrain. (d) Farmland 4. (e) Buildings.

TABLE I
RMSEs OF THE DATASETS WITH DIFFERENT TEMPLATE WINDOW SIZE

Category	The size of template window (pixels)								
	20	30	40	50	60	70	80	90	100
Habitation	0.6492	0.6401	0.6240	0.5617	**0.5292**	0.5382	0.5353	0.5481	0.5413
Bare land	0.7880	0.7756	0.7654	0.7670	0.7698	**0.7556**	0.7733	0.7785	0.7754
Mixed terrain	0.7779	0.7844	0.7750	0.7542	**0.7461**	0.7622	0.7849	0.7569	0.8250
Farmland	0.5895	0.6340	0.6440	0.5847	**0.5583**	0.5858	0.6054	0.5813	0.5827
Buildings	0.8413	0.8302	0.8261	0.8279	**0.8224**	0.8237	0.8391	0.8342	0.8420

100×100 pixels and the template window size and Canny operator of SSSF was set to 15×15 pixels and 0.2, respectively. After the mismatched points were eliminated, the matching criteria, including CMR, RMSE, and time, were calculated.

As shown in Fig. 11 and Table III, the proposed method gets the best matching accuracies compared with the other methods.

The SURF algorithm relies too much on the gradient direction of the local area pixels, which restricts matching accuracies. The ECC is invariant to photometric distortions, whereas is sensitive to nonlinear deformation and illumination change over UAV images. Therefore, its matching accuracies are deeply affected. The DLSS also achieves better matching performances. However, the registration accuracies of DLSS are lower than the proposed method. The reasons are that DLSS is built by collecting the self-similarities from all pixel not just shape properties, and is easily interfered by external factors such as occlusion, shadows. The distortions and rotations of buildings caused by viewpoint change between adjacent UAV images affect the registration accuracy of SSSF algorithm. That evaluates the effectiveness of the proposed method.

Additionally, the time cost of all methods is taken into consideration. From Fig. 12, it can be seen that the proposed method needs more computation time than the SSSF method. The reason is that the proposed method is defined by combing SSSF with morphological processing. The computational complexity of the SURF algorithm is high and the "1-to-*N*" matching approach is very time consuming. ECC algorithm is an area-based subpixel matching method and directly uses the intensity values across the image, which needs much time. Besides, the matching strategy of ECC is very time-consuming due to the combination of forward additive approach and the inverse compositional methods. Therefore, the speed of ECC is slower than the proposed method and DLSS. DLSS is defined by collecting the self-similarities of all pixels over the template window, which needs more time than counting the location relationship of edge information to feature points to build SSSF descriptor.

E. Discussion

In this study, the influence of the template window size on the proposed method has been analyzed. As shown in Table I, the RMSEs of all tests are lower than 1 pixel. This indicates the stability of the proposed method to the size of template window.

Tables II and III evaluate the effectiveness of the morphological processing. In other words, the feature points on the buildings are not robust to be used to match images. Regardless of the size and density of the building, the feature points on the building should be removed. The designed structural elements are used to fill "holes" and smooth edges. No matter what structural elements are used, the ultimate goal is to get the complete building range. From the global point of view, the influence of the designed structural elements may be very small. What really affects the registration accuracy is the removing of the feature points on the building. Besides, when using morphological algorithm to remove the feature points on the building, there is a limitation. It may not be able to extract enough points for densely populated areas, which is also the limitation of the proposed method. Therefore, we can choose whether to apply this method according to the features.

As shown in Table II, experimental results prove the effectiveness of each improvement of the proposed method. In addition,

TABLE II
RMSEs OF THE PROPOSED METHOD WITH DIFFERENT VARIABLE

Category	Y	N1	Category	Y	N2
Habitation	0.47	0.53	Habitation	0.47	0.68
Bare land	0.76	0.98	Bare land	0.76	0.87
Mixed terrain	0.72	0.78	Mixed terrain	0.72	0.76
Farmland	0.55	0.59	Farmland	0.55	0.61
Buildings	0.82	0.85	Buildings	0.82	0.84

Y: proposed method with morphological processing and phase congruency;
N1: with morphological processing but without phase congruency;
N2: without morphological processing but with phase congruency.

