

Detecting Designated Building Areas From Remote Sensing Images Using Hierarchical Structural Constraints

Fukun BI, Mingyang LEI*, Zhihua YANG, Jinyuan HOU, and Yanyan QIN

School of Electronic Information Engineering, North China University of Technology, Beijing 100144, China

*Corresponding author: Mingyang LEI E-mail: 752225185@qq.com

Abstract: Automatic detection of a designated building area (DBA) is a research hotspot in the field of target detection using remote sensing images. Target detection is urgently needed for tasks such as illegal building monitoring, dynamic land use monitoring, antiterrorism efforts, and military reconnaissance. The existing detection methods generally have low efficiency and poor detection accuracy due to the large size and complexity of remote sensing scenes. To address the problems of the current detection methods, this paper presents a DBA detection method that uses hierarchical structural constraints in remote sensing images. Our method was conducted in two main stages. (1) During keypoint generation, we proposed a screening method based on structural pattern descriptors. The local pattern feature of the initial keypoints was described by a multilevel local pattern histogram (MLPH) feature; then, we used one-class support vector machine (OC-SVM) merely to screen those building attribute keypoints. (2) To match the screened keypoints, we proposed a reliable DBA detection method based on matching the local structural similarities of the screened keypoints. We achieved precise keypoint matching by calculating the similarities of the local skeletal structures in the neighboring areas around the roughly matched keypoints to achieve DBA detection. We tested the proposed method on building area sets of different types and at different time phases. The experimental results show that the proposed method is both highly accurate and computationally efficient.

Keywords: DBA detection; local structural constraint; multilevel local pattern histogram (MLPH); similarity of the local structure; scale invariant feature transform (SIFT)

Citation: Fukun BI, Mingyang LEI, Zhihua YANG, Jinyuan HOU, and Yanyan QIN, "Detecting Designated Building Areas From Remote Sensing Images Using Hierarchical Structural Constraints," *Photonic Sensors*, 2020, 10(1): 45–56.

1. Introduction

With technological advances in sensor platforms and observation loads, remote sensing technology applications have gradually expanded. Detecting a designated building area (DBA) from a remote sensing image fulfills an urgent need in important fields such as illegal building monitoring, dynamic land use monitoring, antiterrorism efforts, and military reconnaissance [1–3].

Remote sensing images are typically large and contain complex scenes. Therefore, a DBA detection method must be both highly efficient and accurate. X. Yao *et al.* [4] used a target-oriented saliency model and a learned condition random field (CRF) model to achieve accurate detection of differently scaled airport targets in remote sensing images. Considering the symmetric nature of circular oil depots, A. O. Ok *et al.* [5] proposed an automated

Received: 2 October 2018 / Revised: 18 April 2019

© The Author(s) 2019. This article is published with open access at Springerlink.com

DOI: 10.1007/s13320-019-0558-5

Article type: Regular

thresholding method that focused on circular regions and designed a new measure and a circle support ratio to verify the detected circles. Because oil tanks appear in remote sensing images as circles and this information is insufficient to separate the targets from the complex backgrounds, L. Zhang *et al.* [6] proposed a hierarchical oil tank detector that used deep surrounding features extracted by a deep learning model. However, the aforementioned methods are strongly dependent on the feature sets for specific target types such as airports or oil tanks. In addition, the characterizations of the surroundings of each building area in different remote sensing images will differ considerably due to the incidental angle of the satellite platform, the lighting conditions, weather, and other factors. To solve this problem, some scholars have recently attempted to use local descriptor technologies to match and detect building areas. B. Sırmaçek *et al.* [7] used the Gabor filter to extract local feature points and then detect urban areas from remote sensing images. C. Tao *et al.* [8] used segmented region information, prior knowledge, and scale invariant feature transform (SIFT) keypoints to detect airport areas. SIFT [9] is one of the most effective local descriptor-matching techniques, which makes target detection more robust and reliable [10, 11]. However, due to the rich variety of ground objects in large remote sensing images, the above methods may produce large numbers of redundant local descriptors in nonbuilding areas that have abundant image textural features. On one hand, the large number of features adversely affect the efficiency and precision of matching and detection. On the other hand, the traditional local descriptor-matching methods are not effectively designed to represent building area characteristics; thus, their matching efficiency and accuracy can suffer. Deep learning models have recently been applied to object detection [12]. However, these methods need to scan and identify the whole image with huge computational costs, and the building area belongs to the distributed target

without specific appearance characteristics, which is not suitable for detecting by deep learning based methods.

To address the limitations of the above detection methods on large and complex remote sensing image scenes, this paper presents an effective DBA detection method that uses hierarchical local structural constraints. This method makes full use of the local structural information of building areas. We design a hierarchical constraint strategy to screen the most reliable candidate keypoints of building areas and match the keypoint pairs precisely. The method presented in this paper both greatly reduces the number of required feature description calculations in the redundant nonbuilding areas and effectively decreases the number of matching point errors to achieve rapid and reliable DBA detection.

