Journal of the Indian Society of Remote Sensing (June 2020) 48(6):831–839
https://doi.org/10.1007/s12524-020-01117-4

RESEARCH ARTICLE

Suburban Building Detection from Optical Remote Sensing Images Based on a Deformation Adaptability Model

Fukun Bi[1] · Jie Zhang[1] · Fengqian Pang[1] · Mingming Bian[2] · Yanping Wang[1]

Received: 4 February 2020 / Accepted: 28 May 2020 / Published online: 10 June 2020

Abstract
Automatic detection of suburban building areas (SBAs) is one of the research hotspots for remote sensing images (RSIs), which are widely used in the dynamic monitoring of land use, illegal building monitoring, anti-terrorism, etc. However, because the buildings are distributed targets and their appearances are quite different, the current mainstream detection methods have difficulty obtaining good detection results. To improve the detection performance, this paper presents an SBA detection method based on a deformation adaptability model. It can be divided into two main stages. (1) During key structure (KS) extraction, we first obtain building potential area by Mask R-CNN, and then we extract KS from building potential area based on roof parallel structures (RPSs) to achieve the rapid extraction of KSs. (2) In candidate region identification, to make full use of the relationship among the key structures of buildings, we present a deformation adaptability model, which adapts well to distributed targets with different appearances, and it also has strong resistance to intra-class deformations and a strong ability to eliminate false alarms. We test the proposed method on both general and complex datasets, and the experimental results show that our method has a higher detection accuracy and efficiency than typical methods.

Keywords Deformation adaptability model · Building detection · Remote sensing image · Deep learning

Introduction

With the rapid development of remote sensing technology, remote sensing image processing plays a crucial role in various fields. Regarding the development of the national economy and owing to the sharp increase in commercial land resources and the rapid decrease in cultivated land resources, dynamic monitoring of land use is particularly important. For anti-terrorism, because most terrorist camps are located on plateaus, the traditional patrol methods are slow and inefficient. Therefore, designing an automatic method is challenging.

Over the last few decades, research towards the automatic detection of SBAs from optical remote sensing images has gained significant attention, and the main methods are summarized as follows:

The most common SBAs detection method is based on line segment detection. von Gioi et al. (2010) proposed a line segment method that combines the gradient information between pixels to form line segment. A line segment extraction method, proposed by Burns et al. (1986), divides the pixels with the same gradient direction into the same linear support region and then uses the correlation structures to determine the location and properties of the edge. However, in RSIs there are many line segments, and the combinations of these segments are complex and variable; hence, the detection performance of the above methods is low. Moreover, this kind of method has difficulty distinguishing the excessive interference produced by vegetation regions, and it will greatly increase the false alarm rate. In addition, because a certain SBA is composed of several monomer buildings, and the monomer buildings are usually distributed targets, the ordinary line segment detection method cannot adequately describe the topological structures, which easily leads to miss targets.

✉ Jie Zhang
13683590123@163.com

[1] School of Information Science and Technology, North China University of Technology, Beijing 100144, China

[2] Beijing Institute of Spacecraft System Engineering, Beijing 100094, China

Research on local descriptor-based matching methods is also deeply needed. An airport detection method using clustered SIFT key points and region information to detect airports from large IKONOS images, was proposed by Tao et al. (2011), this method uses a series of SIFT key points to describe an airport, and an improved SIFT matching strategy is used to detect airports. However, it considers only the local structural information of buildings and does not consider whole structures. This method is also time consuming because it needs to scan the whole input image to obtain the key points. Moreover, when SBAs are not large targets, unlike airports, a large number of false alarms will be generated when it is applied to objects that are similar to the local features of the target. Recently, we have also used related technologies for designated building area, but it is very time consuming and can easily attain many false alarms.

Recently, a method based on the strict spatial topology in RSI building detection has been developed. Chaabouni-Chouayakh and Datcu (2010) proposed a 'coarse-to-fine' building detection method, which divides the scene into the same characteristic region and then trains the structure extractor to complete building detection. However, the limitation of this method is that it can detect only building area patterns in the strict spatial topology. Sirmacek and Unsalan (2010) used local features and spatial voting to detect buildings, extracted features such as edges and key points from different angles, and then vote on the candidate areas to determine the location of buildings. However, the above two methods are strictly constrained by the azimuth topological relationship of SBAs, and the detection results are affected by the density and arrangement of the buildings. Because each detection method applies a special topology, it is not resistant to intra-class changes, and it cannot adapt to multi-topologically distributed targets, which will ultimately affect the detection efficiency.

