A Morphological Building Detection Framework for High-Resolution Optical Imagery Over Urban Areas

Article *in* IEEE Geoscience and Remote Sensing Letters · July 2016

DOI: 10.1109/LGRS.2016.2590481

3 authors, including:

Qian Zhang
East China Normal University
12 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Xin Huang
Wuhan University
163 PUBLICATIONS **4,500** CITATIONS
SEE PROFILE

A Morphological Building Detection Framework for High-Resolution Optical Imagery Over Urban Areas

Qian Zhang, Xin Huang, *Senior Member, IEEE*, and Guixu Zhang

***Abstract*—This letter proposes an efficient framework for building detection from coarse to fine using morphological technique for high-resolution optical satellite imagery over urban areas. First, the preliminary result of building regions is obtained by the recently developed morphological building index (MBI) method, which is able to detect potential building structures. However, the raw results derived from the MBI can be subject to a number of false alarms, which are caused by bright soil, roads, and open areas. In this letter, we propose to use morphological spatial pattern analysis as a postprocessing to further optimize the MBI result and remove the commission errors. The original MBI result is then separated into seven mutually exclusive categories—core, islet, loop, bridge, perforation, edge, and branch—by applying a series of morphological transformations such as erosions, geodesic dilation, reconstruction by dilation, anchored skeletonization, etc. The objects corresponding to the generic categories are then analyzed, and the categories corresponding to building parts are maintained, while the others are abandoned. After this postprocessing, the small noisy patches and narrow roads, which were wrongly extracted by the MBI, can be removed. In addition, the shape of the buildings can also be regularized by removing the branches, and the holes contained in the building objects can be identified and filled. Extensive experiments performed on GeoEye-1 and WorldView-2 images confirm the effectiveness and robustness of the proposed morphological building detection framework.**

***Index Terms*—Building detection, high resolution, morphological, postprocessing.**

I. Introduction

BUILDINGS are a key factor in region management and planning, change monitoring, etc., as buildings can reflect the status and trends of a region's development. Therefore, having reliable and accurate information about building objects is critical, particularly in rapidly developing countries such as China. Extensive research has been devoted to building extraction in recent decades. The emergence of available high spatial resolution images from satellite sensors has provided more spatial information and opportunities for building extraction but has also resulted in more challenges. This has prompted scholars to break through the traditional thinking and look for new solutions to exploit the spatial details and extract accurate building information [1]. High spatial resolution images have now become one of the most important input sources for rapidly obtaining building information.

A number of studies have used additional information as a supplement for building verification or distinction. Sources of additional information have included light detection and ranging (LiDAR) images, digital elevation models, digital surface models (DSMs) [2], [3], stereo imagery [4], and so on. Qin and Fang [2] used both a DSM and aerial imagery to hierarchically extract buildings. Zhang *et al*. [3] used ultrahigh-resolution orthophotographs and a DSM to obtain a classification map based on a dual morphological top-hat profile. Shaker *et al*. [4] used IKONOS stereo imagery and digital surface and terrain models to extract building planimetric positions and heights, respectively. Due to the great contribution of the additional information, the aforementioned methods generally perform well. However, a major problem is that auxiliary data sources (e.g., the LiDAR point cloud) are not available in most situations.

On the other hand, image primitives such as shadows [5]–[7], corners, edges, or line segments have also been considered for building extraction. Ok *et al*. [5] proposed a new fuzzy landscape generation approach (denoted as Shadow-based) to model the directional spatial relationship between buildings and their shadows, and the final arbitrarily shaped building regions are automatically detected by a GrabCut partitioning approach from monocular QuickBird and GeoEye-1 imagery. Ok [6] then extended the previous approach [5] and proposed an effective shadow postprocessing step (denoted as Shadow-GC, GC = graph cuts) to refine the building shadow and a two-level graph theory approach to detect building regions. Manno-Kovacs and Ok [7] also proposed an approach for building detection from monocular very high resolution optical images by integrated urban area knowledge.