2. Detection algorithm

This paper proposes a DBA detection method by using the hierarchical constraints of local structures, based on building area features and spatial distribution characteristics. Our method is divided into two parts. Firstly, to reduce the number of keypoints in the nonbuilding areas, we propose a keypoint-screening method for suspected building areas based on structural pattern descriptions. The local pattern feature of the initial keypoints is described by a multilevel local pattern histogram (MLPH) [13]; then, we use one-class support vector machine (OC-SVM) [14–17] to screen the keypoints that exhibit building attributes. Secondly, based on local structural similarities, we propose a keypoint-matching method to detect the DBA. To improve the reliability of keypoint matching in large building areas after this rough matching, we achieve precise matching by calculating the similarities in the local skeletal structures around the keypoints. Finally, the DBA can be obtained by using these reliable matching pairs of keypoints. Figure 1 shows the workflow of our algorithm.

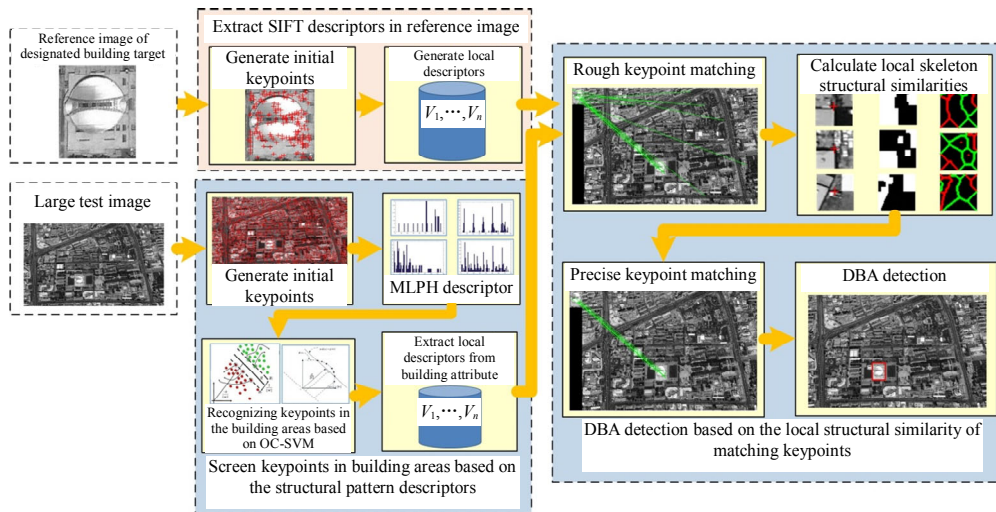


Fig. 1 Workflow of the proposed detection algorithm.

2.1 Screening of keypoints in building areas based on the structural pattern descriptors

SIFT keypoint matching is a classical object-detection method used in natural image processing [18], however, it is difficult to apply these SIFT keypoints directly to detect the DBA from remote sensing images because remote sensing images generally present two major difficulties: large sizes and complex scenes. During the local descriptors extraction stage, the SIFT algorithm produces a large number of keypoints in nonbuilding areas. These abundant descriptors not only increase the amount of calculation required during the subsequent matching, but also affect the detection accuracy. In this paper, we first filter the keypoints in suspected building areas of the remote sensing images. This process is divided into the following three main steps.

2.1.1 Generating initial keypoints

We use the SIFT algorithm to extract the initial keypoints from both the reference image and the remote sensing test image. SIFT features possess valuable properties such as invariance to rotation, scaling, affine transformation, illumination, and view transformation. The steps to extract keypoints are as follows: extreme point detection, precise

keypoint location, keypoint directional distribution, and local descriptor generation [9]. Because remote sensing images often include complex and diverse types of ground objects, SIFT keypoint extraction produces numerous SIFT keypoints. Typical scenes include building areas, wooded areas, green spaces, and bare ground, as shown in Fig. 2.

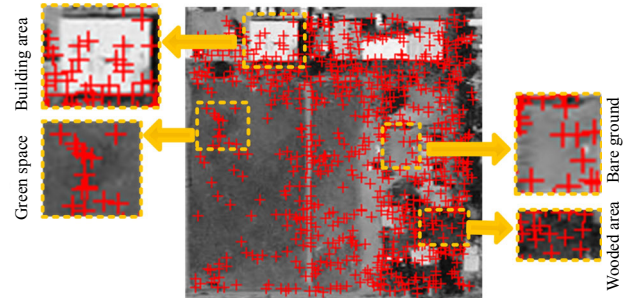


Fig. 2 Generation of initial keypoints in a typical remote sensing scene.

2.1.2 MLPH descriptor

Multilevel local pattern histogram (MLPH) [13] is an effective local descriptor of local image structural patterns. Because the MLPH features of neighboring areas around keypoints have obvious differences around different ground objects, the MLPH descriptor can be used to distinguish the keypoints in building areas from those in other areas. Based on the initial SIFT keypoints obtained from the entire image, we extract the MLPH descriptor of

each neighboring area around the keypoints to evaluate its corresponding local pattern features.