Currently, deep learning models have been widely applied to object detection (Girshick et al. 2014). However, these methods need to scan and identify each local area in the whole image with high computational costs, and the building area belongs to the distributed target without specific appearance characteristics, which is not suitable for detection by deep learning-based methods. In addition, this kind of methods requires a very large training sample. Although the deep learning method is not suitable for distributed targets detection in large-scale suburban remote sensing images, we might can use deep learning networks to segment remote sensing images to quickly obtain the building potential area, and it can greatly reduce the amount of computation and improve the detection efficiency.

To overcome these limitations, we propose a novel building area detection method based on deformation adaptability model that includes two main contributions. (1) During key structure(KS) extraction, we use Mask R-CNN (He et al. 2017) to obtain building potential area firstly and then extract KS from building potential area based on roof parallel structures (RPSs); this strategy can quickly extract candidate building areas from the whole image, reduce the computational complexity of subsequent steps, and improve the detection efficiency. (2) In identification stage, we propose a DAM, we first build primitives descriptor based on topological configuration, then we present the pattern matching based on azimuth multi-structure, and we last present the pattern judgment based on adaptability model. The DAM covers most of the topological relationships of primitives to target candidate areas, and it also has strong resistance to intra-class deformation, which occurs because of the appearance.

Proposed Method

This method consists of the following two parts: extraction of key structures from building potential area and candidate region identification based on a deformation adaptability model.

Extraction of Key Structures from Building Potential Area

In the RSIs, buildings are sparsely distributed and usually appear in the form of building areas. To facilitate the subsequent description, we defined a monomer SBA as a target to be detected, and each monomer building in the SBA is defined as a KS. Because of the sparse distribution, most areas in the field to be detected are not building areas; hence, we first use Mask R-CNN to extract the building potential area from the sampled input images, and then extract the line segments using LSD, quickly acquiring the KSs of the candidate areas. Regardless of the satellite observation angle, the buildings always have parallel line structures. Therefore, we conduct a rapid screening method based on roof parallel structures (RPSs), and apply it to building potential area to obtain complete KSs of building areas.

Building Potential Area Extraction Based on Mask R-CNN

Although the deep learning method is not suitable for distributed targets detection in large-scale suburban remote sensing images, we can use deep learning networks to segment remote sensing images to quickly obtain the building potential area. Firstly, we apply Mask R-CNN on the sampled input images to obtain building potential area, and this strategy can reduce the amount of calculation of the subsequent stages and improve the detection efficiency. The method process and the building potential area are shown in Fig. 1.

Mask R-CNN is a target segmentation convolution neural network by adding mask branches on the basis of Faster-Rcnn. The network process consists of three parts: RPN candidate area extraction network, FPN target detection network, FPN-Mask segmentation network, as shown in Fig. 2.

RPN Candidate Area Extraction Network The purpose of RPN candidate extraction network is feature extraction and building potential area extraction. RESNET network is used for main feature extraction; this feature is not only used for extracting RPN building potential area, but also used for subsequent FPN target detection network and FPN-Mask segmentation network, realizing the sharing of main network features. In the extraction of RPN candidate region, ROIAlign is used instead of ROIPooling operation, which uses bilinear interpolation to complete pixel level alignment, making subsequent segmentation more accurate.

FPN Target Detection Network FPN target detection network references FCN segmentation network and SSD target detection network, which uses feature pyramid and level by level fusion to detect the target and regress the target box, among them, the features pyramid processes the specific level of main network to get pyramid features of different levels, which are used for subsequent target detection and box regression after fusion.

FPN-Mask Segmentation Network FPN-Mask segmentation network also uses FPN pyramid features for further convolution, up-sampling and other operation, to generate *K* (number of categories) segmentation masks for each ROI, which can effectively avoid intra-class competition. Finally, through the FPN target detection network results and non maximum suppression to confirm and locate the segmented target, the final segmentation results can be obtained.

Line Segment Extraction Based on LSD

LSD is an automatic line detection and segmentation algorithm that has strong anti-interference robustness and low algorithm complexity. We conduct LSD detection on the building potential area extracted in Sect. 2.1.1.

The LSD operation is as follows. (1) Fourfold Gaussian downsampling is performed on the original image. (2) Gradient calculation of image on 2 × 2 template, the template is chosen to be as small as possible to reduce the dependence of each pixel in the gradient calculation process. (3) The gradient values are sorted, and the points with larger gradient values are selected as seed points in each bin. (4) During the calculation process, pixel points whose gradient amplitude is less than the threshold will not be allowed to participate in the formation of the line support region. (5) The region growth algorithm is applied to generate a line support region. (6) A rectangular box containing a series of discrete line support region points will be determined in step five. (7) The NFA is calculated.

