

Morphological Building/Shadow Index for Building Extraction From High-Resolution Imagery Over Urban Areas

Xin Huang and Liangpei Zhang, *Senior Member, IEEE*

Abstract—Morphological building index (MBI) is a recently developed approach for automatic indication of buildings in high-resolution imagery. However, MBI is subject to commission errors due to the similar characteristics between buildings, bare soil and roads. Furthermore, omission errors occur in dark and heterogeneous roofs. In this study, a systematic framework for building extraction from high-resolution imagery is proposed, aiming to alleviate both commission and omission errors for the original MBI algorithm. The improvements include three aspects: 1) a morphological shadow index (MSI) is proposed to detect shadows that are used as a spatial constraint of buildings; 2) a dual-threshold filtering is proposed to integrate the information of MBI and MSI; 3) the proposed framework is implemented in an object-based environment, where a geometrical index and a vegetation index are then used to remove noise from narrow roads and bright vegetation. The proposed framework was validated on an Ikonos image of Washington DC Mall with 1-m resolution and an 8-channel WorldView-2 image of Hangzhou, east of China, with 2-m resolution. By comparison with the ground truth references, it was shown that our method achieved over 90% overall accuracy for discrimination between buildings and backgrounds for both datasets. In the comparative study, it was revealed that the proposed method improved the original MBI significantly. Furthermore, the proposed method was more accurate than the support vector machine interpretation with the differential morphological profiles (DMP) and multiscale urban complexity index (MUCI).

Index Terms—Building detection, high resolution, mathematical morphology, shadow detection, WorldView-2.

I. INTRODUCTION

THE location and identification of buildings is one of the most important tasks for territorial planning in the developing countries such as China, since the socio-economic and environmental issues resulting from the surprisingly high housing-price, have received increasing attention in recent years. The availability of high-resolution Earth Observation data makes it possible to precisely locate and identify buildings. However, the current data bases of urban buildings with high spatial resolution

are scarce. Therefore, in this study, the building extraction approach is investigated, in order to indicate the presence of buildings and update the data bases of urban buildings from high-resolution imagery.

Although the high-resolution remotely sensed imagery provides a new data source for building identification, it poses challenges to the traditional information extraction techniques. The traditional spectral-based approaches are inadequate for discrimination between spectrally similar classes in high-resolution imagery, such as buildings, roads, parking lots, and open areas, since they are made from similar materials. It is generally agreed that spatial information (e.g., texture, structure, and context) is able to complement the spectral feature space and discriminate the spectrally similar targets. Some commonly used spatial features include the pixel shape index (PSI) [1], morphological profiles [2]–[4], wavelet texture [5], and the gray-level co-occurrence matrix (GLCM) [6]. However, many of the aforementioned algorithms refer to the supervised approaches, where a large number of training samples are needed to precisely model the feature distribution of information classes. The supervised methods are subject to collection of training samples and process of machine learning. To this aim, some researchers proposed automated approaches for building detection from high-resolution data, such as the procedure for the calculation of a texture-derived built-up presence index (PanTex) [7]. The basic assumption of the PanTex is based on the fact that buildings cast a shadow that is producing high local contrast. Hence, the PanTex is built on the Contrast measure of GLCM and the directional components of the textural signal are taken into account, which produces a rotation-invariant built-up index. In [8], two improved options of the PanTex were presented. In the first option, the vegetation component was subtracted from the built-up areas presence index in order to remove the effects of bright vegetation; in the second option, a morphological filtering was used to highlight the morphological characteristics of the built-up structures before calculation of PanTex. The PanTex can be automatically calculated without sample collection and machine learning; therefore, it is very suitable for data base construction and update of urban buildings. PanTex has been proved to be effective for indication of built-up presence in high-resolution remotely sensed images [7], [8] and SAR images [9]. Nevertheless, the optimal spatial resolution of the PanTex is 5 meters because 5-m spatial resolution is considered sufficient for the discrimination of built-up areas using the textural approach.

Jin and Davis [16] presented an automated building extraction strategy that simultaneously exploited structural, contex-

Manuscript received May 16, 2011; revised July 26, 2011; accepted August 26, 2011. Date of publication October 20, 2011; date of current version February 29, 2012. This work was supported by the National Natural Science Foundation of China under Grant 41101336 and Grant 40930532, the Fundamental Research Funds for the Central Universities (3101016), and the LIESMARS Special Research Funding.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (corresponding author, e-mail: huang_wuhu@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/JSTARS.2011.2168195

tual, and spectral information. First, image structural information is extracted from the morphological profiles with a series of opening and closing operations. Second, shadows that are also extracted from the morphological profiles are used to provide reliable contextual information for buildings. Next, spectral information is used to detect bright buildings. The final result was obtained by integrating the results of the three different information sources.