TABLE III
REGISTRATION PERFORMANCE OF METHODS USED IN THIS STUDY

Category	Methods	Match points	CMR (%)	RMSE (pixels)	Time (s)
Habitation	Proposed	**1578/1680**	**93.97**	**0.47**	**59.71**
	SURF	1259/1680	74.97	0.87	70.86
	ECC	1373/1680	81.73	0.69	63.31
	DLSS	1434/1680	85.37	0.75	60.77
	SSSF	1463/1680	87.12	0.72	56.14
Bare land	Proposed	**1107/1206**	**91.84**	**0.76**	**44.28**
	SURF	1030/1206	85.41	1.33	55.13
	ECC	1060/1206	87.90	0.96	60.91
	DLSS	1067/1206	88.53	0.92	50.24
	SSSF	1072/1206	88.97	0.98	41.22
Mixed terrain	Proposed	**1244/1342**	**92.71**	**0.72**	**59.51**
	SURF	1127/1342	84.02	0.89	72.48
	ECC	1185/1342	88.32	0.81	61.72
	DLSS	1186/1342	88.37	0.76	63.42
	SSSF	1196/1342	89.18	0.78	57.36
Farmland	Proposed	**1323/1364**	**96.99**	**0.55**	**49.93**
	SURF	1057/1364	77.47	0.94	56.92
	ECC	1206/1364	88.91	0.72	59.32
	DLSS	1229/1364	90.15	0.63	55.61
	SSSF	1244/1364	91.23	0.68	46.87
Buildings	Proposed	**783/856**	**91.52**	**0.82**	**31.34**
	SURF	643/856	75.12	1.12	37.18
	ECC	715/856	83.61	0.91	40.03
	DLSS	756/856	88.38	0.84	33.95
	SSSF	748/856	87.41	0.87	29.78

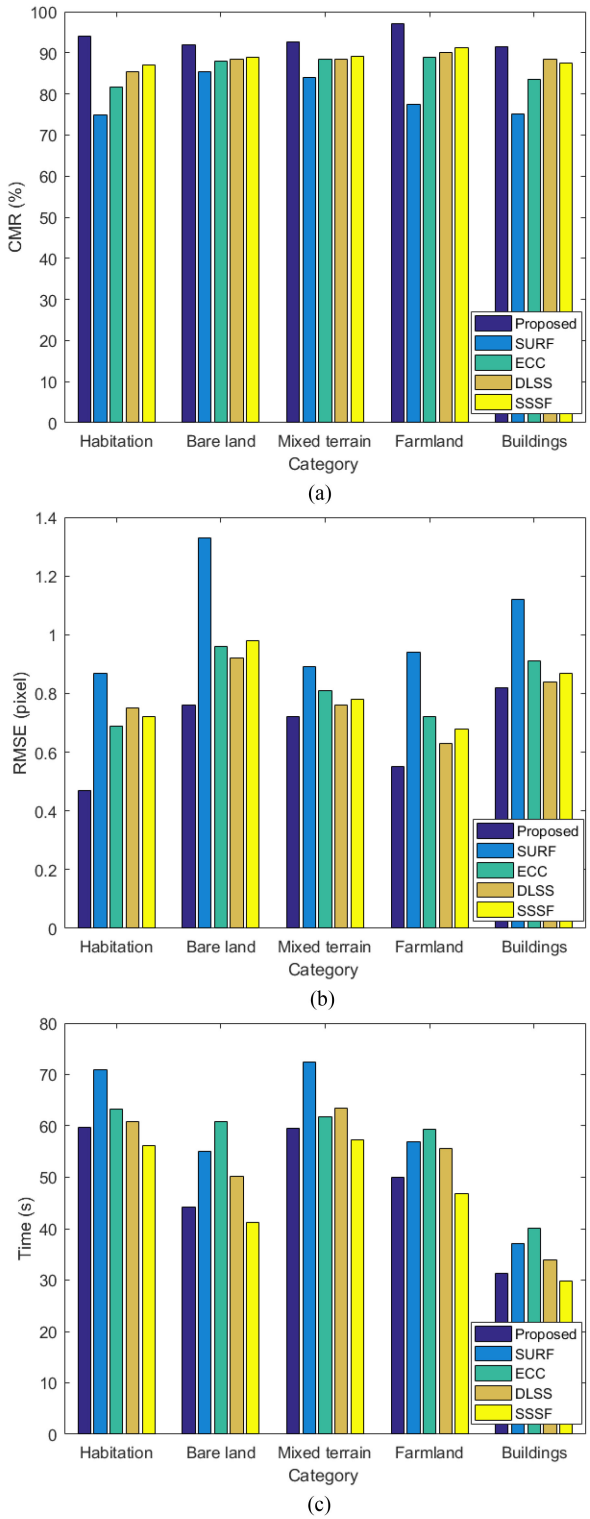


Fig. 12. (a) CMR, (b) RMSE, and (c) time of all methods.