KS Rapid Screening Based on Roof Parallel Structures

The shape and size of the building in the RSIs of building areas are quite different, and because of the influence of satellite observation angle, light, occlusion and other factors, most of the roofs of the buildings do not have a standard rectangular shape, but most have parallel lines. After LSD, a large number of line segments will be generated in the image, and we can quickly screen building potential area by using special morphological structures: RPSs.

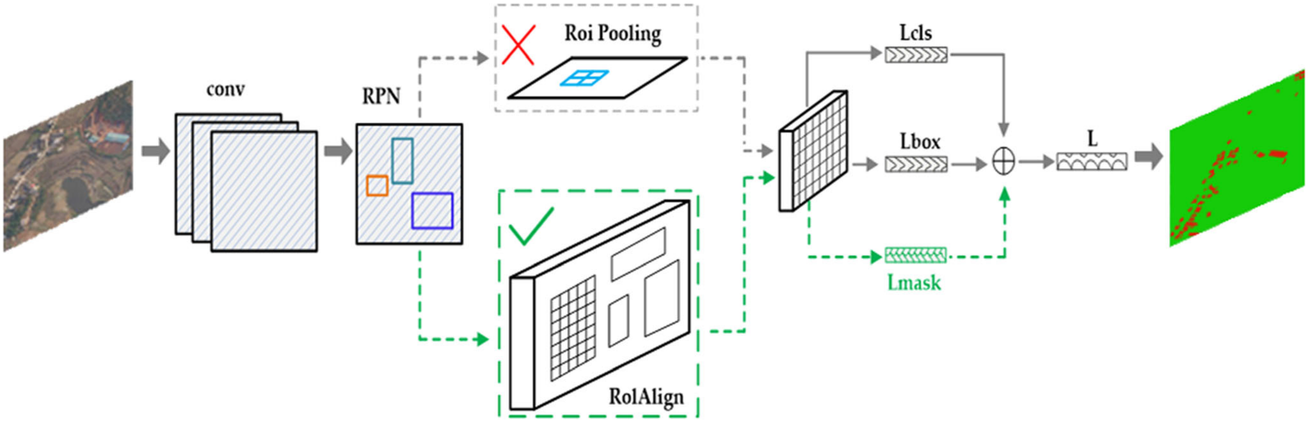

Fig. 1 The process of Mask R-CNN, and the red box represents building potential area (color figure online)

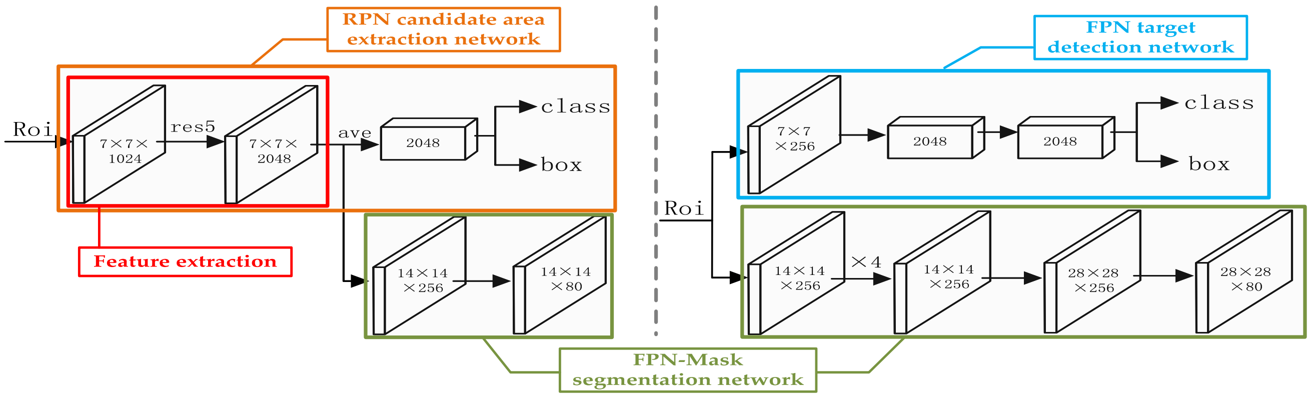

Fig. 2 The three parts of the network process: RPN candidate area extraction network, FPN target detection network, FPN-Mask segmentation network

To analyse the geometric relations in the set of line segments, the following definitions are given first. Let L be a set of line segments, where $l_i, l_j \in L$, and $i,j \in N$. Let D be a set of distances between two straight lines, where $d_{i,j} \in D$. T_d is the threshold value, i.e., the maximum distance between the two lines. The specific steps are as follows: the distance is valid if and only if $d_{i,j} \leq T_d$; otherwise, it is invalid. According to the definition, if there are two parallel lines $l_i \in L$, $l_j \in L$ and $d_{i,j} \leq T_d$, the area composed of l_i, l_j satisfies the assumption of the building shape, and then the ends of the two lines connect to form a complete quadrilateral structure, this area will eventually be identified as a KS, as shown in Fig. 3.