More recently, Huang and Zhang [8] proposed the morphological building index (MBI), which can be quickly implemented, does not need training samples, and has proved to be effective for automatic building extraction from high-resolution imagery. Buildings are extracted by performing a threshold (T) operation on the MBI feature image, which is extracted by a set of morphological operators (e.g., top-hat by reconstruction, granulometry, and directionality) according to the intrinsic properties of the buildings (e.g., brightness, size, contrast, directionality, and shape). However, there are drawbacks to the MBI, as analyzed in [8]. Due to its brightness hypothesis, i.e., building regions are brighter than their surroundings, bright objects such as bright soil, open areas, and roads can cause a number of false alarms since they are brighter than their neighborhoods and show similar spectral properties to buildings.

Manuscript received March 14, 2016; revised June 7, 2016 and June 26, 2016; accepted July 9, 2016. Date of publication July 27, 2016; date of current version August 5, 2016. This work was supported in part by the National Natural Science Foundation of China under Grants 41301472 and 91338111, by the China National Science Fund for Excellent Young Scholars under Grant 41522110, and by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201348.

Q. Zhang and G. Zhang are with the Department of Computer Science and Technology, East China Normal University, Shanghai 200241, China.

X. Huang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2016.2590481

In this letter, an effective morphological building extraction framework is proposed for high-resolution optical imagery. According to the morphological characteristics of buildings, the MBI is constructed to extract coarse building structures. A morphological component classification approach is then used to remove the MBI commission errors (CEs). The main contribution of this letter is the proposal of an automatic morphological framework for building detection in high-resolution optical imagery from coarse to fine.

The rest of this letter is organized as follows. Section II describes the MBI features, morphological spatial pattern analysis (MSPA), and the postprocessing. Section III analyzes and compares the experimental results, which is followed by the conclusion in Section IV.

II. Methodology

A. MBI

The MBI [8] can be viewed as a high-level feature for building extraction, which possesses multiscale and multidirectional characteristics, free parameters, and low computational complexity. The basic principle of the MBI is that building regions are brighter than their surroundings (building shadows). Building objects are extracted by the MBI according to the spectral–spatial properties of the buildings (e.g., brightness, size, contrast, directionality, and shape) by a set of morphological operators. The calculation of the MBI features is briefly introduced as follows.

1) Brightness image. Each pixel's maximum value in the visible bands is kept as the brightness. This is because the visible bands make the most significant contribution to the spectral property of buildings.
2) Top-hat morphological profiles. The differential morphological profiles of the top-hat transformation ($\mathrm{DMP_{TH}}$) represent the spectral–structural property of the buildings

$$\mathrm{DMP_{TH}} = [\Delta \mathrm{TH}(d,s), d \in D, s \in S],$$
$$\text{with, } \Delta \mathrm{TH}(d,s) = |\mathrm{TH}(d,s) - \mathrm{TH}(d, s - \Delta s)|$$

where TH represents the top-hat by reconstruction of the brightness image, s and d indicate the scale and direction of a linear structural element (SE), respectively, S and D represent the sets of scales and directions, respectively, and Δs is the interval of the profiles. The top-hat transformation can highlight the locally bright structures with a size up to a predefined value and is used to measure the contrast.
3) Calculation of the MBI. The MBI is defined by

$$\mathrm{MBI} = \frac{\sum_d \sum_s \mathrm{DMP_{TH}}(d,s)}{N_d \times N_s} \quad (1)$$

where N_d and N_s are the numbers of directions and scales, respectively. The building index is defined as the average of the multiscale and multidirectional $\mathrm{DMP_{TH}}$ since building structures have larger feature values in most of the scales and directions in the morphological profiles due to their local contrast and isotropy.