MLPH calculation involves three steps: image quantization, matrix splitting, and pattern histogram generation. Each initial keypoint is compared with its neighboring pixels within a window with a size $h \times h$. We denote the intensity of the central pixel as g_c . Then, all the pixel intensities within the window are quantized using the following formula to produce a pattern matrix:

$$s(i) = \begin{cases} +1 & g_i > g_c + t \\ 0 & |g_i - g_c| \leq t \\ -1 & g_i < g_c - t \end{cases}, \quad i \in [1, 2, \dots, h^2] \quad (1)$$

where g_i is the intensity of pixel i in the local window, and t is a predefined threshold.

The pattern matrix is split into three matrices: a “positive matrix” (PM), an “equal matrix” (EM), and a “negative matrix” (NM), respectively defined by the following functions:

$$PM(i) = \begin{cases} 1, & s(i) > 0 \\ 0, & \text{others} \end{cases}, \quad i \in [1, 2, \dots, h^2] \quad (2)$$

$$EM(i) = \begin{cases} 1, & s(i) = 0 \\ 0, & \text{others} \end{cases}, \quad i \in [1, 2, \dots, h^2] \quad (3)$$

$$NM(i) = \begin{cases} 1, & s(i) < 0 \\ 0, & \text{others} \end{cases}, \quad i \in [1, 2, \dots, h^2]. \quad (4)$$

For each matrix, a subhistogram is calculated based on the local pattern. The subhistogram is constructed as follows:

$$bins(k) = \sum_{n=1}^N \delta[num(n) = k] \quad (5)$$

where $bin(k)$ is the value of the k th bin, N is the number of local patterns in the matrix, and $num(n)$ is the number of pixels in the n th local pattern. The function $\delta[\cdot]$ yields one if its argument is true and zero otherwise.

To reduce the histogram dimensions and increase the identifiability of the local pattern

histogram, the histogram is merged according to the following formula:

$$vol(k) = B \times vol(k-1), \quad k \in [2, 3, \dots, K] \quad (6)$$

where $vol(k)$ denotes the “volume” of the k th bin in the simplified subhistogram, and B is a parameter to control the growth rate of $vol(k)$. Our experimental value is $B = 2$. The local pattern histogram is obtained by concatenating three simplified subhistograms.

Different scales correspond to different thresholds of t . The MLPH is formed by concatenating local pattern histograms extracted at multiple scales. The growth rate of t is defined as follows:

$$\begin{aligned} t_m &= T \times t_{m-1}, \quad m \in [2, 3, \dots, M] \\ s.t. \quad t_M &< C < t_{M+1} \end{aligned} \quad (7)$$

where T is a parameter to control the growth rate of t , M is the total number of levels, and C is the maximum contrast value in the image ($C = 255$). For our applications, our experimental values are $T = 2$ and $M = 5$. Thus, the total dimension of the MLPH is

$$M \times 3 \times K.$$

The general framework of the method is shown in Fig. 3.

2.1.3 Recognizing keypoints in the building areas based on OC-SVM

Due to the relatively stable structural characteristics of building areas, their MLPH attributes of the corresponding keypoints are also both similar and stable. In contrast, the MLPH attributes of other kinds of ground objects in the nonbuilding area are diverse and unstable. Using the MLPH attribute extracted by the above steps, the outstanding OC-SVM classifier [14–17] is applied to build a model of the keypoint neighborhoods. Firstly, a large number of initial keypoints are generated from the data set in the keypoints extraction step of the SIFT, and some blocks of the neighboring area around keypoints in building area are manually selected for OC-SVM training from

them. Then, during the keypoint-screening stage, we use the trained OC-SVM model to determine

whether the keypoints belong to the building or nonbuilding area.

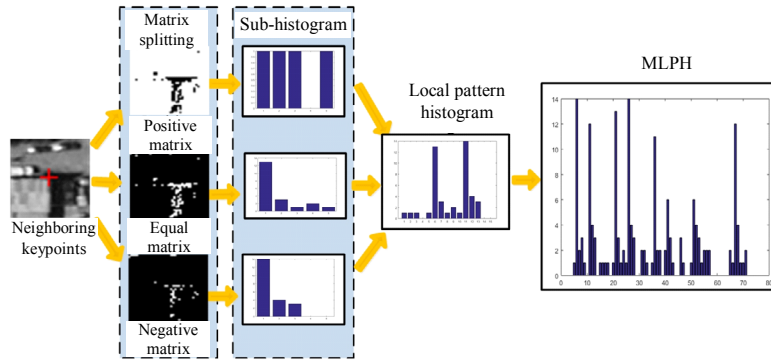


Fig. 3 Extracting MLPH features. From left to right, the panels show the original image, the matrix splitting procedure, the subhistograms derived from the split matrices, the local pattern histogram, and the MLPH, respectively. The local pattern histogram is formed by concatenating three subhistograms, and the subhistograms are computed based on (6) with $B = 2$ and $K = 5$. The MLPH is formed by concatenating the local pattern histograms of different scales, and the local pattern histograms are computed based on (7) with $T = 2$ and $M = 5$.

