IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 5, 2018, accepted March 12, 2018, date of publication March 27, 2018, date of current version May 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2819705

Building Extraction From RGB VHR Images Using Shifted Shadow Algorithm

XIANJUN GAO[1], MINGWEI WANG[2], YUANWEI YANG[1], AND GONGQUAN LI[1]
[1]School of Geoscience, Yangtze University, Wuhan 430100, China
[2]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

Corresponding author: Yuanwei Yang (yyw_08@whu.edu.cn)

This work was supported in part by the CRSRI Open Research Program under Grant CKWV2017537/KY, in part by the Open Fund of Key Laboratory for National Geographic Census and Monitoring, National Administration of Surveying, Mapping and Geoinformation under Grant 2016NGCM07 and Grant 2017NGCM07, and in part by the Yangtze Youth Fund under Grant 2016cqn04.

ABSTRACT Building extraction from RGB VHR images is an important and popular topic for mapping, disaster emergency responding, and city management. The automation of most methodologies cannot meet the need for applications. In this paper, based on classification and optimization, we propose a novel methodology using shadows to automatically extract building samples and verify buildings accurately to improve automation and accuracy. On one hand, in order to acquire various and reliable building samples automatically for classification, detected shadows first are shifted opposite to the direction of illumination to extract building shadows. Furthermore, each building shadow will be shifted again in the same way. Then according to the distribution of classes in these customized shifted regions, building samples can be filtered out by removing those recognized objects. On the other hand, besides the common measures to optimize the initial building during post-processing; a new, original, and an efficient shadow-based index for building verification is also designed. Shadow rate on the intersect boundary between the expanding edge of candidate regions and their shifted regions following the illumination direction can efficiently recognize buildings. When the proposed method is compared to other sample acquisition methods based on shadow, experimental results show that the approach for building samples acquisition is helpful to get accurate initial building results. Moreover, in comparison with other building extraction methods, the proposed building verification method can distinguish buildings from non-buildings. This significantly improves the accuracy of the final results. Numerical assessments performed on a series of test images indicate that our proposed approach for building extraction is efficient and feasible, especially in suburban areas.

INDEX TERMS Building extraction, classification and post-processing, shifted shadow algorithm, automatic building samples extraction, shadow-based verification.

I. INTRODUCTION

Buildings automatically or semi-automatically extracted from very-high-resolution (VHR) images can minimize the human work burden in the application of continuous updates on city planning and cartographic mapping, rapid responses to civilian and military emergencies, disaster planning and management or many other types of urban simulation [1]. It has been a popular research topic since sensors develop so quickly that the newest, detail information of buildings can be obtained conveniently. However, automatic building extraction from high-resolution images must address how to use automatic strategies to replace manual assistance to provide complex building information. Previous studies focusing on building detection in monocular images can be categorized from three main aspects, including graph-based methodology, active contour approaches and supervised machine learning approaches.

Graph-based approaches have been introduced by several researchers to detect buildings in monocular optical images. Kim and Muller [2] divided the process of building into four stages including extracting lines, generating line-relation-graph based on lines geometric relationship, detecting building hypothesis by the graph and lastly building verification by ground-level, medium-level and vertical-level lines. The approach was restrained by certain building shapes. Sirmacek and Unsalan [3] used scale-invariant feature transform (SIFT) in multiple subgraph matching to detect urban areas and then combined graph-cut algorithms to detect separate buildings. Both of them can only extract building hypothesis rather than accurate building boundaries.

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Ok [4] utilized shadow information to extract building candidate regions and then proposed a graph theory framework consisting of two partitioning levels to obtain accurate building regions. This approach relied on NIR band data and solar angles. If the source image cannot supply relative prior knowledge, it may not work efficiently.

Active contour is another effective approach for building detection. Ruther *et al.* [5] proposed a semiautomatic approach to delineate buildings by optimizing their approximate building contour position based on snake models and using elevation on DSM to verify. However, DSM was hard to acquire. Ahmadi *et al.* [6] proposed a novel active contour model to automatically detect buildings. The new model based on level-set formulation by adding two terms to Chan and Vese energy function was initialized by selecting sample data. However, it also needed priori knowledge about the number of building and background class. To sum up, the accuracy of active contour methodology is restricted by the capability of extracting building boundaries. The main reason is that feature points of those buildings without obvious difference from adjacent objects are difficult to detect. In addition, they are often used for buildings with regular shape especially rectangle [7]. Because of these restrictions, the approach is limited in application and the performance in different images or scenes is not very robust.

In comparison, supervised machine learning approaches are less limited by the scene, priori knowledge and shapes. Then they are widely studied and popularly used to detect buildings combining classification with optimization [1], [8]–[15]. Specifically, spectral, contextual, structural features (e.g. shape, size, height, morphological building index) and differential morphological profiles [16]–[18] extracted as example data of definite categories are entered into classifiers such as SVM, rule-based classification and fuzzy for training [19], [20]. Then the trained classifier can classify the image to extract initial buildings. After that, they are optimized through several post-processing strategies, such as region growth, morphology operations and so on [21], [22]. Additionally, shadow, vegetation, height and prior shape information such as rectangular or circular buildings, can be regarded as supportive information to verify buildings in the optimization process [22]–[24]. For instance, Sumer and Turker [25] proposed a GA-based building detection approach. Fisher's linear discriminant analysis and adaptive fuzzy logic controller operation were combined to find each potential building regions. Then, morphological post-processing were aimed to remove artefacts. The performance was better for urban and suburban buildings than rural test scenes in IKNOS image tests. Turker and Koc-San [19] proposed an integrated approach for rectangular- and circular- shape buildings. Building patches were detected from the image by SVM classification. Then the boundaries were constructed by Hough transformation and perceptual grouping by edge detection. This approach can be applied in any area but fails may still occur when buildings are too close or roofs are occluded.

Generally speaking, compared with other kinds of approaches, supervised machine learning has lesser limitation in principle. What is more, owing to the advantages of machine learning principle, the automation is promoted greatly than the ones of the other two approaches. However, training samples for each class are usually selected by manually, decreasing the automation of the approach. As a result, in order to improve the automatic of building extraction, it is important to find methods to acquire samples automatically. Besides, the initial building results always contain many non-buildings. Building verification in post-processing is vital to improve the accuracy. Given that shadow is adjacent to its relative building, it is useful to locate the building position. Moreover, it is easily to be detected automatically based on their spectral signatures. If the detected shadow location can be used to extract building samples properly, the automation would be improved. Additionally, as a common and significant evidence, shadow has been employed in many approaches for verification in post-processing [26]–[28]. Iwasaki and Yamazaki [29] measured shadow lengths before and after an earthquake to estimate building heights and detect the change in buildings. In the process, collapsed buildings can be identified automatically. Sirmacek and Unsalan [30] applied multiple building cues, including edge information, shadow information and invariant color features, in order to detect buildings whose roofs have the same color. However, these verification indexes were not definite, easy enough to be calculated to measure the relationship between building and its shadow. If suitable indices for building verification based on shadow can be designed, the automation and accuracy combining with supervised machine learning can be improved.