Building Potential Area Identification Based on a Deformation Adaptability Model

The DAM is composed of three parts: primitives descriptor based on topological configuration, pattern matching based on azimuth multi-structure and pattern judgment based on adaptability model, as shown in Fig. 4. The monomer primitive descriptor is proposed to describe the topological configuration, and to reduce the problem of intra-class deformation. The pattern matching based on adaptability model elastically describes the positional relations among the monomers, to improve the ability of anti-intra-class deformation.

Primitives Descriptor Based on Topological Configuration

Due to the sparse distribution of KSs in RSIs and the large differences in appearances, it is necessary to describe various topological structures with simplified and reliable models. SBAs are usually composed of several monomer KSs; therefore, a monomer primitive is proposed to describe the monomer KS, and then the intra-class deformation caused by the appearance of the KSs can be reduced. The method is as follows: we have a series of KSs after that in Sect. 2.1.2, determine the centroid of each KS, and then replace the position and azimuth of KSs with the centroid to form the monomer primitive, as shown in Fig. 5.

Pattern Matching Based on Azimuth Multi-structure

Primitive usually have a typical layout, such as an L-type, I-type, M-type layouts, as shown in Fig. 5b. It is found by analysis that if we construct models for each arrangement,

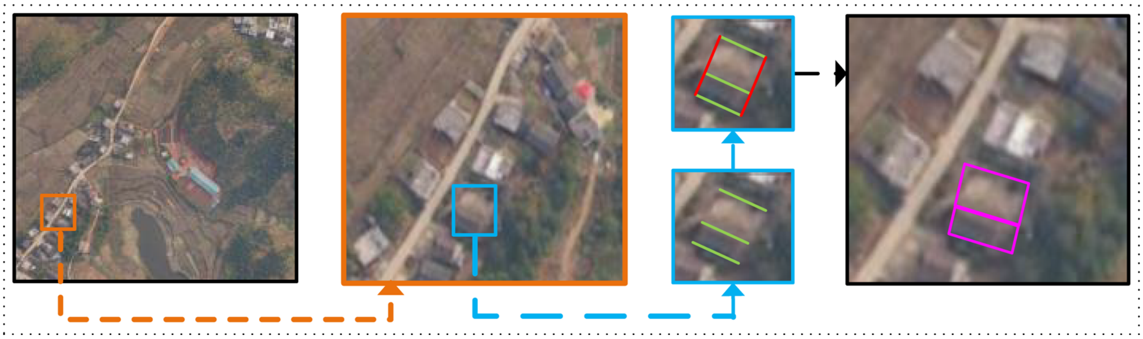

Fig. 3 The process of generating KSs. The green lines show the detected parallel lines, the red lines are obtained by connecting the ends of the green lines, and the pink quadrilateral structure represents the detection results (color figure online)

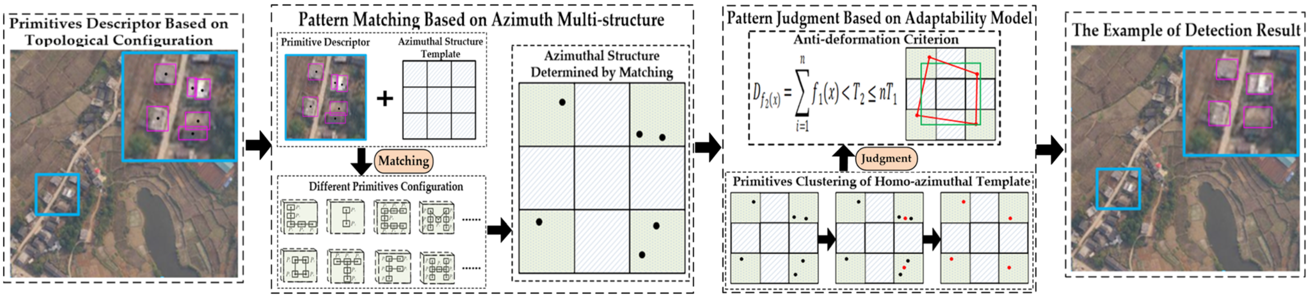
Primitives Descriptor Based on Topological Configuration
Pattern Matching Based on Azimuth Multi-structure
Primitive Descriptor
Azimuthal Structure Template
+
Matching
Different Primitives Configuration
Azimuthal Structure Determined by Matching
Pattern Judgment Based on Adaptability Model
Anti-deformation Criterion
$D_{f_2(x)}=\sum_{i=1}^{n} f_1(x)<T_2\leq nT_1$
Judgment
Primitives Clustering of Homo-azimuthal Template
The Example of Detection Result