Compared with the initial keypoints distribution in Fig. 2, after the screening stage, the keypoints in the nonbuilding areas in Fig. 4 have been greatly suppressed.

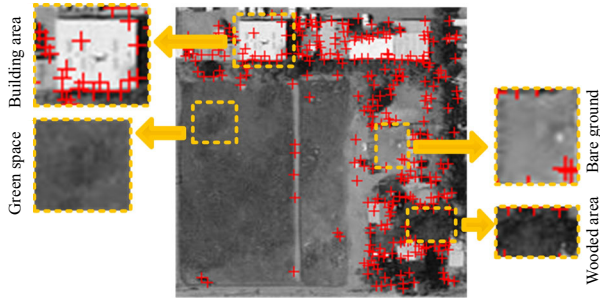


Fig. 4 Keypoints distribution after screening.

2.1.4 DBA detection based on the local structural similarities of matching keypoints

After the keypoints in suspected building areas are obtained, we propose a stable keypoint-matching method based on local structural similarity to improve the matching accuracy in large and complex remote sensing images. Firstly, based on Euclidean distance, we use a keypoint-screening method to roughly identify matching pairs. Then, because the initial matching keypoints that characterize the same building area should contain similar local information at the same scale – particularly local skeletal structure similarity – we

adopt this feature to further identify stable matching pairs with high structural similarity to obtain the DBA. The specific steps are listed below.

2.1.5 Rough keypoint matching

As discussed above, we extract the DBA SIFT keypoints from the reference image and simultaneously extract and screen out the keypoints with building attributes from the remote sensing test image. Then, we calculate the SIFT matching features (128 dimensions) for each extracted keypoint. Next, based on the Euclidean distance, we perform a rough matching of the filtered keypoints to obtain matching point pairs. The matched pairs satisfy the following formula:

$$\frac{d(R_i, S_j)}{d(R_i, S_k)} < thr \quad (8)$$

where R_i and S_j are the SIFT descriptors in the reference image and the remote sensing test image, respectively, $d(R_i, S_j)$ is the minimum Euclidean distance, $d(R_i, S_k)$ is the second smallest distance, and thr is a threshold whose value is generally 0.8 in our application.

2.1.6 Precise keypoint matching based on local skeletal structural similarities

Because of the size and complexity of remote

sensing images, many mismatches may appear in the keypoint pairs after the rough matching process. To solve this problem, this paper proposes a precise keypoint-matching strategy based on the similarity of the local skeletal structure as described below.

(1) Calculating the skeletal structures of neighboring areas around keypoints

In this paper, a bright and dark linear skeletal structure is proposed to describe the features of the neighboring areas around matched keypoints. The set of matched keypoints in the reference image after the rough matching process is defined as: $P = \{p_i\}$, $i = 1, 2, \dots, N$, where N is the number of roughly matched pairs, and each keypoint p_i has three attributes: a position $[(x, y)]$, a scale (σ), and a direction (o).

For each keypoint, we calculate the bright and dark linear skeletal density of its neighboring areas at the corresponding scale and position. The radius of the keypoint neighborhood is equal to the radius of the corresponding descriptor, which is determined as follows:

$$radius = \frac{3\sigma_{oct} \times \sqrt{2} \times (d+1) + 1}{2} \quad (9)$$

where $d = 4$, and σ_{oct} is the scale of the keypoint.

The bright and dark linear skeletons are extracted from the neighboring areas around the matched keypoints; then, we calculate the density, which represents the structural characteristics of the local area using the following steps. Firstly, the binary image (BM) of the keypoints is calculated by using the adaptive threshold segmentation method:

$$BM = [\text{Otsu}(MoP) \circ b] \bullet b. \quad (10)$$

We adopt the opening and closing operation for the binary image to remove spot noise, fill the holes, and obtain the binary image data that reflects the main contour. Here, the Otsu is an efficient operation for image binarization, $f \circ b$ denotes using the structural elements b to perform the morphological opening operation for the image, and $f \bullet b$ denotes using the structural elements b to

perform the morphological closing operation.

The bright linear skeleton (morphological skeletal operation) is extracted from the foreground parts of the neighboring areas around the keypoints (those parts whose pixel values are “1”). Then, we calculate the density $density_{br}$ as follows:

$$density_{br} = \frac{\sum Mor[Mor(BM, skel), spur]}{row(BM) \times col(BM)} \quad (11)$$

where $Mor(BW, opt)$ is defined as a morphological skeletal operation on the binary image, $row(\bullet)$ is the number of rows in the image, and $col(\bullet)$ is the number of columns. When $opt = skel$, the skeletons have been extracted, and when $opt = spur$, the burrs in the skeletons have been cleared.

Accordingly, the extracted dark linear skeletal area is the corresponding background part of the image (the part whose pixel values are “0”), and the density $density_{dr}$ is calculated as follows:

$$density_{dr} = \frac{\sum Mor[Mor((1-BM), skel), spur]}{row(1-BM) \times col(1-BM)} \quad (12)$$

where “1” indicates an all-1 matrix with the same size as BM .