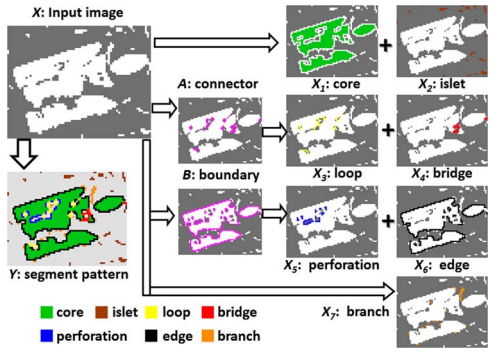

Fig. 1. MSPA segmentation of binary patterns. X: Input binary pattern (white: foreground; grey: background). X_i: Intermediate steps. Y: Resulting segmentation.

B. MSPA

In a binary map (foreground and background), MSPA is used to analyze the geometry and connectivity of the image components. MSPA consists of a series of sequential mathematical morphological operators, such as erosion, geodesic dilation, reconstruction by dilation, and anchored skeletonization. Each pixel of the foreground is segmented as one of the mutually exclusive sets of structural classes: core, islet, edge, loop, perforation, bridge, and branch. In different applications, the corresponding objects in the generic MSPA categories are different [9]–[11]. MSPA borrows the notion of path connectivity to determine the 8-connected or 4-connected regions of the foreground (FGconn). There are two key parameters in MSPA, i.e., the foreground connectivity (FGconn) and the size parameter (Ew). Parameter Ew defines the width or thickness of the noncore classes in the pixels.

Fig. 1 shows the calculation process of MSPA. Core pixels (X_1) are groups of foreground pixels whose distance to the background is greater than the given size parameter Ew. The core pixels are detected by the erosion of input image (X) with a Euclidean disk of radius equal to Ew. In practice, the core pixels X_1 are obtained by thresholding the Euclidean distance transform (EDT) of the foreground by a threshold value Ew

$$X_1 = T_{d \geq Ew}\left[\mathrm{EDT}(X)\right].$$

Islet pixels (X_2) are connected components of the foreground that do not include any core pixels. The islet pixels are the difference between the input image and the reconstruction by the dilation of the core pixels

$$X_2 = X - R_X^{\delta}(X_1)$$

where $R_X^{\delta}(X_1)$ represents the reconstruction by the dilation of X using X_1 as the seed.

Both bridge and loop are connectors (A) which are foreground pixels linking the core connected components. The anchored skeleton of core pixels is first extracted. The connector pixels are then the noncore pixels whose geodesic distance from this skeleton is less than s. Bridge pixels (X_3) are those pixels emanating from more than one core connected component, while loop pixels (X_4) are from the same single connected component.

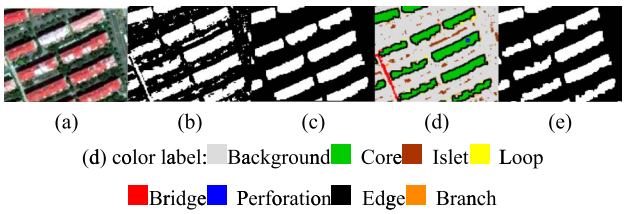

Fig. 2. Example of the proposed approach. (a) Original image. (b) Original MBI features. (c) Ground reference. (d) MSPA result, $Ew = 2$. (e) Final building result extracted from (d).

Boundaries (B) consist of edge and perforation pixels and are the unclassified foreground pixels whose distance to the core pixels is less than or equal to Ew

$$B = T_{0<d\leq EW}\left[\text{EDT}\left(X_1^C\right)\right] \cap [X - (X_1 \cup X_2 \cup X_3 \cup X_4)]$$

where the boundary pixels (B) are subdivided into outer and inner boundaries, i.e., edge (X_5) and perforation (X_6), respectively.

Branch pixels are those pixels that do not belong to any of the previously defined classes, and they emanate from boundaries or connecters.

MSPA is size dependent due to it being based on a size parameter corresponding to a Euclidean distance threshold value. Thus, the size, area, extent, and existence of each category are affected by the value of the size parameter Ew [9].

C. Proposed Scheme

The process of the proposed morphological building detection framework consists of the following three steps.