Fig. 4 The process of candidate region identification based on DAM

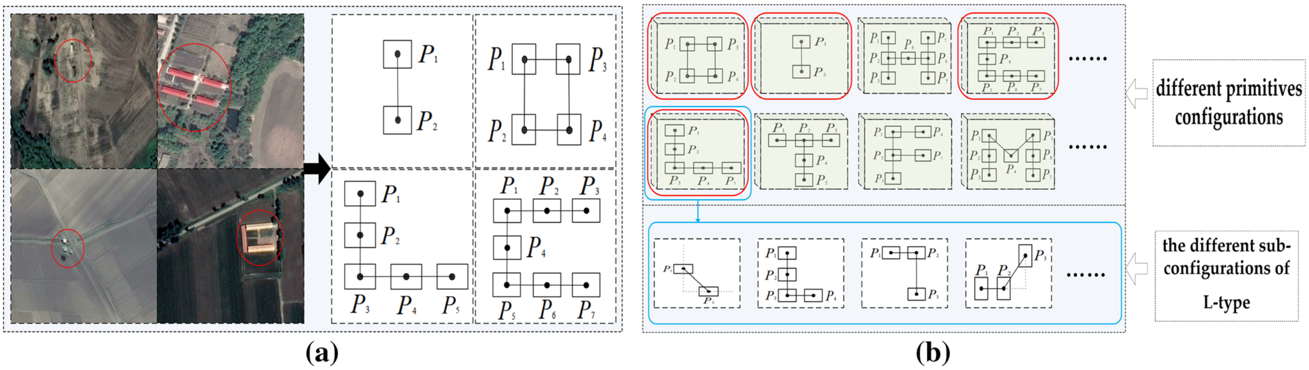
P1
P2
P3
P4
P5
P6
P7
different primitives configurations
the different sub-configurations of L-type
(a)
(b)

Fig. 5 The process of primitives descriptor of topological configuration. **a** Represents the primitive descriptors corresponding to different remote sensing images. **b** Represents a series of different primitives configurations, and each primitive configuration has several sub-configurations. The red box denotes topological configurations in **a**, and the blue box denotes the different sub-configurations of L-type (color figure online)

the number of models will be large. To describe the topological relationships concisely and to cover as many types of topological configurations as possible, we construct an azimuth multi-structure pattern-matching procedure.

We construct a nine-lattice azimuthal structure template (AST), which consists of nine azimuth templates (ATs) of the same size. It is activated when the primitives are matched in each AT. All activated ATs form an AST, and all ASTs form the azimuth structure patterns. The azimuth structure patterns of primitives can be obtained by training. Even if there are many topological modes in an SBA, they can be expressed by the simplified model after learning. Because of the different resolutions of the RSIs, the size and scale of building area patches are different. To express these patterns uniformly, we normalize the extracted primitives to the same scale.

Pattern Judgement Based on Adaptability Model

Primitives Clustering of Homo-azimuthal Template Azimuth primitives represent the azimuth topological relationship between primitives, and primitives located within the same AT have similar azimuthal descriptions. When the number of primitives to be matched in each AT is greater than one, we cluster them, and the specific operation is as follows: when more than one primitive is detected in each AT, obtain the clustering centroid of these primitives, and then this centroid becomes the new primitive.

Deformation Adaptability Judgement After the above steps, there exists at most one fixed primitive in each azimuth template. Because the observation angle is changeable, the matching model needs to have a certain deformation adaptability; hence, an adaptability model is designed to improve the ability to resist intra-class deformation in this paper, as shown in Fig. 4.

For quantitative descriptive variables, the single primitive deformation and cumulative primitive deformation are defined as follows:

$$D_{f_1(x)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

$$D_{f_2(x)} = \sum_{i=1}^{n} f_1(x) \qquad (2)$$

where the single primitive deformation $D_{f_1(x)}$ is expressed as the displacement in the position between the actual primitive of a monomer and the standard primitive of a monomer, the accumulation primitive deformation $D_{f_2(x)}$ is expressed as the accumulation displacements of all primitives in the SBA.

Based on the above two deformations, we propose an anti-deformation criterion, which is shown below:

$$D_{f(x)} = \sum_{i=1}^{n} f_1(x) < T_2 \leq nT_1 \tag{3}$$

When the offset between each primitives and standard position is less than threshold T_1, and the accumulation offset of all primitives is less than threshold T_2, all primitives to be detected are determined as a target.