When this process completes, the bright linear skeletal densities $density_{br}$ and $density_{bs}$, and the dark linear skeletal densities $density_{dr}$ and $density_{ds}$ have been extracted for each candidate matching keypoint in both the reference and test images.

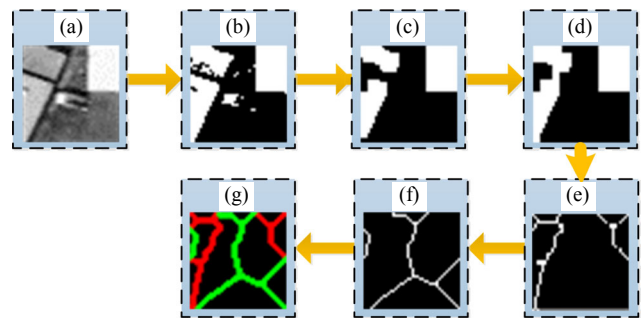


Fig. 5 Process of skeletal structural extraction from the neighboring areas around matching keypoints: (a) original image, (b) threshold segmentation, (c) opening operation, (d) closing operation, (e) skeleton operation, (f) antiskelton operation, and (g) bright and dark line extraction.

(2) Calculating the local skeletal structural similarity for the matching keypoint pairs

After the rough matching, because there are “many-to-one” matching point features in the reference image and the test image, this paper calculates the bright and dark linear skeletal similarity of each matching point pair to perform further screening. We define the similarity of the bright and dark linear skeletons among p_{ri} , the keypoint p_{si}^1 , and the corresponding matching keypoint (here, the superscripted ¹ represents the first matching point) from the reference image and the test image, respectively as follows:

$$S(p_{ri}, p_{si}^1) = \frac{1}{2} \times \frac{\min(\text{density}_{br}, \text{density}_{bs})}{\max(\text{density}_{br}, \text{density}_{bs})} + \frac{1}{2} \frac{\min(\text{density}_{dr}, \text{density}_{ds})}{\max(\text{density}_{dr}, \text{density}_{ds})}. \quad (13)$$

Greater similarities in the neighboring skeletal structures between the test and reference images reflect smaller density value differences (S values closer to 1).

(3) Screening of reliable matching pairs based on the similarity

The unreliable matching pairs are removed, including the “many-to-one” and unstable matching pairs. Then, we use the maximum value of the skeletal similarity S to identify the reliable matching pairs. Assume that a keypoint p_{ri} in the reference image corresponds to multiple matching points $p_{si}^1, p_{si}^2, \dots, p_{si}^{N_0}$ (where N_0 is the number of matching points, and $N_0 \in [1, N]$) in the test image. We obtain the N_0 corresponding skeletal similarities $S(p_{ri}, p_{si}^1), S(p_{ri}, p_{si}^2), \dots, S(p_{ri}, p_{si}^{N_0})$ and find the maximum value:

$$S(p_{ri}, p_{si}^r) = \max[S(p_{ri}, p_{si}^1), \dots, S(p_{ri}, p_{si}^{N_0})], \quad r \in [1, N_0] \quad (14)$$

which corresponds to the r th similarity in N_0 . We define that the keypoint p_{si}^r is the only proper corresponding point to the matching point p_{ri} .

The similarity of skeletal density characterizes

the structural similarity of the two neighboring areas around the keypoints and is used to set the threshold T . When a matching pair satisfies

$$S(p_{ro}, p_{so}) < T \quad (15)$$

the matching pair is unstable and removed. Our experimental value is $T = 0.5$. Finally, the most reliable matching pairs are obtained, and we then conduct a precise matching process by using these reliable matching pairs. When all the values of S are less than the threshold T , we assume that the building area does not exist in the test image.

2.1.7 Extracting the DBA by using the reliable matching point pairs

The structural similarity is sorted according to the matching points retained after the similarity process, and three pairs with the highest similarity are selected as affine transformations to obtain the DBA.

We denote the matching points with the highest local skeletal similarity S in the reference picture as (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) , and the corresponding matching points in the test image as (x_1', y_1') , (x_2', y_2') , and (x_3', y_3') . Then, we substitute them into the affine transformation model as follows:

$$\begin{bmatrix} x_1 & y_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & y_1 & 0 & 1 \\ x_2 & y_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_2 & y_2 & 0 & 1 \\ x_3 & y_3 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_3 & y_3 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \end{bmatrix}. \quad (16)$$

For convenience, the above equation is abbreviated as: $\mathbf{Ax}=\mathbf{b}$. \mathbf{A} , \mathbf{x} , and \mathbf{b} represent three matrix terms in (16), respectively. Then, the least squares solution of the above-determined system is

$$\mathbf{x} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{b} \quad (17)$$

where T denotes the transposition operation.

Using the affine transformation matrix \mathbf{x} , the boundary values of the given reference image are

substituted into the transformation model to obtain the DBA values in the test image. Finally, the DBA area is obtained.