Therefore, in this paper, an original approach is proposed for building detection based on automatic SVM classification and post-processing [Figure 1]. An image is divided into four classes, i.e. *shadow*, *vegetation*, *building* and *bare land*.

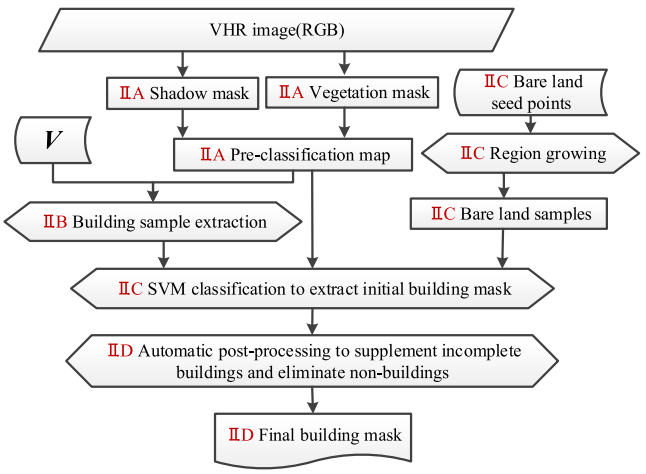

FIGURE 1. Flow chart of the proposed methodology for building extraction.

Training samples of shadows, vegetation and buildings are automatically acquired. Specifically, we design a novel and robust method that utilizes building shadow regions for building samples acquisition, as well as an estimated shadow shift vector to extract various supplementary building samples. In addition to the bare land samples collected by seed point selection and region growth, accurate samples of the four classes are trained to establish a suitable SVM classifier and to classify other unlabeled pixels. An initial building mask can be extracted from the classified results. This mask is then optimized furtherly via automatic post-processing. In the post-processing scheme, morphological operations, region growth and a newly designed building verification strategy based on shadow evidence are integrated to supplement some incomplete buildings and to remove non-buildings.

In this study, our main contribution with respect to building extraction includes two aspects:

(1) A novel method is developed to automatically acquire accurate building samples. Using automatic detected shadows, they are shifted to judge building shadows and furtherly extract each building regions where shifted building shadows are covered to acquire various building samples automatically.

(2) A new index to verify buildings with shadow information is designed. Making full use of the common border between buildings and its shadows, the shadow rate on such border can be effective to inspect whether the region is building or not.

II. THE PROPOSE METHOD

The proposed approach utilizes single ortho-rectified images that can provide RGB band data. Building extraction actually consists of two processes, i.e. classification to extract candidate buildings and post-processing to improve accuracy. Initially, vegetation and shadow masks are extracted automatically. The shadow mask is then used to acquire building samples by the new proposed shifted region strategy. Thereafter, the automatically obtained building, shadow and vegetation samples are used as the training data for the SVM classifier, along with the manually acquired bare land samples. The image can be classified into four classes $\{C_{SD}, C_{VE}, C_{BL}, C_B\}$, which represent shadow, vegetation, bare land and building respectively. Then, the initial building mask can then be extracted using the trained SVM model. Lastly, the final building mask can be constructed through post-processing operations, which consist of morphology operations to delete small non-buildings and to fill in building holes, region growth to supplement incomplete buildings and inspection to eliminate non-building regions. Each stage of our methodology is elucidated in subsequent sections.

A. AUTOMATIC VEGETATION AND SHADOW DETECTION TO FORM A PRE-CLASSIFICATION MAP

Shadows and vegetation are common objects existed in many images can be detected easily according to their unique spectral signatures respectively. Then we should conduct a pre-classification by assigning an initial label to each image pixel to produce a pre-classification map $C = \{C_{SD}, C_{VE}, C_U\}$, where C_U represents unlabelled pixels.

Shadows in images generally have low brightness, higher saturation and other spectral signatures. We use the methodology which combines typical spectral and newly designed shadow signatures with automatic threshold strategy to detect shadows accurately and automatically, as shown in equation (1). All the detected shadow pixels would be labelled as C_{SD}.

$$C_{SD} = \left\{ (i,j) \left| \begin{array}{l} \left(B'(i,j) > T\_B' \&\& I(i,j) < T\_I\right) \\ \left\| \left(Q(i,j) > T\_Q \&\& G'(i,j) < T\_G'\right) \right. \\ \| A(i,j) > T\_A \end{array} \right. \right\} \quad (1)$$

Where (i, j) is the pixel coordinate in the image. I represents the intensity feature in HIS space. B', G' are nominalized B and G feature respectively. Q and A are combined features designed by above simple features [defined in Equation (2) and (3)]. The thresholds of each feature are automatically calculated by Otsu algorithm [31].

$$Q = B' - I \quad (2)$$

$$A = \begin{cases} 2B' - I - G', & G' <= T\_G' \\ 2B' - I - 2G', & G' > T\_G' \end{cases} \quad (3)$$

Based only on R, G and B band data, *NDVI* cannot be used. Experimental results show that G' in vegetation is greater than other objects. Therefore, combining this feature and its Otsu threshold by equation (4), the vegetation can be automatically detected. Thereafter, all the detected vegetation would be labelled as C_{VE}.

$$C_{VE} = \{(i,j) | G'(i,j) > T\_G'\} \quad (4)$$

Then the rest unlabelled pixels would be labelled as C_U, and the pre-classification map can be obtained without any manual assistance. It would be used in the following process. In addition, shadow and vegetation samples can be easily extracted from automatic detected results.

B. SHIFTED SHADOW REGION ANALYSIS TO AUTOMATICALLY EXTRACT BUILDING SAMPLES

In this study, owing to class distribution in shifted shadow regions which reflects the category of the object where it belongs to, we proposed a shifted region strategy to automatically extract building samples. Firstly, shift vector $\boldsymbol{V}$ can be estimated according to direction of illumination in an image. All shadow regions would be shifted according to their customized $\boldsymbol{V}_k$. Secondly, since high rate of vegetation in such regions indicates these shadows belong to vegetation, by analyzing the class distribution in the shifted shadow regions according to pre-classification map, vegetation shadow can be filtered out to extract building shadows. In addition, shifting the building shadows again, except vegetation and shadow, the other part of the shifted shadow regions are likely to be buildings. Thereafter, the initial

building samples can be automatically obtained. Then through obtaining the intersection of the C_U in each shifted regions, the candidate building samples can be extracted. Lastly, they would be furtherly inspected with its spectral features to remove some interrupted non-building samples, especially vegetation. Finally, all the building sample regions for classification can be obtained.