the effectiveness of morphological processing and phase consistency algorithm are also reflected in the experimental results. First, as shown in Table III, the CMR differences of all methods in habitation and farmland (including buildings) are larger than bare land and mixed terrain. This demonstrates the effectiveness of morphological processing. Second, compared with the original SSSF method, the modified method still gets better accuracy in bare land and mixed terrain. Therefore, registration results demonstrate that the proposed method is robust for illumination and viewpoint change of UAV images.

From the experimental results, the CMR of the five algorithms is always Proposed $>$ SSSF $>$ DLSS $>$ ECC $>$ SURF. In general, affected by similar textures and illumination changes between UAV images, the SURF algorithm failed to achieve good matching results. Then, ECC algorithm is an enhanced NCC algorithm, which is robust to the linear differences between images rather than nonlinear intensity differences between UAV images. Differently, DLSS and SSSF algorithms are methods based on shape similarity. The sum of square differences of DLSS algorithm is sensitive to the gray changes caused by illumination and noise. As a result, DLSS algorithm cannot describe the similar texture between UAV images well. Compared with DLSS, SSSF algorithm can extract local structure information from the macro perspective. However, when the Canny operator is used to extract edge, it is inevitably affected by illumination changes. In addition, the feature points on the buildings are not stable because of local distortion caused by the viewpoint change. This study uses morphological processing and phase congruency to smooth the influence of illumination change, obvious rotation, viewpoint change and similar textures. Therefore, the proposed method achieves more reliable and comprehensive matching performance compared with DLSS, ECC, and SSSF. Theoretically, the registration result should be this sort, but there may be different precision for lots of images.

IV. Conclusion

This study utilizes morphological algorithm and phase congruency to improve the UAV image registration based on structural shape similarity. Considering the influence of central projection and viewpoint change on the image registration of UAV images, the feature points on buildings are eliminated by morphological processing. The phase congruency is utilized to detect structural features when calculating the SSSF descriptor, which is robust for the illumination change between adjacent UAV images. Then, the NCC between SSSF descriptors is used as similarity measure to detect tie points. The matching results indicate that the proposed method is reliable for viewpoint change, illumination change and obvious rotation between adjacent UAV images.

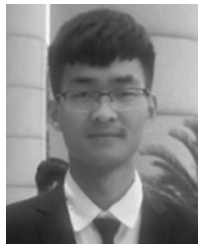
Besides, as shown in Fig. 10(b), there are few tie points distributing in bare land because it contains few terrain features. The proposed method is affected by the terrain shape structure of UAV images. Thus, an image enhancement method could be utilized to enhance the shape properties.

References

[1] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vision Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

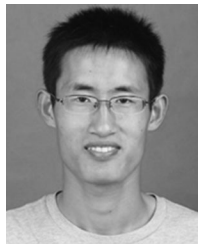
[2] M. Hao, W. Z. Shi, H. Zhang, and C. Li, "Unsupervised change detection with expectation-maximization-based level set," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 210–214, Jan. 2014.