3. Experiments

3.1 Dataset and experimental setting

To demonstrate the performance and efficiency of the proposed detection method, we designed qualitative and quantitative experiments to test the method in real and complex remote sensing scenes. An optical remote sensing image of Beijing Institute of Technology (BIT) taken by Google Earth on June 28, 2009 was obtained as the DBA template to be detected. We identified the stadium (area ①, a single large-scale building), multiple student apartments (area ②, multiple buildings), and the Qiushi lab (area ③, large building) as shown in Fig. 6. Under the influence of the incidence angles, illumination conditions, weather, and other factors related to the satellite platform at different times, the characterization of the DBA and its surrounding

environment will differ greatly between remote sensing images, as illustrated in Fig. 6. Therefore, to select a large test image for detection, we acquired from Google Earth at different times a total of 30 remote sensing images of the BIT and their vicinities, in addition to 10 non-BIT remote sensing images. These 40 images represented different seasons, different incident angles, different illumination conditions, etc. The size of each image was 1920×1080 pixels. Three DBAs (the template areas to be detected shown in Fig. 6) were detected from each image to verify the effectiveness of the method proposed in this paper. Using SIFT, a large number of initial keypoints were generated from the data set, and 4000 blocks of the neighboring area around keypoints in building area were manually selected for OC-SVM training from them. All the experiments were programmed by using MATLAB 2015a. The experimental platform was a personal computer (PC) with a 3.70GHz Intel Core i3 CPU and 4 GB of RAM.

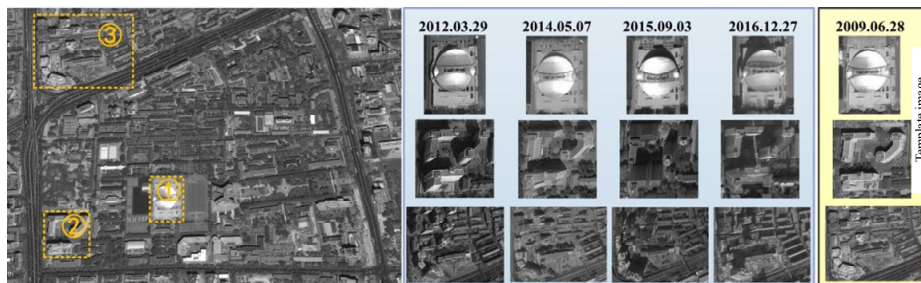


Fig. 6 DBA image examples of the BIT at different times.

3.2 Results and evaluation

3.2.1 Performance analysis of the proposed method

This section presents and discusses the experimental results to demonstrate the effectiveness of the proposed method. We selected the panoramic images of the BIT at different times and detected three DBAs from each image by using the approach discussed in Section 3.1. The results are shown in Fig. 7 (for display convenience, the order of the

described areas is adjusted to ② ③ ①). As shown in the figure, the initial number of keypoints [Fig. 7 (a)] in these complex images at different times is large. After screening based on the local structural pattern features, the number of keypoints with nonbuilding attributes is greatly reduced [Fig. 7(b)]. However, many mismatched pairs exist in the initial matching pairs: the initial number of matching pairs is shown in Fig. 7(c). After the precise matching process based on the local structural similarity of keypoints, the number of the wrong matching pairs [Fig. 7(d)] is

reduced considerably.

In addition, this paper analyzes the detection results of all the 30 large test images of the BIT via a quantitative statistics method, as shown in Tables 1 and 2. It can be observed that the number of keypoints has been reduced by approximately 50% on average by screening the keypoints of the suspected building attributes in Table 1, and since most of the removed keypoints are in nonbuilding areas, the final successful matching points are hardly affected. In addition, compared with the traditional RANSAC algorithm in Table 2, the number of correct matching pairs (CP) extracted by using the similarities of skeletal structures is larger, and the reliability is higher for the same keypoints.

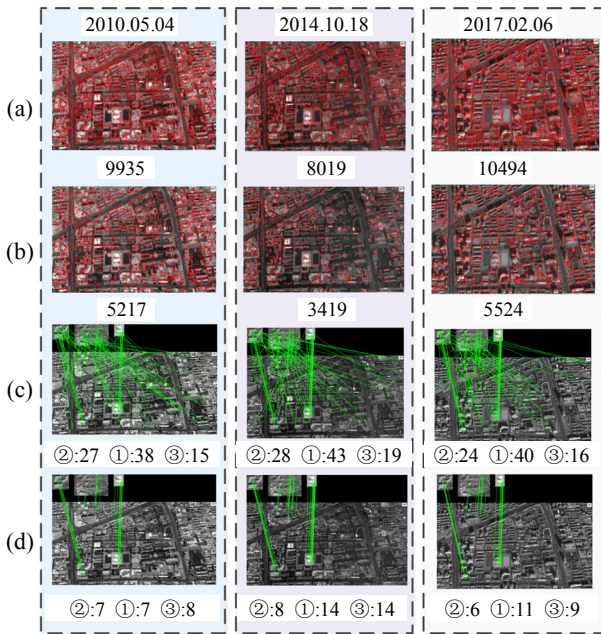


Fig. 7 Algorithmic performance: (a) the initial number of keypoints, (b) the number of keypoints with nonbuilding attributes, (c) the initial number of matching pairs, and (d) the number of precisely matched pairs.