1) PROPOSED SHIFTED REGION ALGORITHM USING SHADOW

Define a unit normal vector $\boldsymbol{V}_i$, parallel to the opposite direction of illumination representing, is the shifted unit vector. Actually it can be estimated using some corner points on the shadow edge and their corresponding points on the object edge. Generally, building corner and its shadow are easier to be found in an image. Therefore we propose a method to customize approximate shift vector $\boldsymbol{V}_k$ for each shadow region. These vectors can conduct the shadow region to a new location where part of the original object is covered. It is expressed in equation (5).

$$\boldsymbol{V}_k = \sqrt{\lambda l_k} \cdot \boldsymbol{V}_i, \quad k \in \{1, 2, \ldots, K\} \tag{5}$$

Where λ is a rate coefficient, k represents the index of the shadow region, K is the total number of shadow regions; and l_k is the perimeter of this shadow region. λ can be set by empirical value, [0.5,1.5] will be recommended. Then, a set of shift vectors $\boldsymbol{V}_S = (\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_K)$ in an image can be used to obtain corresponding shifted regions.

2) BUILDING SHADOW EXTRACTION

In fact, when a shadow region is shifted by a suitable distance to the opposite direction of illumination, the number of source object to which a shadow belongs is greater than the number of other object classes in the shifted shadow region. Therefore, in the shifted regions of vegetation shadows, vegetation would be the greatest class in the non-shadow objects. Then, combined with a pre-classification map where vegetation and shadow are detected firstly, vegetation shadows can be recognized by comparing the number between C_{VE} and C_U in each shifted shadow region. Most non-building shadows can be filtered out by eliminating these vegetation shadows and then building shadows can be obtained. Furthermore, building shadows will be shifted again and their distribution are analyzed to automatically extract building training samples.

3) INTEGRATION OF THREE DIFFERENT KINDS OF SHIFTED SHADOW REGIONS TO REMOVE C_{BL} *PIXELS*

In the shifted building shadow region, the class distribution is clarified with the aid of the pre-classification map. Most C_U pixels in this region denote a building, but sometimes a few of them belong to the bare land adjacent to the building. Therefore, as long as these interference regions can be removed, the remaining C_U regions can be regarded as building samples. Then, we design a strategy which integrates three different shifted shadow regions to determine their intersection in the C_U region. These three regions are determined using $\boldsymbol{V}_i$, $\boldsymbol{V}_{i\_left}$ and $\boldsymbol{V}_{i\_right}$. As displayed in Figure 2(a), $\boldsymbol{V}_{i\_left}$ and $\boldsymbol{V}_{i\_right}$ are unit vectors obtained by rotating $\boldsymbol{V}_i$ clockwise and counter-clockwise at the same angle θ, respectively. Hence, they can be calculated using equation (6).

$$\begin{aligned} \boldsymbol{V}_i &= (\sin\beta, \cos\beta) \\ \boldsymbol{V}_{i\_left} &= (\sin(\beta+\theta), \sin(\beta+\theta)) \\ \boldsymbol{V}_{i\_right} &= (\sin(\beta-\theta), \sin(\beta-\theta)) \end{aligned} \tag{6}$$

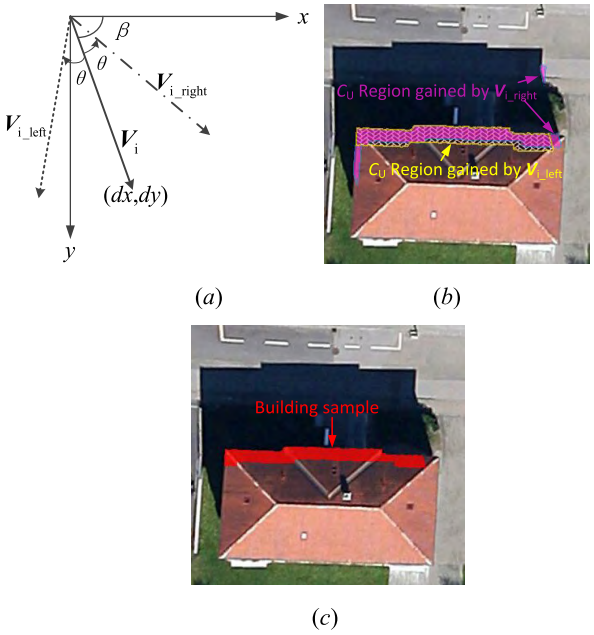

FIGURE 2. **(*a*) Schematic diagram that describes the relationship among shift vectors V_i, V_{i_left} and V_{i_right}. (*b*) Region in pink with an azure boundary denotes the C_U region in the shifted shadow regions as obtained using V_{i_right}. The region with a yellow boundary represents C_U regions in shifted shadow regions determined with V_{i_left}. (*c*) Region in red indicates the building sample gathered by integrating the intersection of the three kinds of C_U regions determined usingV_i, V_{i_left} and V_{i_right}.**

For example, the two kinds of C_U regions are determined using $\boldsymbol{V}_{i\_left}$ and $\boldsymbol{V}_{i\_right}$, as illustrated in Figure 2(b). The building sample can be then acquired by integrating the intersection of the three kinds of C_U regions [Figure 2(c)]. These regions indeed correspond to buildings, and the C_{BL} pixels near the building can thus be easily removed. In addition, some C_U regions which are far smaller than the other sample regions extracted in a single shifted building shadow region are also deleted to efficiently lower C_{BL} interference. Therefore, most C_{BL} samples can be eliminated in this step, and we can then obtain the initial building samples.

4) REMOVAL OF SOME VEGETATION SAMPLES

On occasion, a few vegetation shadows cannot be removed after building shadow extraction, resulted in a few vegetation samples may be misrecognized as building samples.

Therefore, each initial sample is inspected further according to the condition listed in equation (7) to remove these vegetation samples and to extract the most accurate building samples.

$$SAM_B = \{R_s||\bar{G}'_s - \bar{G}'_{\mathrm{All}}| < |\bar{G}'_s - T'_G|\&\&\bar{I}_s \geq \bar{I}_{\mathrm{All}}\} \quad (7)$$

Where s is the index number of the sample region R_s; SAM_B represents the set of building sample regions;$\bar{G}'_s$ and $\bar{I}_s$, $\bar{G}'_{All}$ and $\bar{I}_{All}$ are the mean values of features G' and I in the sample region and in the entire sample region, respectively; and T'_G is the automatic Otsu threshold of G'. The vegetation sample can thus be removed through this inspection. By applying the aforementioned strategy to each building shadow, we can automatically detect various and accurate building sample regions.

C. BUILDING INITIAL MASK EXTRACTION BASED ON SVM CLASSIFICATION

SVM algorithm is used to divide C_{U} into four classes, i.e. *shadows*, *vegetation*, *buildings* and *bare land*. The samples of each class would be gained and their features would be extracted and input to SVM classifier for training to extract initial building mask.