1) The original MBI features are obtained by (1) from the original image.
2) Morphological classification using MSPA. The original MBI feature is the input data for MSPA, and all the pixels whose values are larger than 0 are considered to be foreground; all the other pixels are considered to be background. All the foreground pixels are then segmented and allocated by MSPA as one of the seven structural categories.
3) Postprocessing. This step aims to extract the building regions from the MSPA classification result. According to the MSPA structural categories, some of the classes are kept as building regions, and the others are abandoned.

In order to illustrate the process of the proposed framework in detail, an example is shown in Fig. 2, which is cut from a large-area image with a 1-m resolution. Many nonbuilding objects with high brightness, such as cars, bare land, open spaces, roads, etc., are denoted as buildings in the original MBI [see Fig. 2(b)]. Additional errors are associated with the structures connected with the buildings.

It can be seen from Fig. 2(a) and (d) that the core pixels are the interior of the building areas, excluding the foreground perimeter. The islet pixels are the disjoint building objects, which mainly consist of cars and the small areas of roads between buildings. The loop pixels are connected at more than one end to the same building core area. The bridge pixels represent the connections between buildings and roads. The perforation pixels are the internal building object perimeters.

TABLE I
NUMBERS OF GROUND TRUTH SAMPLES FOR THE FIVE VALIDATION AREAS

Data	Satellite	No. of samples		Landscape
		Building	Background	
R1	GeoEye-1	17603	19497	urban
R2	GeoEye-1	15,476	15,786	urban
R3	GeoEye-1	13,125	14,690	urban
R4	GeoEye-1	14,855	14,672	urban
R5	Worldview-2	15,368	15,215	campus

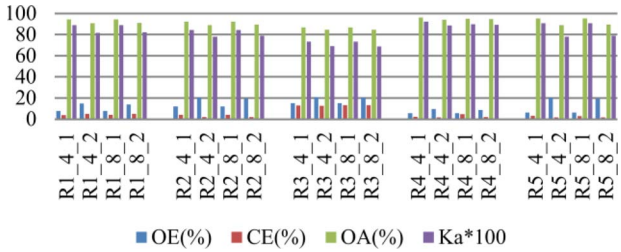

Fig. 3. Parameter sensitivity analysis for the proposed method. "R1_4_1" denotes the result of the proposed method with test image R1, with MSPA parameters $Ew = 4$ and FGconn $= 1$.

The edge pixels are the external building object perimeters. The branch pixels are connected at one end to the edge, perforation, bridge, or loop pixels.

Thus, according to the definition of the MSPA categories and for the purpose of building object extraction, the categories of islet, bridge, and branch are removed, and the core, loop, edge, and perforation classes are kept as the correct building structures, as shown in Fig. 2(e). In addition, the areas of the core smaller than a threshold (t) are removed in order to remove the small nonbuilding objects. A threshold length–width ratio ($t2$) is also used to remove the long and narrow objects (i.e., roads).

III. EXPERIMENTS

A. Image Data Set

The proposed building detection framework was assessed over five test sites selected from three different images acquired by the GeoEye-1 and WorldView-2 satellites, as shown in Fig. 4 (row 1). All five images were of 645 × 564 pixels, with a 2-m resolution. Among the test images selected, four of them (R1–R4) were urban regions, with dense buildings, selected from a single GeoEye-1 image of Wuhan, central China. One test image (R5), with sparse and large-area buildings, was selected from a WorldView-2 image of East China Normal University in Shanghai, China. Table I lists the numbers of ground-truth samples, where "background" indicates the nonbuilding areas. The ground truth was manually annotated.