Table 1 Numbers of keypoints.

Parameter	Number
Initial SIFT keypoints	9765
Keypoints after screening	4845

Table 2 Algorithmic performance statistics in matching stage.

Parameter	Area ①	Area ②	Area ③
Initial matching pairs	18	26	43
CP of RANSAC	10	11	7
CP of the proposed method	11	13	10

3.2.2 Performance comparison with typical matching detection methods

To further verify the detection performance of the proposed method for the DBA, we compared it with the classical SIFT and other typical detection methods applied in [19, 20]. We detected three test building areas (described in Section 3.1) in the 40 remote sensing image scenes taken at different times (including 30 remote sensing images of the BIT and their vicinities, and 10 non-BIT remote sensing images).

When more than 90% of the DBA area was detected, we considered the building area to have been detected correctly. We denoted the number of correct target detections as N_c , the number of mislabeled targets as N_f , and the total number of the hand-marketed real targets included in all the test images as N_t . The true detection rate (TD) is defined as follows:

$$TD = N_c / N_t \times 100\%. \quad (18)$$

The false alarm rate (FA) is defined as

$$FA = N_f / (N_c + N_f) \times 100\%. \quad (19)$$

We calculated the numbers of keypoints, the values of CP, TD, FA, and the computing times of the four compared algorithms. Table 3 reports the averages of all these test image indicators.

Table 3 Performance comparison of different algorithms.

Method	Keypoints	CP	TD (%)	FA (%)	Time (s)
Classical SIFT	9508	10	93.3	6.6	5.8
Method [19]	10452	10	82.2	17.7	3.6
Method [20]	12785	6	51.1	48.8	1.9
Proposed method	4643	11	94.4	5.5	3.3

As indicated in Table 3, the number of keypoints generated by the proposed method for the test images, which are large and complex scenes, is noticeably less than that of the other compared methods, and the proposed method achieved the largest number of correct matching pairs. This result was attributed to the strong correlation with the structural pattern description and local structural similarity proposed in this paper. In contrast with the other methods, we note that the method proposed in [19], which was described the building area neighborhood based on 64-dimensional descriptors, did not perform as well as the classical SIFT 128-dimensional descriptors. The descriptors generated by the method presented in [20] were single-scale and contained no directional information. Thus, the descriptors produced by the existing methods were not effectively designed for building area characteristics. In addition, because only the keypoints with building attributes retained in the screening step were hierarchically matched by the rough and precise matching processes, the proposed method achieved the best performance (the highest true detection rate and the lowest false alarm rate). In contrast, in the method proposed in [19], the keypoints were matched by using the Euclidean distance and the Hessian matrix trace; however, this approach resulted in a low detection rate because of the many mismatching pairs produced from the complex remote sensing scenes. In terms of computational efficiency, the proposed method selected only the keypoints with building attributes to participate in the subsequent characterization and matching operations. Compared with classical SIFT and the method proposed in [19], the runtime of the proposed method was relatively short. In addition, although the method presented in [20] had the shortest computing time, there was a substantial gap between its performance and that of the other compared methods.

Overall, from a timeliness perspective, the

proposed method had both higher performance and greater reliability.

4. Conclusions

The DBA automatic detection technology has a wide range of applications; consequently, it has a high research value in the remote sensing field. However, because remote sensing images are large and complex, the traditional building area detection methods will produce a large number of redundant local descriptors, and the traditional local descriptor-matching methods are not designed effectively for the characteristics of building areas. To address these problems, we presented a hierarchical matching method for DBA detection. The proposed method first used a rough keypoint-matching method based on the MLPD feature description and OC-SVM classifier screening to identify the points with building attributes. Then, we used a precise method that used the skeletal density of the neighboring areas around keypoints to represent the keypoint's local structure and match local structures with high similarities. To demonstrate the performance of this method, we established a multitype test database of DBAs at different times based on remote sensing data from Google Earth. The experimental results showed that the keypoint-screening method using structural patterns could effectively reduce the computational cost of the subsequent local descriptors and largely avoid the problem of matching keypoints in nonbuilding areas. The precise matching method based on local structural similarity effectively removed unreliable matching pairs and achieved an accurate DBA detection. The proposed method not only guaranteed a high detection rate and a low false alarm rate but also greatly improved the efficiency of DBA detection.