1) SAMPLE REGIONS EXTRACTION

In the previous supervised classification, samples of each class are selected manually. In response, we propose an approach to gather samples regions to reduce human effort. On the basis of shadow and vegetation automatic detection, their samples can be selected automatically using a pre-classification map. The collection of bare land sample regions alone requires some manual assistance by determining some seed points and integrating an automatic region-growth algorithm. Moreover, the most pivotal building sample regions are then obtained automatically and accurately through our designed method based on the shifted regions principle.

2) AUTOMATIC SVM CLASSIFICATION TO OBTAIN INITIAL BUILDING EXTRACTION RESULTS

SVM is a non-parametric classification technique based no statistical learning, which can identify and optimize the separating hyperplane of classes by adopting the structure risk minimization (SRM) principle [32]. We applied the SVM classification algorithm and selected feature vector to establish a classification model by training these samples. The trained classifier would be further used to distinguish which class the unlabeled pixel in the pre-classification map belongs to.

To begin with, we use Radial Basis Function (RBF) as the kernel function. The SVM algorithm parameters are set with empirical values to realize automatic classification. Secondly, direct spectral features are selected based on R, G, B bands. Apart from them, some transformation and operations such as subtraction, texture(LAWF) and HIS color transition are used to provide other spectral and texture attributes [25]. Then through principle component analysis, we choose suitable features to construct feature vector. Thirdly, according to the samples of four classes gathered above, we adopted a random extraction strategy to accelerate classification. Random function is utilized to automatically select partial sample pixels with the same features as final training samples. Then the total number of samples would decrease, but the sample species would not change. Furthermore, via inputting sample data, SVM classifier would be trained. Lastly, it would be used to classify the remaining unrecognized objects. Then an initial building mask can then be extracted from the SVM classification result [Figure 4(b)].

D. POST-PROCESSING FOR FINAL BUILDING EXTRACTION BASED ON SHADOW

Given the advantage of classification and various building samples, most buildings can be detected in the initial building mask. However, there are two deficiencies: (1) some buildings are poorly in completeness because they are recognized as other classes; (2) some non-buildings, especially bare land, may be recognized as buildings. Therefore, we conduct post-processing to improve completeness and correctness.

1) MORPHOLOGICAL OPERATIONS TO DELETE SMALL NON-BUILDING PIXELS AND TO FILL IN BUILDING HOLES

Small, scattered and incorrect building pixels can be removed by combining simple dilatation, erosion, closing and opening operations reasonably. Some holes in building roofs which are occasionally erroneously recognized can then be processed by filling in the internal contours attributed to buildings. Therefore, morphological operations can generally be used in the entire process of building extraction to delete small non-building pixels and fill in building holes.

2) REGION GROWTH TO SUPPLEMENT INCOMPLETE BUILDINGS

With regard to some buildings which are recognized partly, we utilized the region growth algorithm to supplement these incomplete building regions. We choose I, G and *Grey* as the feature parameters, and their values can be adjusted manually to obtain as better results as possible. Figure 4(c) indicates the results improved by morphological operations and with the region growth method. The findings suggest that the completeness of building regions has clearly improved.

3) PROPOSED BUILDING VERIFICATION INDEX BASED ON SHADOW

Because bare lands are more similar to some buildings in spectral features compared with other classes, they are the most non-buildings among initial building mask extracted from classification results. So we design a novel index to distinguish buildings from the candidate regions.

Given that buildings usually cast shadows whereas bare lands do not, we design a newly index η to recognize building based on shadow evidence. Specially, each candidate

building region (CBR) is shifted according to its shift vector $\boldsymbol{V}_R$ described in equation (8) with the same strategy of shifting shadow region.

$$\boldsymbol{V}_R = -\sqrt{\lambda l_R} \cdot \boldsymbol{V}_{\mathrm{i}} \tag{8}$$

Where $\boldsymbol{V}_R$ is opposite to the direction of illumination, l_R is the perimeter of CBR.

Define the intersection edge between shifted candidate building region (SCBR) and the boundary of enlarged candidate building region (ECBR) as $\mathrm{Edge}_{\mathrm{RSD}}$. As for any CBR to be judged, η represents the rate of the shadow on $\mathrm{Edge}_{\mathrm{RSD}}$, which is defined as follows

$$\eta = \frac{N_{SD}}{N_{RSD}} \tag{9}$$

Where N_{RSD} and N_{SD} are the total pixels number and shadow pixels number on $\mathrm{Edge}_{\mathrm{RSD}}$. As shown in [Figure 3], ECBR can be gained by dilate CBR once. If CBR is attributed to a building, part of the building shadow certainly covers most of $\mathrm{Edge}_{\mathrm{RSD}}$. By contrast, a corresponding shadow does not fall on its $\mathrm{Edge}_{\mathrm{RSD}}$ if the region represents bare land. It leads to η of buildings is much higher than that of bare lands. Thus, the η index can efficaciously recognize buildings. Then these non-buildings can be removed.

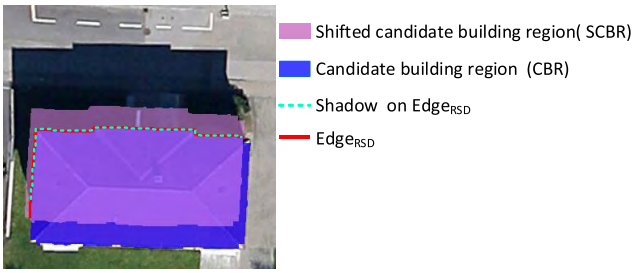

FIGURE 3. Schematic that describes the method of obtaining $\mathrm{Edge}_{\mathrm{RSD}}$ and the evidence that η efficaciously recognizes buildings. The blue region represents CBR, whereas the purple region (SCBR) is determined by shifting the building region according to V_{B}. The red solid line denotes $\mathrm{Edge}_{\mathrm{RSD}}$, which is the boundary of the dilated building region in the shifted building region. The azure dashed line is the shadow on $\mathrm{Edge}_{\mathrm{RSD}}$. If the region is bare land, then the length of the shadow line is almost 0.

Lastly, each candidate building region would be verified by the aforementioned index η. Those regions with low η would be removed. Experimental results show that 0.5 is an reasonable threshold for η. Moreover, we preserve all of the regions which have building samples to avoid erroneously removing some building regions. The final building mask can be extracted by integrating the aforementioned operations to post-process the initial results. Figure 4(d) shows that the proposed two level post-processing can automatically optimize the initial building results.

III. EXPERIMENTS AND RESULTS

A. INPUT DATASET AND ACCURACY ASSESSMENT STRATEGY

The results are assessed using the reference building extracted by a professional human operator.