[3] T. Zeng, W. Yang, and H. Wu, "UAV remote sensing image processing and classification in Wenchuan earthquake district," *Proc. SPIE*, vol. 7498, 2009, Art. no. 749823.
[4] D. Bulatov *et al.*, "Context-based automatic reconstruction and texturing of 3D urban terrain for quick-response tasks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 93, pp. 157–170, Jul. 2014.
[5] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath, "Brain functional localization: A survey of image registration techniques," *IEEE Trans. Med. Imag.*, vol. 26, no. 4, pp. 427–451, Apr. 2007.
[6] H. D. Tagare and M. Rao, "Why does mutual-information work for image registration? A deterministic explanation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1286–1296, Jun. 2015.
[7] V. Argyriou and T. Vlachos, "Estimation of sub-pixel motion using gradient cross-correlation," *Electron. Lett.*, vol. 39, no. 13, pp. 980–982, Jun. 2003.
[8] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 12, pp. 1156–1168, Dec. 1994.
[9] W. K. Pratt, *Digital Image Processing: PIKS Inside*. Hoboken, NJ, USA: Wiley, 2001.
[10] G. D. Evangelidis and E. Z. Psarakis, "Parametric image alignment using enhanced correlation coefficient maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1858–1865, Oct. 2008.
[11] F. Zhou, W. M. Yang, and Q. M. Liao, "A coarse-to-fine subpixel registration method to recover local perspective deformation in the application of image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 53–66, Jan. 2012.
[12] M. L. Uss, B. Vozel, V. A. Dushepa, V. A. Komjak, and K. Chehdi, "A precise lower bound on image subpixel registration accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3333–3345, Jun. 2014.
[13] S. M. M. Kahaki, M. J. Nordin, and A. H. Ashtari, "Contour-based corner detection and classification by using mean projection transform," *Sensors*, vol. 14, no. 3, pp. 4126–4143, Mar. 2014.
[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[15] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, New York, NY, USA, 2006, pp. 1978–1983.
[16] Y. Ke, R. Sukthankar, and I. C. Society, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, 2004, vol. 2, pp. 506–513.
[17] H. Bay *et al.*, "Speeded-up robust features (SURF)," *Comput. Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[18] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 438–469, 2009.
[19] B. Li and H. Ye, "RSCJ: Robust sample consensus judging algorithm for remote sensing image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 574–578, Jul. 2012.
[20] V. A. Zimmer, M. A. G. Ballester, and G. Piella, "Multimodal image registration using Laplacian commutators," *Inf. Fusion*, vol. 49, pp. 130–145, Sep. 2019.
[21] Y. X. Ye, L. Shen, M. Hao, J. Wang, and Z. Xu, "Robust optical-to-SAR image matching based on shape properties," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 564–568, Apr. 2017.
[22] S. M. M. Kahaki *et al.*, "Deformation invariant image matching based on dissimilarity of spatial features," *Neurocomputing*, vol. 175, pp. 1009–1018, Jan. 2016.
[23] X. Xie *et al.*, "A novel extended phase correlation algorithm based on Log-Gabor filtering for multimodal remote sensing image registration," *Int. J. Remote Sens.*, vol. 40, no. 14, pp. 5429–5453, Jul. 18, 2019.
[24] M. Hao *et al.*, "Robust multisource remote sensing image registration method based on scene shape similarity," *Photogrammetric Eng. Remote Sens.*, vol. 85, no. 10, pp. 725–736, Oct. 2019.
[25] B. Barman and S. Patra, "Hyperspectral image analysis using neighborhood rough set and mathematical morphology," in *Proc. Int. Conf. Accessibility Digit. World*, 2016, pp. 75–80.
[26] A. Jain *et al.*, "Comparison of edge detectors," in *Proc. Int. Conf. Med. Imag., m-Health Emerg. Commun. Syst.*, 2014, pp. 289–294.
[27] Y. Ye, S. Jie, 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.
[28] S. M. M. Kahaki, S. L. Wang, and A. Stepanyants, "Accurate registration of *in vivo* time-lapse images," *Proc. SPIE*, vol. 10949, 2019, Art. no. 109491D.
[29] S. Drazic, "Advanced morphological distances based on dilation and erosion," *Fundamenta Informaticae*, vol. 164, no. 1, pp. 17–39, 2019.
[30] Y. Zhang *et al.*, "Joined fragment segmentation for fractured bones using GPU-accelerated shape-preserving erosion and dilation," *Med. Biol. Eng. Comput.*, vol. 58, pp. 155–170, 2020.
[31] S. Belongie, J. Malik, and J. Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, vol. 13, pp. 831–837.
[32] X. M. Liu, J. B. Li, and J. S. Pan, "Feature point matching based on distinct wavelength phase congruency and Log-Gabor filters in infrared and visible images," *Sensors*, vol. 19, no. 19, Oct. 2019, Art. no. E4244.
[33] M. H. Shi *et al.*, "Conformal monogenic phase congruency model-based edge detection in color images," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 10701–10716, Apr. 2019.



Jian Jin received the B.S. degree in 2014 from the China University of Mining and Technology, Xuzhou, China, where he is currently working toward the master's degree in surveying engineering with the School of Environment Science and Spatial Informatics.

His research interests include multisource remote sensing image registration and UAV image mosaic.



Ming Hao received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2015.

He is currently an Associate Professor with the School of Environment Science and Spatial Informatics, China University of Mining and Technology. He has authored or coauthored more than 30 peer-reviewed articles in international journal such as the IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, *Remote Sensing*, *International Journal of Remote Sensing*, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. His research interests include change detection, image registration, and image fusion for remote sensing images.