Recently, Huang and Zhang [10] proposed a morphological building index, namely MBI, which has proved to be effective for automatic building extraction from high-resolution imagery with 2-m spatial resolution. The basic idea of MBI is to build a relationship between the implicit characteristics of buildings (e.g., brightness, size, and contrast) and the morphological operators (e.g., top-hat by reconstruction, granulometry, and directionality). Buildings are extracted by performing a threshold on the MBI feature image. As analyzed in [10], the original MBI algorithm is subject to commission and omission errors. The commission errors are related to bright soil, open areas and roads since they are also brighter than their neighborhoods and show similar spectral characteristics as buildings. On the other hand, most of omission errors refer to heterogeneous and dark roofs. The objective of this study is to propose a self-contained system for automatic building extraction from high-resolution imagery. The proposed framework is expected to alleviate the aforementioned commission and omission errors of the original MBI algorithm. To this aim, the following strategies are employed:

- A morphological shadow index (MSI) is proposed for automatic shadow extraction from an image. Shadows are spatially adjacent or close to buildings. As a consequence, the distance to shadows is used as a spatial constraint of buildings, based on which some commission errors such as open areas and bright soil can be removed.
- Most of heterogeneous and dark roofs correspond to low MBI values that are always viewed as noise and hence filtered out in binarization of the MBI image. As a response to this problem, in this study, the MBI image is separated into low-MBI and high-MBI regions. A strong constraint, i.e., a small threshold on the distance between buildings and shadow, is applied to the low-MBI regions, while a weak condition, corresponding to a relatively large threshold on the distance between buildings and shadow, is imposed on the high-MBI regions. This dual-threshold processing is able to reduce omission errors but not to increase commission errors.
- A geometrical index (GI) is defined to represent the geometrical characteristics of buildings (e.g., rectangular fit and length-width ratio). GI is able to remove noise such as roads. In addition, the normalized difference vegetation index (NDVI) is used to reduce the commission errors of bright vegetation.

The proposed building extraction system is constructed on the aforementioned four spectral and spatial indices (e.g., MBI, MSI, GI, and NDVI). The main differences between the proposed method and the work presented by Jin and Davis [16] lie in the following aspects: 1) In our study, the building or shadow information is described by a one-band feature image

(MBI and MSI, respectively), which is more straightforward for representation of buildings or shadows compared to the direct use of the high-dimensional morphological profiles in [16]; 2) The proposed MBI and MSI indices are built on a series of linear structuring element (SE) that are more appropriate than the disk-shaped SE used in [16] for detecting buildings with different spatial characteristics such as directionality and shape; 3) The proposed method is implemented on an object-based image processing framework since it is convenient to define contextual relationship and provide vector output for database construction and updating.

The rest of this paper is organized as follows. Section II describes the proposed morphological building/shadow index. The building extraction system is presented in Section III. Section IV analyzes and compares the experimental results, followed by the conclusion in Section V.

II. MORPHOLOGICAL BUILDING/SHADOW INDEX

A. Morphological Building Index (MBI)

The basic idea of MBI is to build the relationship between the spectral-structural characteristics of buildings and the morphological operators.

- *Brightness*: The maximum value of a pixel at all the visible bands is recorded as brightness of the pixel. The visible bands are focused on since they have the most significant contribution to the spectral information of buildings. Furthermore, the white top-hat transformation is used to highlight bright structures with a size up to a pre-defined value.
- *Local contrast*: The relatively high reflectance of roofs and the spatially adjacent shadows lead to a high local contrast of buildings. The differential morphological profiles (DMP) [2] of the white top-hat are used to represent the local contrast of bright structures.
- *Size and Directionality*: A challenging task for construction of a building index is to automatically filter out roads that have very similar spectral reflectance as buildings. Roads are always elongated in one or two directions while buildings are more isotropic. Consequently, the morphological building index is implemented on a series of linear structural elements (SE) that are able to measure the size and directionality of structures [11].
- *Shape*: Taking into account the rectangularity of buildings, the length-width ratio is used to remove narrow and elongated structures.

Buildings can be automatically extracted by the following steps:

- Step 1) Calculation of brightness: The maximum of multispectral bands for pixel x is recorded as its brightness value:

$$b(x) = \max_{1 \leq k \leq K} (\text{band}_k(x)) \quad (1)$$

where $\text{band}_k(x)$ indicates the spectral value of pixel x at the k th spectral band and K is the number of multispectral bands.