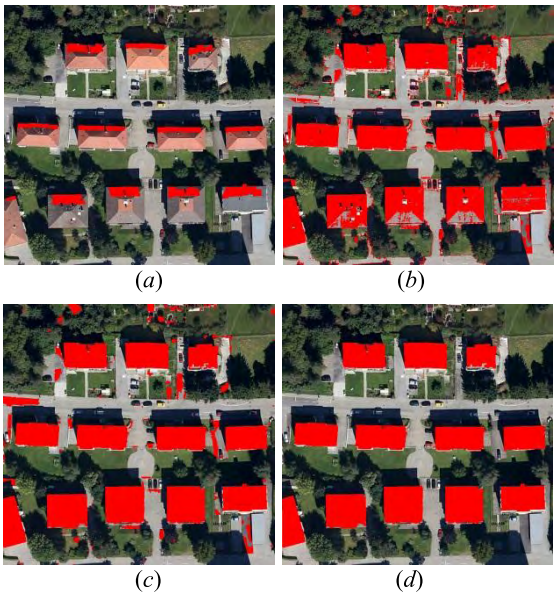

FIGURE 4. (*a*) The result for the building samples (Red). (*b*) corresponds to the initial building mask (Red) extracted from the classification result. (*c*) The result (Red) optimized by optimized through morphological operations and region growth. (*d*) The final result of building extraction, which are obtained by inspecting each region in (*c*) according to η.

To evaluate the result effectively, correctness, completeness and F-measure [33] illustrated in Equations (10)–(12), are used to evaluate the pixel-based and object-based performance of the results. |TP|, |FP| and |FN| represent the number of true positive (detected correctly), false positive (detected erroneously) and false negative (undetected) respectively. In the calculation of the object-based performance, if the pixel overlap ratio of each building object to its reference building is ranged in (0.6, 1.0], (0, 0.6) and 0, they will be labelled as TP, FP and FN respectively.

$$\text{correctness} = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FP})} \tag{10}$$

$$\text{completeness} = \frac{(\mathrm{TP})}{(\mathrm{TP}) + (\mathrm{FN})} \tag{11}$$

$$\text{F-measure} = \frac{2 \times (\text{correctness}) \times (\text{completeness})}{(\text{correctness}) + (\text{completeness})} \tag{12}$$

B. RESULT ANALYSIS

1) ASSESSMENT OF THE MAIN PROCESS RESULT OF THE PROPOSED METHOD

Figure 5 shows a series of results generated in the process of extracting buildings, including sample regions of four categories, the initial building mask and the final building results. The images in the first column are the original VHR images. Those in the second column indicate the automatic results of shadow detection, vegetation detection, building sample regions and the manually obtained sample regions of bare land. The initial buildings can be extracted from the classification results generated by inputting samples of four classes into SVM classifier, as illustrated in the images of the third column. The final building extraction results can then be

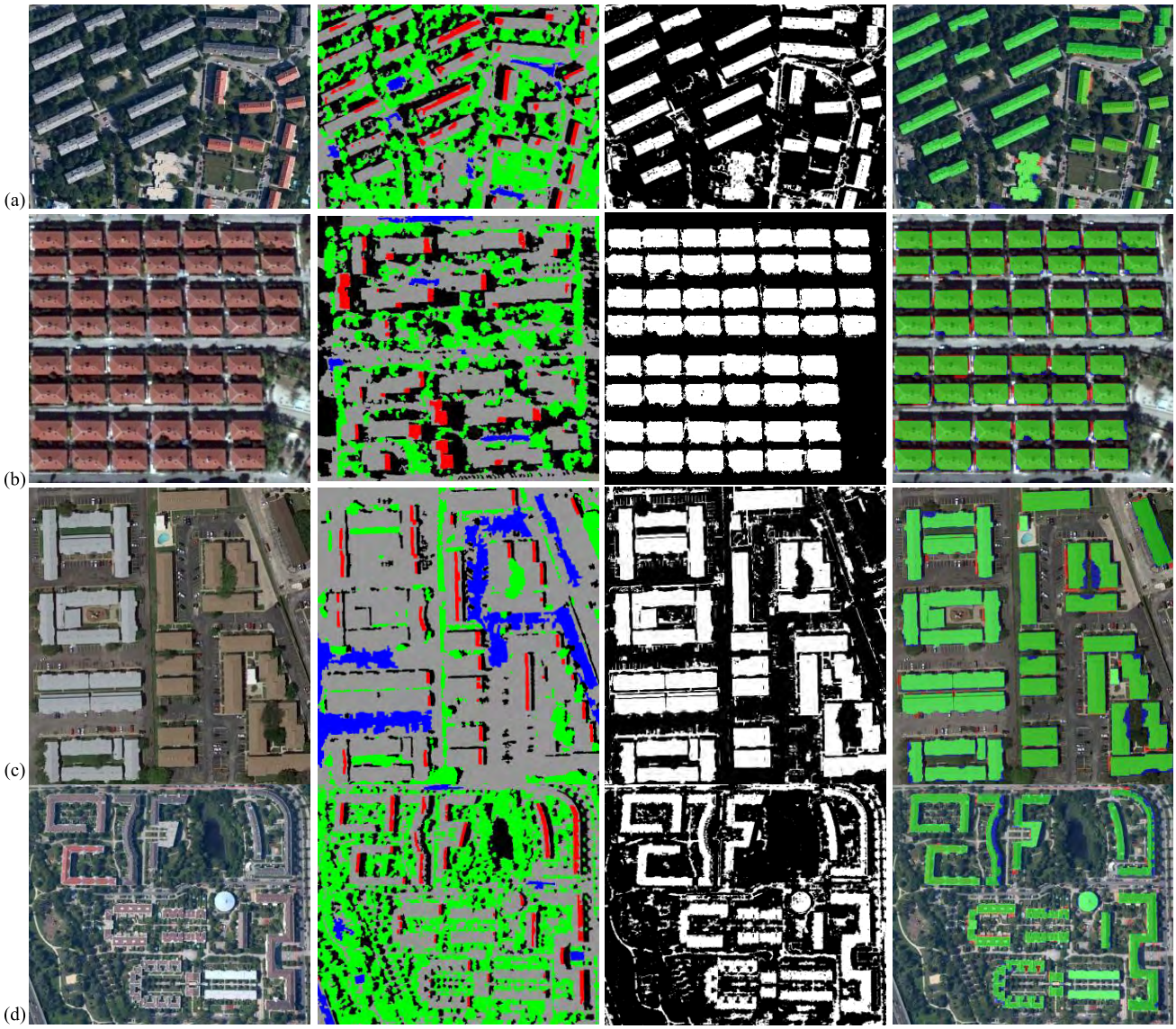

FIGURE 5. (First column) Test patches (a)–(d). (Second column) Images with four different classes as prepared for classification, where colours green, black, red, blue and gray indicate the regions of vegetation, shadows, building samples, bare land samples and unlabelled pixels, respectively. (Third column) Initial building masks extracted from the classification results. (Fourth column) Final building extraction results, where the colours green, red and blue represent TP, FP and FN pixels, respectively.

generated after post-processing, as presented in the images of the fourth column.