Results and Discussion

Datasets and Indicators Definition

In this paper, our experimental datasets consist of publicly available Google Earth and GaoFen-2 satellites with 1200 optical images, resolution ranges from 0.8 to 2 m, and size ranges from 2000 × 2000 pixels to 15,000 × 15,000 pixels. There are 820 images in the general datasets, most of which are single monomer or simple-structured buildings, the remaining 380 images in the complex datasets, and most SBAs in complex datasets are clustered by several monomer buildings. Two hundred images are randomly selected from the two datasets at a ratio of 3:1, respectively. And the building area patches are captured as positive samples; moreover, the confusing not building patches are labelled as negative samples, such as paths and farmland; some typical examples are shown in Fig. 6. All the experiments were programmed using MATLAB 2016a. The experimental platform was a PC with a 3.6 GB CPU and 64 GB of server memory.

We use the precision and recall as indicators for evaluating our method.

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{4}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{5}$$

where TP represents the number of real buildings extracted by this method, FN represents the number of real buildings labelled manually but not detected by this method, and FP represents the number of incorrect buildings extracted by this method.

Comparative Experiments and Analysis

We compare the typical methods of building area detection mentioned above with the proposed method, to show the feasibility and competitiveness of our method. All methods are trained and tested on the datasets described in Sect. 3.1, and we adjust the parameters of each method to the optimum. Specifically, we also added a comparative experiment, comparing our method with a DPM, which is a kind of method with strong anti-intra-class deformation ability. The test results are shown in Fig. 7.

As shown in Fig. 7, our method achieves good test results in both types of datasets. von Gioi et al. (2010) proposed that the line segment detection method has a higher false alarm in both datasets, which may be because this kind of method has difficulty distinguishing the excessive interference produced by vegetation regions, and it will greatly increase the false alarm rate. Tao et al. (2011) suggested that the local description method easily misses targets because this kind of method only acquires local regions and lacks overall structures, and the building areas of distributed targets are easily missed. The DPM

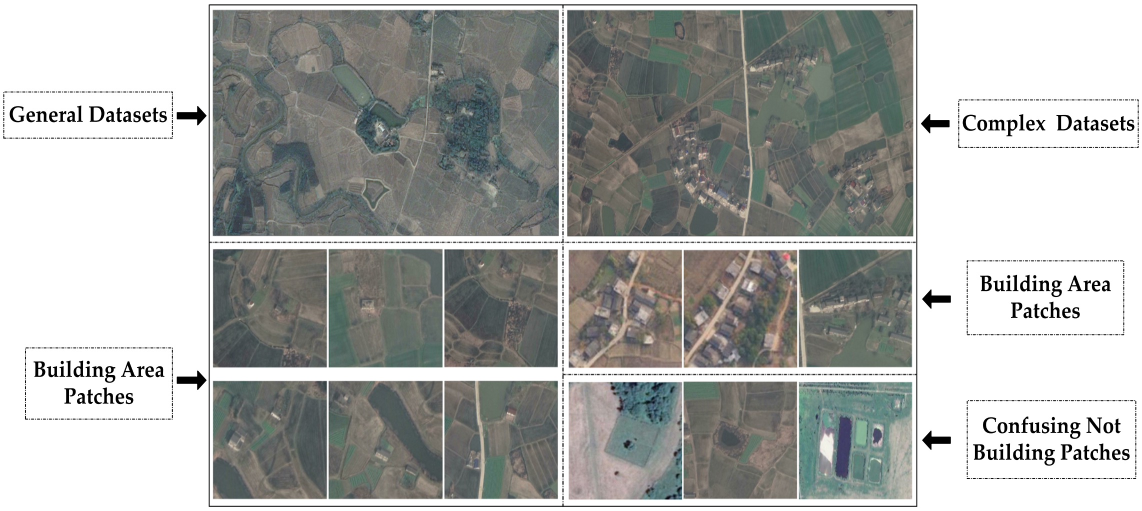

Fig. 6 General datasets and complex datasets. The example of building area patches and confusing not building patches