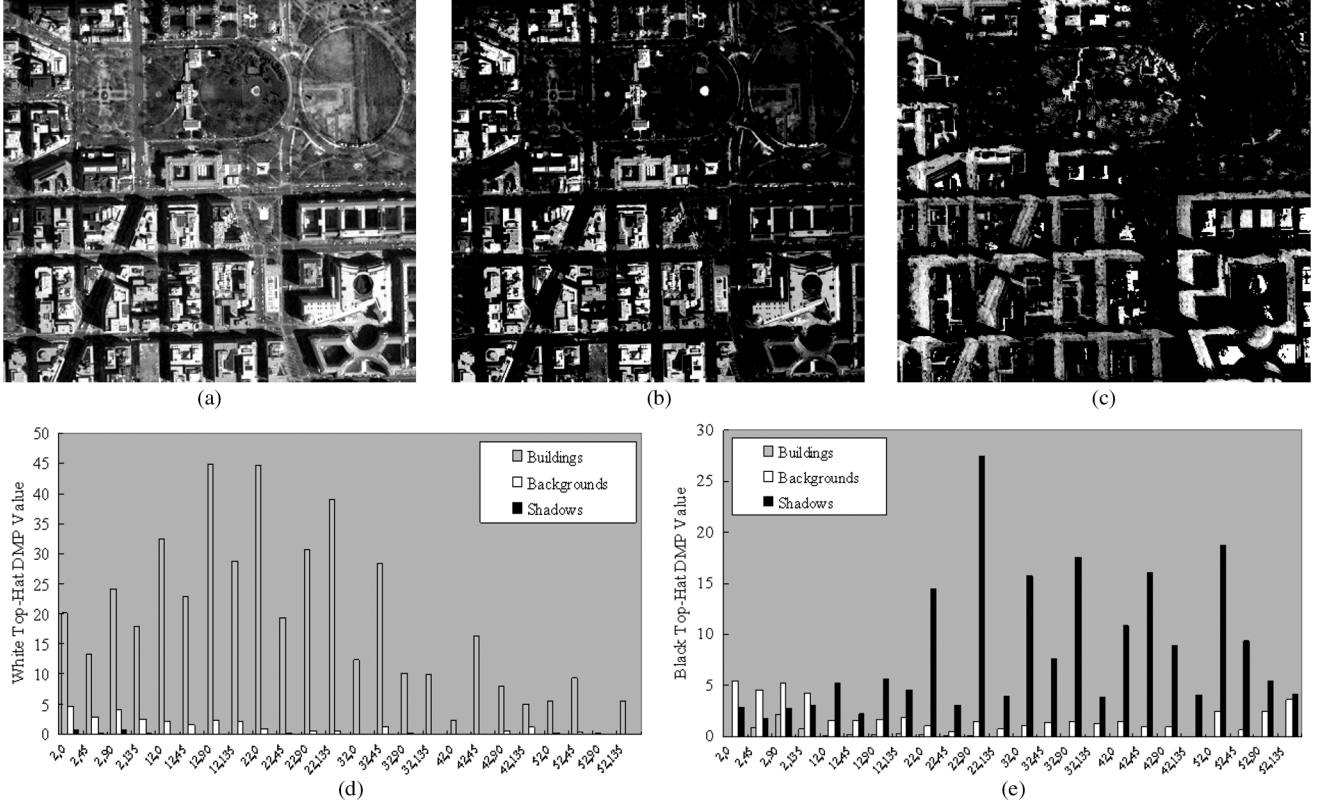


Fig. 1. (a), (b), and (c) are the original image, MBI and MSI feature images, respectively. (d) and (e) show the white and black top-hat histograms that are used to calculate the MBI and MSI, respectively.

— Step 2) Construction of MBI: The spectral-structural characteristics of buildings (e.g., contrast, size, and directionality) are represented using the DMP of the top-hat by reconstruction with a series of linear SE. The morphological operators used for construction of MBI are introduced below.

1) White top-hat by reconstruction (W-TH):

$$\text{W-TH}(d, s) = b - \gamma_b^{re}(d, s), \quad (2)$$

where γ_b^{re} represents the opening-by-reconstruction [12] of the brightness image, and s and d indicate length and direction of a linear SE, respectively.

2) Morphological profiles (MP): MP of the white top-hat is defined as,

$$\begin{cases} \text{MP}_{\text{W-TH}}(d, s) = \text{W-TH}(d, s) \\ \text{MP}_{\text{W-TH}}(d, 0) = b \end{cases} \quad (3)$$

3) Differential morphological profiles (DMP): The DMP of the white top-hat is defined as

$$\text{DMP}_{\text{W-TH}}(d, s) = |\text{MP}_{\text{W-TH}}(d, (s + \Delta s)) - \text{MP}_{\text{W-TH}}(d, s)| \quad (4)$$

where Δs is the interval of the profiles and $s_{\min} \leq s \leq s_{\max}$.