The samples automatically selected by the proposed method are very comprehensive. As shown in the second column in Figure 5, red regions are building samples extracted by the proposed method. The samples almost cover each building, which supplies sufficient building information for classification. Therefore, it is significant for improving the accuracy of the initial building result. In comparison, some initial building results generated by combining manually selected samples with SVM classifier are complemented. 13 images are tested by using different sample selecting measures. Figure 6 shows the comparison between the pixel-based performances of the initial results gained by these two methods, including correctness and completeness. The averages of correctness, completeness and F-measure in the automatic method are 74.7%, 81.3% and 76.8%, while those of the manual method are 64.1%, 82.1% and 71.2% respectively. In comparison, the correctness of the automatic method is a bit higher than the manual method and the completeness is similar. In fact, F-measure combining correctness with completeness can reflect the performance accurately. In sum, as illustrated in Figure 6(c), it is obvious that F-measure of automatic method is generally higher than the ones of compared method. Thus, our proposed method gained more accurate initial buildings. It is mainly due to building samples, other shadow, and vegetation samples extracted automatically, variously, and completely.

TABLE 1. Performance levels of building extraction results in Figure 5.

Test image	Number of reference		Pixel-based performance of the initial result (%)			Pixel-based performance of the final result (%)			Object-based performance of the final result (%)		
	Pixels	Objects	Completeness	Correctness	F-measure	Completeness	Correctness	F-measure	Completeness	Correctness	F-measure
(a)	140911	28	93.10	70.03	79.93	93.32	97.11	95.18	96.43	100.0	98.18
(b)	80763	52	94.17	92.63	93.39	93.49	95.53	94.50	100.0	100.0	100.0
(c)	128301	32	86.23	79.82	82.90	89.78	96.74	93.12	100.0	100.0	100.0
(d)	96946	17	91.08	56.85	70.00	89.80	94.90	92.28	100.0	100.0	100.0

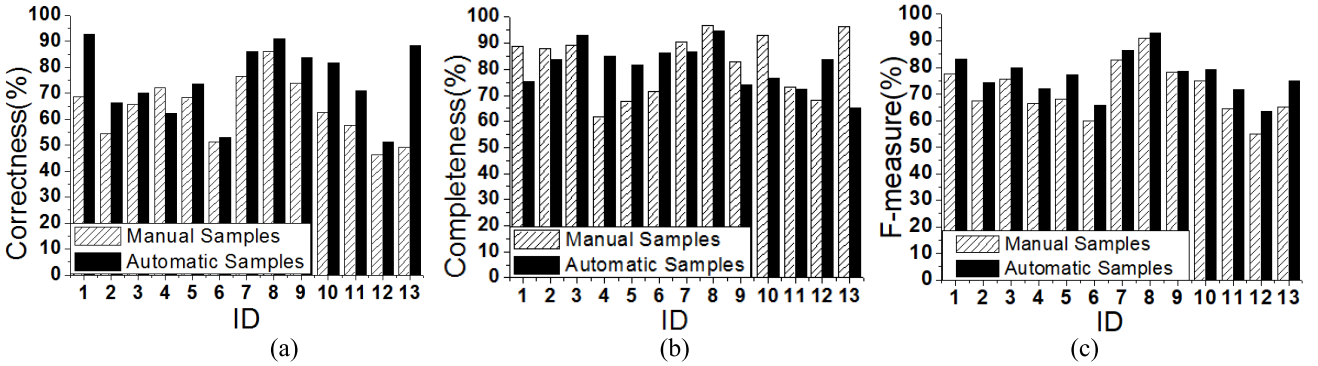

FIGURE 6. The initial results acquired by SVM classification with manual samples and automatic samples, shown in (a) pixel-based correctness, (b) pixel-based completeness, and (c) F measure performance.

Therefore, the proposed method assured the classification to provide an accurate initial building result.

Table 1 illustrated the performance of the initial and final building results of the proposed method from pixel-based and object-based perspectives respectively. Obviously, the performance of the final results is better than that of the initial results. The correctness and completeness are improved greatly. For example, as for 5(a), the initial results almost detect most of the buildings and some non-buildings such as road. Therefore, its completeness reaches 93% while its correctness is only 70%. After the promotion of the post-processing, each index especially correctness increases greatly which reaches 97%. Similarly, results of other images are optimized obviously by the post-processing. The reason is that bare lands which is similar with buildings in spectral features, are the main misrecognition. The proposed strategies uses shadow index to distinguish between bare land and building in the post-processing, so most non-buildings especially shadow are removed. Therefore, the correctness is improved significantly. In addition, adopting suitable mathematical morphology and region growth processing, the completeness of buildings can be improved partly. Seen from the object-based performance of final results, almost all buildings can be found correctly and with little wrong recognition. In sum, the post-processing method improves the correctness and completeness of the final results obviously, and the final performance is good in pixel-based and object-based evaluation.

If non-building objects adjacent to a building have similar spectral signature, sometimes they are prone to be recognized as part of building as complemented. That is why there are a few of erroneous building detection. Therefore, our post-processing strategy can also be used to solve the problems induced by classification. The overall performance of these test images indicates that our approach efficiently detects buildings. Therefore, the proposed method is efficient to detect buildings.

2) ASSESSMENT OF THE COMPARISON TO OTHER BUILDING EXTRACTION RESULTS

Another building extraction method given by Ghaffarian [34], using shadow to detect training areas for parallelepiped classification(i.e. SD-PAC) is introduced to compare with the proposed approach in this paper. IV depicts a set of clear visual comparison results. The images in the first column are the original VHR images. Those in the second column are the results of the proposed method in this paper. The results of SD-PAC are illustrated in the images of the third column. Their performance is listed in Table 2.

TABLE 2. Performance levels of building extraction results.

Test image	Number of reference		Building extracted methods	Pixel-based performance of the final result (%)			Object-based performance of the final result (%)		
	Pixels	Objects		Completeness	Correctness	F-measure	Completeness	Correctness	F-measure
(e)	104800	95	The proposed method	83.66	93.55	88.33	94.74	98.90	96.77
			SD-PAC method	83.91	80.68	82.27	94.73	86.53	90.45
(f)	182658	79	The proposed method	85.67	97.40	91.16	100.0	100.0	100.0
			SD-PAC method	94.63	83.39	88.66	100.0	79.00	88.26
(g)	114607	90	The proposed method	82.17	87.86	84.92	84.44	97.44	90.48
			SD-PAC method	84.55	78.80	81.57	88.88	82.47	85.56
(h)	122292	46	The proposed method	80.41	96.22	87.61	80.43	94.87	87.05
			SD-PAC method	75.02	83.79	79.16	73.91	51.51	60.71
(i)	475794	51	The proposed method	81.38	95.71	87.96	80.39	95.34	87.23
			SD-PAC method	70.51	80.43	75.14	70.58	73.46	72.00
(j)	96946	17	The proposed method	89.80	94.90	92.28	100.0	100.0	100.0
			SD-PAC method	81.73	84.05	82.88	94.11	45.71	61.53
(k)	207138	41	The proposed method	86.27	96.41	91.06	88.37	97.44	92.68
			SD-PAC method	71.97	76.07	73.49	74.41	50.79	60.37
Average	1304235	419	The proposed method	84.19	94.57	89.04	95.28	99.38	97.24
			SD-PAC method	80.33	80.89	80.45	85.23	67.07	74.13
Min	96946	17	The proposed method	80.06	87.86	84.92	80.39	94.87	87.05
			SD-PAC method	70.51	75.07	73.49	70.58	45.71	60.37
Max	475794	95	The proposed method	89.79	97.39	92.28	100.0	100.0	100.0
			SD-PAC method	94.63	84.05	88.66	100.0	86.53	90.45

Table 2 and Figure 7 illustrated the difference between the performance of our proposed approach and SD-PAC method. Obviously, experimental results on the test images show that the proposed approach attained better performance in both pixel-based and object-based accuracy. The main reason lies in the different strategies of classification and post-processing methods.