B. Accuracy Assessment and Parameter Analysis

The assessment of the building extraction performance was undertaken with four statistical measures [12]. The omission error (OE) and CE indicate samples that are wrongly identified

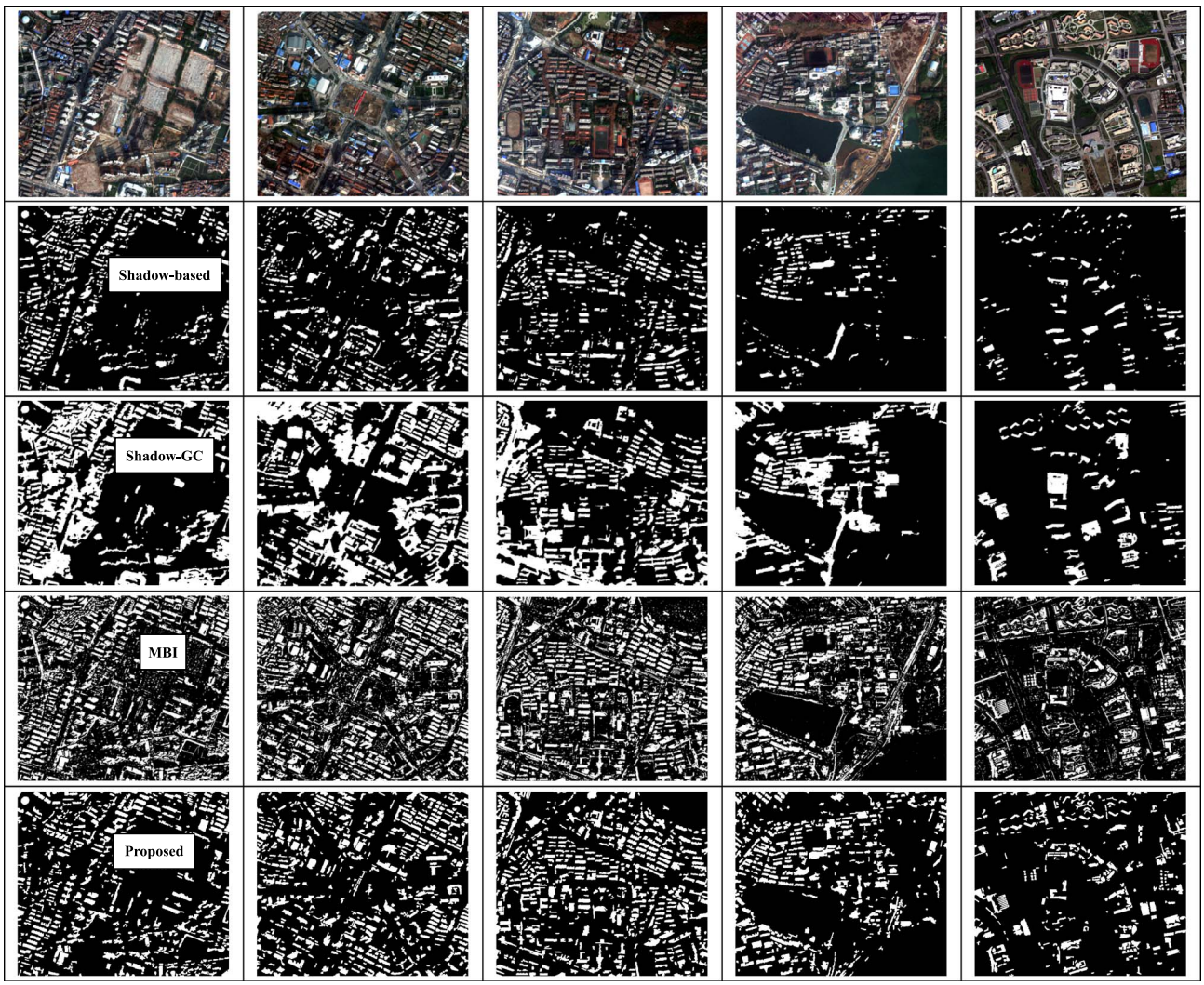

Fig. 4. Results of building detection. Row 1: Columns 1–5 represent the original images of R1–R5, respectively. Rows 2–5 represent the building maps extracted by the Shadow-based, the Shadow-GC, the MBI, and the proposed method, respectively.

as background and samples that are wrongly classified as buildings, respectively. The overall accuracy (OA) and Kappa coefficient (Ka) for the building–background classification were calculated based on the confusion matrix.