The morphological building index (MBI) is defined as the average of DMP of the white top-hat:

$$\text{MBI} = \frac{\sum_{d,s} \text{DMP}_{\text{W-TH}}(d, s)}{D \times S} \quad (5)$$

where D and S denote the numbers of directionality and scale of the profiles, respectively. Four directions are considered in this study ($D = 4$) since increase of D did not result in higher accuracy for building detection [10]. The sizes of SE (s_{\max} , s_{\min} , and Δs) should be determined according to the spatial characteristics of buildings and the spatial resolution (in this study, $s_{\min} = 2$, $s_{\max} = 52$, and $\Delta s = 5$). The number of scales is calculated by $S = ((s_{\max} - s_{\min})/\Delta s) + 1$. Construction of MBI is based on the fact that building structures have larger values in most of the directions of the white top-hat DMP histogram since they show high local contrast in these directions. As a consequence, buildings often refer to large MBI values.

— Step 3) Building extraction: structures that satisfy the following conditions are identified as buildings:

- 1) $\text{MBI}(x) \geq T_B$,
- 2) $\text{NDVI}(x) < t_1$,
- 3) $\text{Ratio}(x) < t_2$,
- 4) $\text{Area}(x) \geq t_3$,

where $\text{MBI}(x)$, $\text{NDVI}(x)$, $\text{Ratio}(x)$, and $\text{Area}(x)$ denote the building index, vegetation index, length-width ratio,

and area for an image structure or object x , respectively. The length–width ratio is effective in suppressing bright and elongated roads, and NDVI is effective in removing bright vegetation. Area is used to delete small noise.

B. Morphological Shadow Index (MSI)

MSI can be viewed as a twinborn index of MBI since shadows show spectrally opposite but spatially similar characteristics as buildings. Consequently, construction of MSI is also based on the relationship between the spectral-structural characteristics of shadows and the corresponding morphological operators:

- *Brightness*: The brightness of a shadow structure should be smaller than a pre-defined threshold due to its low spectral reflectance.
- *Local contrast*: Shadows also show a high local contrast, but differently from buildings, they are significantly darker than neighboring structures. Therefore, the differential morphological profiles (DMP) of the black top-hat (B-TH) are used to represent the local contrast of shadows.
- *Size and Directionality*: Sizes of shadows are affected by factors such as sizes of buildings, the solar elevation, and distance between buildings, while directionality of shadows is related to the sun position. In this study, the parameters of MSI are set to the same values as MBI, because in general these parameters of shadows can be considered similar with their adjacent buildings.

Based on the above analysis, the calculation of MSI can be extended from MBI by replacing the white top-hat (W-TH) with the black top-hat (B-TH). Shadows are extracted by the following three steps:

- Step 1) Calculation of brightness ((1)).
- Step 2) Calculation of MSI: Equation (5) can be converted to

$$MSI = \frac{\sum_{d,s} DMP_{B-TH}(d,s)}{D \times S}. \quad (6)$$

The black top-hat is used to highlight dark structures:

$$B-TH(d,s) = \varphi_b^{re}(d,s) - b \quad (7)$$

where φ_b^{re} represents the closing-by-reconstruction of the brightness image. Large MSI values are more likely to be shadows.

- Step 3) Shadow extraction: Structures that satisfy the following conditions are identified as shadows:

- 1) $MSI(x) \geq T_S$,
- 2) $NDVI(x) < t1$,
- 3) $Bright(x) < t4$,

where $MSI(x)$, $NDVI(x)$, and $Bright(x)$ denote the shadow index, vegetation index, and brightness value for a structure x , respectively. Dark trees are removed by applying a threshold on the NDVI index. The brightness threshold is used to suppress the structures that are relatively darker than their neighborhood but have high spectral reflectance.

A graphical example for calculation of MBI and MSI is demonstrated in Fig. 1, where (a), (b) and (c) show the original image, MBI, and MSI feature images, respectively.