Firstly, in the process of classification, shadow is used to acquire training samples of building but different classifiers are adopted to obtain the initial result. By using shadow to acquire building samples, it is efficient to get an accurate initial building location. If the samples of some isolated buildings are not collected at first, the buildings probably cannot be detected. Then these missing buildings won't be counted in the object-based performance. Therefore, the object-based performance of completeness can indicate the accuracy of classification results from another perspective. As shown in Table 2, it is easy to find the differences in object-based completeness in both methods are small, less than 7%. In some images, they almost got the similar accuracy, such as 7(e), 7(f). This means shadows can provide sufficient samples to find possible buildings. Thus, the object-based completeness has not big difference. Moreover, because we adopt SVM classifier while SD-PAC uses parallelepiped classifier in supervise classification, the correctness of final results can reflect their ability in detecting accurate buildings. Generally speaking, SVM has a better ability in classification in comparison. Our SVM method totally is about higher than the SD-PAC method by 13.4%. Thus, with sufficient samples and appreciate classifier, the performance of the initial results through the proposed method is overall higher than SD-PAC.

Moreover, owing to adopting different post-processing strategies, the accuracy of the final results differs a lot, especially in the object-based correctness. Because SD-PAC just uses mathematics morphology to remove small regions and complete some undetected building holes located inside of detected buildings. Its completeness is improved. However, it lacks a further measure to verify whether the detected regions is building or not leading to a lower correctness. In contrast, in our approach, except morphology optimization, we additionally adopt shadow verification to remove those wrongly detected buildings and region growth to complement those regions adjacent to buildings. In addition, the most non-buildings are bare lands. The proposed index

FIGURE 7. (First column) Test patches (e)–(k). (Second column) The results of the proposed method. (Third column) The results of the SD-PAC (Ghaffarian [34]) where the colours green, red and blue represent TP, FP and FN pixels, respectively.

using shadow is very efficient to distinguish them. With its help, those detected initial non-buildings can be removed effectively. Then, the correctness of our approach can be improved greatly. As shown in Figure 7, red objects are non-buildings which cannot be filtered by post-processing, where such wrong recognition is serious in SD-PAC method.

TABLE 3. Elapsed time of each section in the proposed building detection approach.

Sections	Total time (s)	Average time (s)	Percentage
Shadow and vegetation detection	2.91	0.18	0.38%
Building shadow extraction	5.67	0.35	0.75%
Sample extraction	22.33	1.39	2.93%
SVM classification	693.53	43.35	92.05%
Post-processing	28.95	1.81	3.84%
Grand total	753.39	47.08	100%

In some complex scenes, such as 7(g) and 7(i) there are some FPs in both methods. These phenomenon always occur in some building where its adjacent bare land has similar spectral features with them. They are masked together in the initial results that post-processing cannot distinguish. However, even in this situation, the proposed methods have much less FPs than SD-PAC. Meanwhile, as shown in Table 2, the average of correctness object-based in the proposed method is about 95%, which is 30% up to SD-PAC. Therefore, the performance of F-measure in the proposed method is generally better than SD-PAC.

3) COMPUTATION TIME

The proposed methods were implemented in a VS2010 environment using a hybrid program code based on C++ and OpenCV. All of the experiments were conducted on a desktop computer with an Intel Core i5 CPU, 3.10 GHz and 4 GB RAM. The elapsed time for each section in our proposed methods to extract shadows, vegetation and buildings from 16 test images are indicated in Table 3. The average number of pixels is 512,360 (772 × 663). This indicates that the total and average elapsed time are approximately 753.39s (=12.55 min) and 47.08 s respectively. Moreover, the average duration of shadow and vegetation detection is only 0.18 s, which corresponds to the shortest time in all of the stages. This speed mainly benefits from the principle of spectral signatures with automatic thresholds. By contrast, the SVM classification of building extraction consumes the most time at 92% of the total time. However, the average time required to process an image is approximately 47 s, which is much less than the average processing time required by the method presented by Ok [4]. This rapid processing time is attributed to the random strategy applied to select some samples, which reduces the number of samples and does not influence the classification results.

IV. CONCLUSIONS

In this paper, we propose novel methods to detect buildings in RGB VHR images using shadow sufficiently in the process of collecting samples and building verification. Results are compared with other methods from two aspects. Experimental results show that this proposed sample extraction by shadow is effective and can improve the automation greatly. In addition, the final results are compared with SD-PAC. Both post-processing strategy differ a lot especially in building verification. Because the SD-PAC doesn't have any verification measures, leading to its performance which is much lower than the proposed method. In contrast, integrated with region growth and morphological operations as well as verification through shadow, the results are optimized effectively by our proposed method. Thus, our post-processing strategy can also be used to solve the problems induced by classification and then leads to the overall accuracy improved greatly.

The experimental results on over 16 test images derived from aerial and satellite images demonstrate that the proposed approach is successful in terms of shadow, vegetation and building detection. A significant characteristic of our approach is that it is no longer restricted with respect to image type because it requires only RGB band data, which are basic data in all kinds of VHR images. Therefore, the application range of our approach is broad. Nonetheless, our building extraction approach requires manual operation to obtain bare land samples. In addition, pixel-based classification has some drawbacks to obtain complete buildings. Hence, on one hand, future work should focus on automatically acquiring bare land samples which will promote the automation of our approach. On the other hand, object-oriented segmentation can be furtherly integrated with our proposed strategy using shadow to obtain better results. Last but not least, other pieces of building evidence should also be researched and combined with shadows for building verification.

REFERENCES

[1] X. Huang, W. Yuan, J. Li, and L. Zhang, "A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 654–668, Feb. 2017.

[2] T. J. Kim and J.-P. Müller, "Development of a graph-based approach for building detection," *Image Vis. Comput.*, vol. 17, no. 1, pp. 3–14, Jan. 1999.

[3] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.

[4] A. O. Ok, "Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, pp. 21–40, Dec. 2013.

[5] H. Ruther, H. M. Martine, and E. G. Mtalo, "Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 56, pp. 269–282, Jul. 2002.