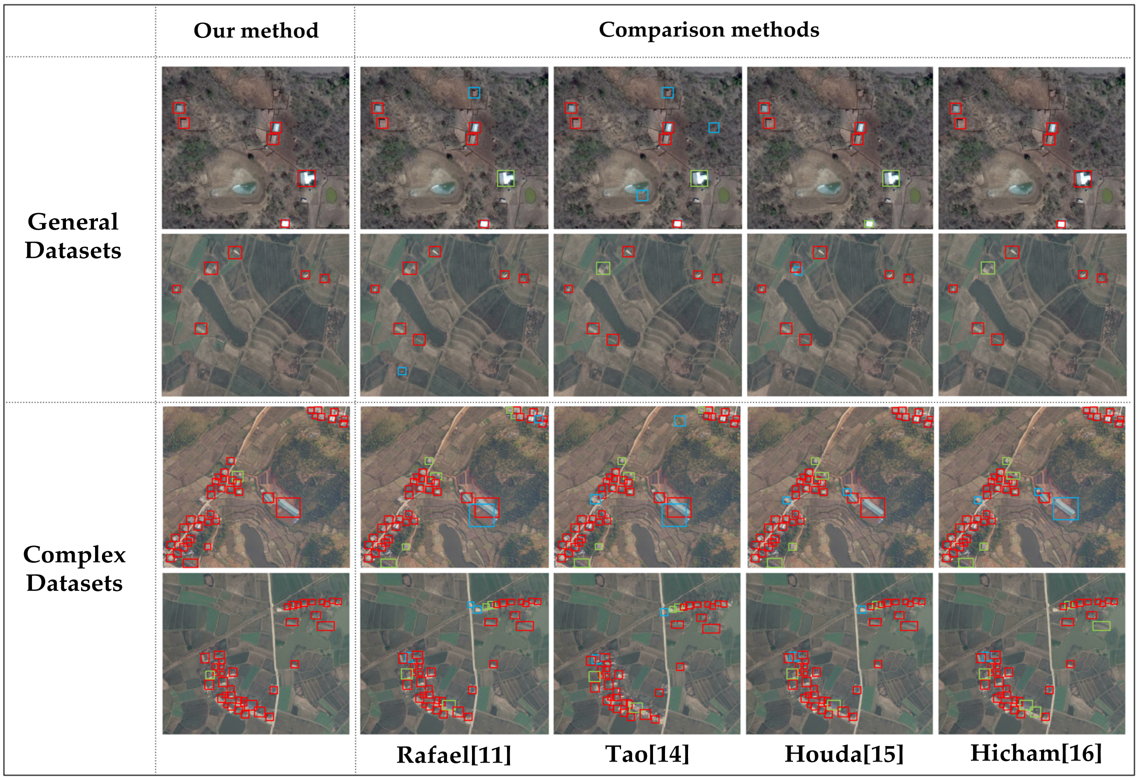

Fig. 7 The typical test results of our method and the comparison method. The red boxes indicate the correctly detected building targets, the green boxes indicate the missed building targets, and the blue boxes indicate the false alarm (color figure online)

method proposed by Randrianarivo et al. (2013) has a low detection rate in complex datasets, and it has a low ability to resist intra-class deformations.

The detection time for the above five methods are shown in Table 1.

As shown in Table 1, our method takes the least amount time on both types of datasets. This phenomenon is related to our efficient building potential area screening strategy and the simple and effective identification strategy. Tao et al. (2011) suggested that the local description method takes the longest time in both datasets, which is probable because the key points need to be extracted and the corresponding descriptors are calculated from the full image. The DPM method proposed by Randrianarivo et al. (2013) also takes a long time in the two datasets, because of the complexity of the DPM model itself.

The PR curves for both two datasets are shown in Fig. 8.

As shown in Fig. 8, our method has good adaptability in both types of datasets. Compared with our method, von Gioi et al. (2010) proposed that the line segment detection method has a higher false alarm in both datasets, which may be because this kind of method extracts the line segments directly from the full image and is disturbed by false alarms in the complex view. The local description method proposed by Tao et al. (2011) has a low detection rate and a high false alarm rate in both datasets, this is probable because this kind of method only acquires local regions and does not consider the overall structures. The accuracy dropped significantly when the target is not a large target, unlike airports. Chaabouni-Chouayakh and Datcu (2010) proposed that the method based on a strict spatial topological structure has a low detection rate in the complex

Table 1 The detection time for our method and the five comparative methods

	Our method	von Gioi et al. (2010)	Tao et al. (2011)	Chaabouni-Chouayakh and Datcu (2010)	Hicham (Randrianarivo et al. 2013)
General (min)	0.562	0.833	3.75	1.875	3.125
Complex (min)	0.697	0.958	5.13	2.625	3.937