The MBI refers to two parameters: directionality (d) and the size (s) of the SE. Eight directions (d) were considered, ranging from 22.5 to 180 with an interval of 22.5. The linear SE size (s) was varied from 2 to 22 with a step size of 5, which was set according to the spatial resolution of the image and the size of the buildings in the study area. Further details of MBI parameter selection can be found in [8].

Fig. 3 shows the sensitivity of the MSPA parameters, namely, FGconn and Ew, to the accuracy of the R1–R5 test images. Clearly, the results are not sensitive to the FGconn parameter. When Ew is kept the same, the results with different FGconn values show only very slight differences, which can be ignored. Based on our extensive experiments, objects can be more easily separated with less computation time when FGconn uses 4-connectivity rather than 8-connectivity. Therefore, FGconn was fixed as 4 in this letter. Parameter Ew has a greater impact on the results than FGconn, as shown in Fig. 3. According to the definition of Ew, the image resolution, and the average size of the buildings, the optimal values of Ew for images with different spatial resolutions will be different. For all the test images, the OA and Ka of $Ew = 1$ are higher than those of $Ew = 2$, and the OE is lower in Fig. 3. Thus, the optimal value of Ew is 1 for the five test images.

The postprocessing involves two parameters: the threshold of the smallest area of objects (t) and the length–width ratio $(t2)$ of narrow objects, which were set to 30 and 9.6, respectively. These parameters were kept constant in all the test data sets.

C. Results and Discussion

The proposed morphological building detection framework was compared with the original MBI features and the Shadow-based and Shadow-GC methods proposed by Ok *et al.* [5] and Ok [6]. Fig. 4 shows the extracted building maps, and the accuracies are given in Table II.

It can be seen that many buildings are not detected by the Shadow-based (see row 2 of Fig. 4), and the average OE of the Shadow-based is the highest, as listed in Table II. The aforementioned problem is significantly reduced in the results of the shadow-GC, but there is slight overcorrecting. Many

TABLE II
ACCURACY ASSESSMENT OF THE BUILDING EXTRACTION FOR THE SHADOW-BASED, THE SHADOW-GC, THE MBI, AND THE PROPOSED METHOD (THE UNITS OF OE, CE, AND OA ARE IN PERCENT)

Site	Accuracy evaluation															
	Shadow-based				Shadow-GC				MBI				Proposed			
	OE	CE	OA	Ka	OE	CE	OA	Ka	OE	CE	OA	Ka	OE	CE	OA	Ka
R1	37.4	2.4	81.5	0.623	8.6	15.7	87.8	0.756	6.1	19.9	86	0.722	**7.9**	**3.8**	**94.5**	**0.89**
R2	39.1	2.8	79.8	0.594	**5.5**	**5.2**	**94.7**	**0.894**	9.7	12.5	88.8	0.776	12	4.1	92.2	0.843
R3	**21.8**	**0.5**	**89.6**	**0.788**	10.8	19.1	85	0.701	11.7	22.8	82.2	0.645	15	13.1	86.8	0.735
R4	58.7	13.7	67.2	0.346	31.8	14.5	78.2	0.564	5.2	18.8	86.4	0.727	**5.6**	**2.2**	**96.1**	**0.922**
R5	56.5	0.2	71.6	0.433	28.5	0.2	85.6	0.712	5.8	16.9	87.5	0.749	**6.4**	**3.1**	**95.3**	**0.906**
Ave	42.7	3.9	77.9	0.557	17.0	10.9	86.3	0.725	7.7	18.2	86.2	0.724	**9.4**	**5.3**	**93.0**	**0.859**

roads are detected as buildings by the Shadow-GC (see row 3 of Fig. 4). Thus, the CE of the Shadow-GC shows a significant increase when compared with that of the Shadow-based. The Shadow-based and Shadow-GC obtain the best results in R3 and R2, respectively.