[6] S. Ahmadi, M. J. V. Zoej, H. Ebadi, H. A. Moghaddam, and A. Mohammadzadeh, "Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, pp. 150–157, Jun. 2010.

[7] A. J. Fazan and A. P. Dal Poz, "Rectilinear building roof contour extraction based on snakes and dynamic programming," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 25, pp. 1–10, Dec. 2013.

[8] X. Jin and C. H. Davis, "Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information," *EURASIP J. Appl. Signal Process.*, vol. 2005, p. 745309, Dec. 2005.

[9] A. Katartzis and H. Sahli, "A stochastic framework for the identification of building rooftops using a single remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 259–271, Jan. 2008.
[10] M. Ghanea, P. Moallem, and M. Momeni, "Automatic building extraction in dense urban areas through GeoEye multispectral imagery," *Int. J. Remote Sens.*, vol. 35, no. 13, pp. 5094–5119, 2014.
[11] Z. Lari and H. Ebadi, "Automated building extraction from high-resolution satellite imagery using spectral and structural information based on artificial neural networks," in *Proc. ISPRS Hannover Workshop*, 2007, pp. 1–4.
[12] L. Abraham and M. Sasikumar, "Automatic building extraction from satellite images using artificial neural networks," *Procedia Eng.*, vol. 50, pp. 893–903, Jan. 2012.
[13] D. Koc-San and M. Turker, "A model-based approach for automatic building database updating from high-resolution space imagery," *Int. J. Remote Sens.*, vol. 33, no. 13, pp. 4193–4218, 2012.
[14] D. K. San and M. Turker, "Building extraction from high resolution satellite images using Hough transform," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 38, no. 8, pp. 1063–1068, 2010.
[15] R. Maurya, P. Gupta, A. S. Shukla, and M. K. Sharma, "Building extraction from very high resolution multispectral images using NDVI based segmentation and morphological operators," in *Proc. Int. Conf. Adv. Eng., Sci. Manage. (ICAESM)*, Mar. 2012, pp. 577–581.
[16] E. Li, S. Xu, W. Meng, and X. Zhang, "Building extraction from remotely sensed images by integrating saliency cue," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 906–919, Mar. 2017.
[17] X. Huang and L. Zhang, "Morphological building/shadow index for building extraction from high-resolution imagery over urban areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, no. 1, pp. 161–172, Feb. 2012.
[18] A. K. Shackelford, C. H. Davis, and X. Wang, "Automated 2-D building footprint extraction from high-resolution satellite multispectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2004, pp. 1996–1999.
[19] M. Turker and D. Koc-San, "Building extraction from high-resolution optical spaceborne images using the integration of support vector machine (SVM) classification, Hough transformation and perceptual grouping," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 58–69, Feb. 2015.
[20] M. Uzar, "Automatic building extraction with multi-sensor data using rule-based classification," *Eur. J. Remote Sens.*, vol. 47, pp. 1–18, Jan. 2014.
[21] M. A. Niveetha and R. Vidhya, "Automatic building extraction using advanced morphological operations and texture enhancing," *Procedia Eng.*, vol. 38, pp. 3573–3578, Jan. 2012.
[22] S. Lhomme, D.-C. He, C. Weber, and D. Morin, "A new approach to building identification from very-high-spatial-resolution images," *Int. J. Remote Sens.*, vol. 30, no. 5, pp. 1341–1354, 2009.
[23] K. Karantzalos and N. Paragios, "Recognition-driven two-dimensional competing priors toward automatic and accurate building detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 133–144, Jan. 2009.
[24] J. Wang, X. Yang, X. Qin, X. Ye, and Q. Qin, "An efficient approach for automatic rectangular building extraction from very high resolution optical satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 487–491, Mar. 2015.
[25] E. Sumer and M. Turker, "An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images," *Comput. Environ. Urban Syst.*, vol. 39, pp. 48–62, May 2013.
[26] J. Yang and Y.-H. Wang, "Towards automatic building extraction: Variational level set model using prior shape knowledge," in *Proc. Int. Conf. Image Anal. Signal Process.*, Nov. 2012, pp. 1–6.
[27] T. Lee and T. Kim, "Automatic building height extraction by volumetric shadow analysis of monoscopic imagery," *Int. J. Remote Sens.*, vol. 34, no. 16, pp. 5834–5850, 2013.
[28] G. Singh, M. Jouppi, Z. Zhang, and A. Zakhor, "Shadow based building extraction from single satellite image," *Proc. SPIE*, vol. 9401, pp. 94010F-1–94010F-15, Mar. 2015.
[29] Y. Iwasaki and F. Yamazaki, "Detection of building collapse from the shadow lengths in optical satellite images," in *Proc. 32nd Asian Conf. Remote Sens.*, 2011, pp. 550–555.
[30] B. Sirmacek and C. Unsalan, "Building detection from aerial images using invariant color features and shadow information," in *Proc. 23rd Int. Symp. Comput. Inf. Sci. (ISCIS)*, Oct. 2008, pp. 1–5.
[31] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[32] T. Joachims, "Making large-scale SVM learning practical," Komplexitätsreduktion multivariaten Datenstrukturen, TU Dortmund, Dortmund, Germany, Tech. Rep. SFB 475, 1998, pp. 499–526.
[33] J. Xiao, M. Gerke, and G. Vosselman, "Building extraction from oblique airborne imagery based on robust Façade detection," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 56–68, Mar. 2012.
[34] S. Ghaffarian, "Automatic building detection based on supervised classification using high resolution Google Earth images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 3, p. 101, 2014.

XIANJUN GAO received the B.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2010 and 2015, respectively. Since 2015, she has been a Lecturer with the School of Geoscience, Yangtze University, Wuhan, China. Her major research interests include aerial images real-time processing and automatic object interpretation from high resolution images.

MINGWEI WANG received the B.S. degree from Hubei Normal University, Huangshi, China, in 2011, and the M.S. degree from the Hubei University of Technology, Wuhan, China, in 2015. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China. His major research interests include hyperspectral image processing, swarm intelligence, and machine learning.

YUANWEI YANG received the B.S. degree in computer science and technology from Hubei Engineering University, Xiaogan, China, in 2007, and the M.S. and Ph.D. degrees in geographic information system from Wuhan University, Wuhan, China, in 2009 and 2016, respectively. He is currently a Lecturer with the School of Geoscience, Yangtze University, Wuhan, China. His research interests include spatial data management and update, remote sensing image processing and analysis, and vector data map matching.

GONGQUAN LI received the Ph.D. degree in geophysical prospecting and information technology from the China University of Mining Industry in 2006. He is currently a Vice-Professor and a Master Advisor with the School of Geosciences, Yangtze University, Wuhan, China, and is also the Vice-Director of the Geographic Information System Department.

His research interests include 3-D geological modeling, spatial analysis, big data, digital image processing and analysis, and geographic information engineering technology.