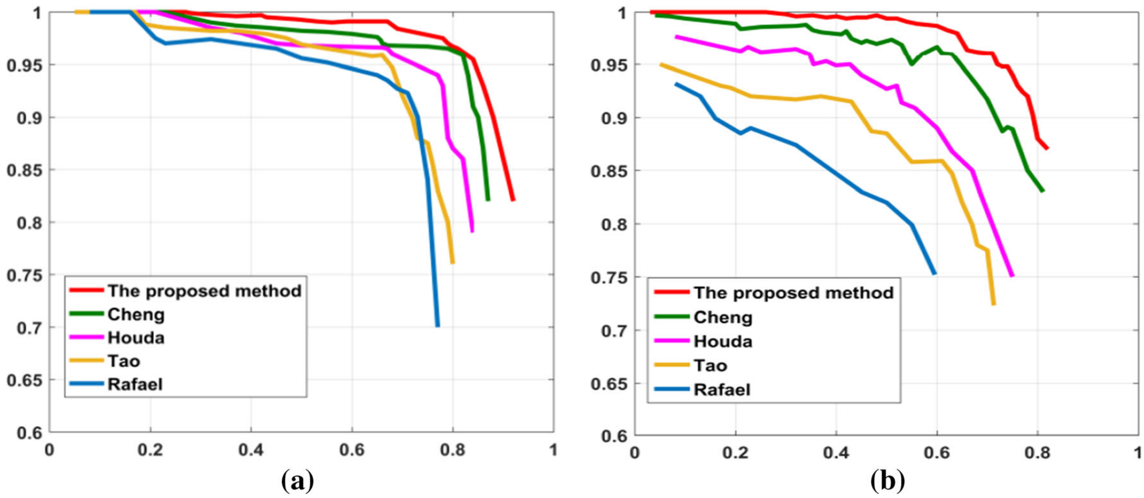

Fig. 8 The PR curves for both two datasets

datasets, and this method has poor resistance to intra-class deformations and is not suitable for distributed target detection with azimuth multi-structures, which may be due to the strict constraints of the azimuth topological relationship. Randrianarivo et al. (2013) showed that the DPM method has a low detection rate in complex datasets, and it is not suitable for distributed target detection with azimuth multi-structures and can easily to miss targets, which might be caused by its limited ability to resist intra-class deformation and its single algorithm model.

Conclusions

In this paper, we proposed an efficient detection method for suburban building areas. Because our method adopts the extraction of KSs, it takes little time and has a high efficiency. We use the pattern matching based on azimuth multi-structure to describe the topological relationship concisely, and to cover as many types of topological configuration as possible, thus reducing the false alarm rate caused by difference in appearance. We then proposed pattern judgement based on adaptability model, to improve the ability to address intra-class deformations; furthermore, our method can better adapt to distributed target detection and improve the detection rate. To demonstrate the performance of this method, we established multitype test datasets. As seen from the experimental results, in terms of detection performance, our method achieve good test results in both types of datasets, because our method has strong resistance to intra-class deformations, and it has good adaptability to distributed targets with different appearances. In terms of the detection time, because we use Mask-Rcnn, which can quickly extract building potential area, and then we propose KS screening method based on roof parallel structures, which can reduce the computational time. Nevertheless, our method still has some limitations. Although our experiment achieved good detection in suburban remote sensing images, its detection performances in other types of images, such as synthetic-aperture radar (SAR) and infrared images, have not been confirmed. Our future work will concentrate on this topic.

Funding This work was supported in part by the National Natural Science Foundation of China (Grant no. 61971006), the National Natural Science Foundation of China (Grant no. 61601006), the Science and technology innovation service capacity building of NCUT (No. 1921/007), the Equipment Pre-Research Foundation (61404130312) and Beijing Natural Science Foundation (No. 4192021).

Compliance with Ethical Standards

Conflict of interest The authors declare no conflicts of interest.

References

Burns, J. B., Hanson, A. R., & Riseman, E. M. (1986). Extracting straight lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(4), 425–455.

Chaabouni-Chouayakh, H., & Datcu, M. (2010). Coarse-to-fine approach for urban area interpretation using TerraSAR-X data. *IEEE Geoscience and Remote Sensing Letters., 7*(1), 78–82.

Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). 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, pp. 580–587.

He, K., Gkioxari, G., Dollar, P. (2017). Mask R-CNN. IEEE international conference on computer vision (ICCV), pp. 2980–2988.

Randrianarivo, H., Saux, B. L., & Ferecatu, M. (2013). Urban structure detection with deformable part-based models. *IEEE*

International Geoscience and Remote Sensing Symposium-IGARSS., 2013, 200–203.

Sirmacek, B., & Unsalan, C. (2010). Urban area detection using local feature points and spatial voting. *IEEE Geoscience and Remote Sensing Letters., 7*(1), 146–150.

Tao, C., Tan, Y., Cai, H., & Tian, J. J. (2011). Airport Detection from large IKONOS images using clustered SIFT keypoints and region information. *IEEE Geoscience and Remote Sensing Letters., 8*(1), 128–132.

von Gioi, R. G., Jakubowicz, J., Morel, J.-M., et al. (2010). LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence., 32*, 722–732.

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