Nevertheless, this method still has some limitations. Although it achieved a good matching and detection effect on panchromatic remote sensing images in our experiment, its detection

performances for other types of remote sensing images such as synthetic-aperture radar (SAR) and infrared images have not been verified. Our future work direction will concentrate on this aspect.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61601006), the Beijing Natural Science Foundation (Grant No. 4192021), and the Equipment Pre-Research Foundation (Grant No. 61404130312).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," in *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, Italy, 2015, pp. 1873–1876.
- [2] S. Ghaffarian and S. Ghaffarian, "Automatic building detection based on purposive fastICA (PFICA) algorithm using monocular high resolution Google earth images," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 97: 152–159.
- [3] A. Manno-Kovács and A. O. Ok, "Building detection from monocular VHR images by integrated urban area knowledge," *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(10): 2140–2144.
- [4] X. Yao, J. Han, L. Guo, S. H. Bu, and Z. B. Liu, "A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF," *Neurocomputing*, 2015, 164: 162–172.
- [5] A. O. Ok and E. Baseski, "Circular oil tank detection from panchromatic satellite images: a new automated approach," *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(6): 1347–1351.
- [6] L. Zhang, Z. Shi, and J. Wu, "A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(10): 4895–4909.
- [7] B. Sirmacek and C. Unsalan, "Urban area detection using local feature points and spatial voting," *IEEE Geoscience and Remote Sensing Letters*, 2010, 7(1): 146–150.
- [8] C. Tao, Y. Tan, H. Cai, and J. W. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geoscience and Remote Sensing Letters*, 2011, 8(1): 128–132.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoint," *International Journal of Computer Vision*, 2004, 60(2): 91–110.
- [10] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg, "Real-time detection and tracking for augmented reality on mobile phones," *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(3): 355–368.
- [11] P. Apostolos and N. Christos, "Vehicle logo recognition using a sift-based enhanced matching schem," *IEEE Transaction on Intelligent Transportation Systems*, 2010, 11(2): 322–328.
- [12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ohio, United States, 2014, pp. 580–587.
- [13] D. Dai, W. Yang, and H. Sun, "Multilevel local pattern histogram for SAR image classification," *IEEE Geoscience and Remote Sensing Letters*, 2011, 8(2): 225–229.
- [14] M. A. Hearst, S. T. Dumais, E. Osuna, J. C. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, 1998, 13(4): 18–28.
- [15] Y. Guerbai, Y. Chibani, and B. Hadjadji, "The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters," *Pattern Recognition*, 2015, 48(1): 103–113.
- [16] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proceedings of the IEEE International Conference on Image Processing*, Thessaloniki, Greece, 2001, pp. 34–37.
- [17] J. Muñoz-Mari, F. Bovolo, L. Gómez-Chova, L. Bruzzone, and G. Camp-Valls, "Semisupervised one-class support vector machines for classification of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(8): 3188–3197.
- [18] S. Nam, W. Kim, S. Mun, J. Hou, S. Choi, and H. Lee, "A SIFT features based blind watermarking for DIBR 3D images," *Multimedia Tools and Applications*, 2018, 77(7): 7811–7850.

- [19] H. Bay, T. Tuytelaars, and G. L. Van, "SURF: speeded up robust features," in *Proceedings of the European Conference on Computer Vision*, Graz, Austria, 2006, pp. 404–417.
- [20] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(1): 105–119.
- [21] R. Velazquez-Pupo, A. Sierra-Romero, D. Torres-Roman, Y. V. Shkvarko, J. Santiago-Paz, and D. Gomez-Gutierrez, "Vehicle detection with occlusion handling, tracking, and OC-SVM classification: a high performance vision-based system," *Sensors*, 2018, 18(2): 374–396.
- [22] H. Chaabouni-Chouayakh and M. Datcu, "Coarse-to-fine approach for urban area interpretation using TerraSAR-X data," *IEEE Geoscience and Remote Sensing Letters*, 2010, 7(1): 78–82.
- [23] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(4): 1156–1167.
- [24] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of lidar data and building object detection in urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014, 87: 152–165.
- [25] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 117: 11–28.
- [26] D. Zhu, B. Wang, and L. Zhang, "Airport target detection in remote sensing images: a new method based on two-way saliency," *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(5): 1096–1100.
- [27] B. Hou, Y. Wang, and Q. Liu, "A saliency guided semi-supervised building change detection method for high resolution remote sensing images," *Sensors*, 2016, 16(9): 1377–1398.
- [28] Y. Hu, X. Hu, P. Li, and Y. Ding, "Building detection from orthophotos using binary feature classification," *Multimedia Tools and Applications*, 2018, 77(3): 3339–3351.
- [29] N. Chandra and J. Ghosh, "A cognitive viewpoint on building detection from remotely sensed multispectral images," *IETE Journal of Research*, 2018, 64(2): 165–175.
- [30] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proceedings of the 9th European Conference on Computer Vision*, Graz, Austria, 2006, pp. 430–443.
- [31] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: a decade survey of instance retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(5): 1224–1244.
- [32] Y. Yu and J. Yang, "Visual saliency using binary spectrum of Walsh-Hadamard transform and its applications to ship detection in multispectral imagery," *Neural Processing Letters*, 2017, 45(3): 759–776.
- [33] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037–2041.
- [34] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proceedings of the IEEE International Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 603–610.
- [35] G. Cheng, J. W. Han, P. C. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Transactions on Image Processing*, 2019, 28(1): 265–278.
- [36] G. Cheng, P. C. Zhou, and J. W. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(12): 7405–7415.