There are a number of nonbuilding objects that are detected as buildings by the MBI in all the test data sets (see row 4 of Fig. 4). The observations are consistent with the CE results listed in Table II. The bare land, roads, and vegetation (grass or farmland) classes are detected as small patches of buildings by the MBI, which has a significant negative effect on the accuracy in all seven test images. However, most of these categories are classified as islet (small patches) or bridge (narrow roads) by MSPA and are therefore discarded. A few wide roads are removed by the length–width ratio, i.e., larger than the threshold (t2). Overall, the CE of the proposed approach is reduced by 12.9% on average compared with the CE of the MBI. The proposed approach is also effective in the rural area where buildings are not dominant, as shown in R5.

The OA of the proposed method shows an average increase of 15.1%, 6.7%, and 7.4%, respectively, compared with the Shadow-based, the Shadow-GC, and the MBI. Furthermore, the Kappa coefficient shows an average increase of 0.303, 0.135, and 0.152, respectively. The proposed approach is also very fast. The average computation times of the Shadow-based and Shadow-GC are about 300 and 480 s, respectively, for a computer with a 2.40-GHz Intel i7 processor and 16-GB RAM. The average computation times of the MBI and the proposed method are 3.3 and 7.7 s, respectively, for a computer with a 2.00-GHz Intel i7 processor and 8-GB RAM. Thus, despite the complexity of the test images, the results are encouraging and prove the proposed method's viability.

IV. CONCLUSION

The main contribution of this letter is that morphological component classification is used with the raw results of the MBI. According to the morphological classification results, the intrinsic CEs of the MBI are removed. The extensive experiments undertaken in this letter indicate that the proposed morphological building detection framework has significant advantages and outperforms the three state-of-the-art methods chosen for comparison. The proposed approach also has a very high computational efficiency. The traditional building detection algorithms are mainly designed for urban areas with dense buildings, and they are not designed for use in rural areas. However, the outstanding performance of the proposed framework is more apparent in rural areas with sparse buildings, compared with urban areas with dense buildings.

ACKNOWLEDGMENT

The authors would like to thank Prof. A. O. Ok for providing the results of his approach for comparison and the anonymous reviewers for the constructive comments.

REFERENCES

[1] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[2] R. Qin and W. Fang, "A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization," *Photogramm. Eng. Remote Sens.*, vol. 80, no. 9, pp. 873–883, 2014.

[3] Q. Zhang, R. Qin, X. Huang, Y. Fang, and L. Liu, "Classification of ultra-high resolution orthophotos combined with DSM using a dual morphological top hat profile," *Remote Sens.*, vol. 7, no. 12, pp. 16 422–16 440, 2015.

[4] I. F. Shaker, A. Abd-Elrahman, A. K. Abdel-Gawad, and M. A. Sherief, "Building extraction from high resolution space images in high density residential areas in the Great Cairo region," *Remote Sens.*, vol. 3, no. 4, pp. 781–791, 2011.

[5] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, Mar. 2013.

[6] 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, 2013.

[7] A. Manno-Kovacs and A. O. Ok, "Building detection from monocular VHR images by integrated urban area knowledge," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2140–2144, Oct. 2015.

[8] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 7, pp. 721–732, 2011.

[9] P. Soille and P. Vogt, "Morphological segmentation of binary patterns," *Pattern Recognit. Lett.*, vol. 30, no. 4, pp. 456–459, Mar. 2009.

[10] P. Vogt, J. R. Ferrari, T. R. Lookingbill, R. H. Gardner, K. H. Riitters, and K. Ostapowicz, "Mapping functional connectivity," *Ecological Indicators*, vol. 9, no. 1, pp. 64–71, 2009.

[11] M. Suarez-Rubio, T. R. Lookingbill, and A. J. Elmore, "Exurban development derived from Landsat from 1986 to 2009 surrounding the District of Columbia, USA," *Remote Sens. Environ.*, vol. 124, pp. 360–370, 2012.

[12] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL, USA: CRC Press, 2008.