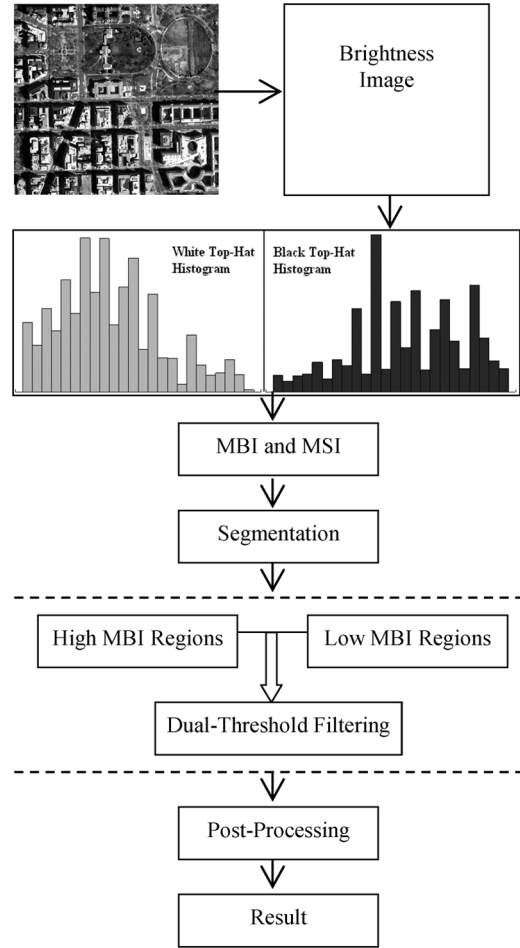


Fig. 2. Flow chart of the proposed building extraction framework.

Fig. 1(d) and (e) are, respectively, the white and black top-hat histograms that are used to calculate the MBI and MSI images. The horizontal axis represents sizes of the linear elements with different directions (0° , 45° , 90° , 135°), and the vertical axis denotes values of the white and black top-hat DMP, respectively, in Fig. 1(d) and (e). For instance, (22, 45) in the horizontal axis means the top-hat DMP values with $d = 45^\circ$ and $s = 22$. The white and black top-hat DMP histograms are computed based on 7,098, 8,802 and 10,248 pixels that are manually chosen from the subset image for buildings, shadow and backgrounds, respectively. From (a), (b) and (c), it can be seen that the proposed MBI/MSI is effective in highlighting buildings/shadows and at the same time suppressing backgrounds (e.g., bare soil, roads, vegetation, and trails). This observation is also supported by Fig. 1(d) and (e) since the feature values of objects of interest are significantly larger than backgrounds in most of scales and directions.

III. BUILDING EXTRACTION FRAMEWORK

The original MBI algorithm is subject to commission and omission errors. The commission errors are usually related to bright objects such as bare soil and bright roads, while the omission errors often refer to dark or heterogeneous roofs. In order to address these problems and improve the original algorithm,

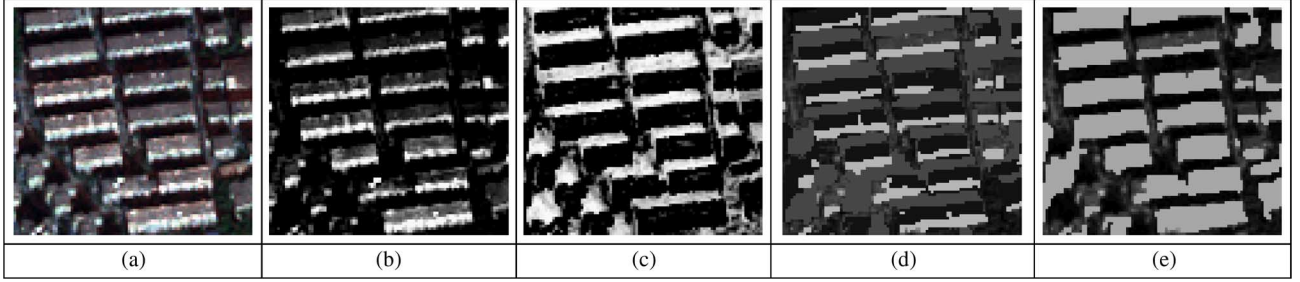


Fig. 3. A small example showing the steps of MBI filtering: (a) the example image, (b) and (c) the MBI and MSI feature images, (d) the result of the MBI filtering, and (e) the final result combining both low and high MBI regions.

TABLE I
PARAMETERS OF THE AUTOMATIC BUILDING EXTRACTION IN IKONOS DC DATASET

	MBI	MSI	Dist	GI	NDVI
High MBI regions	$T_{B_high} = 3$	$T_S = 2$	$D_high = 20$	$T_G = 1.1$	NA
Low MBI regions	$T_{B_low} = 2$	$T_S = 2$	$D_low = 10$	$T_G = 1.1$	NA

a novel building extraction system is proposed by considering the following aspects.

- After the MSI-based shadow detection, the distance to shadows is used as a spatial constraint to suppress bright noise such as soil, open areas and roads.
- The MBI images with high and low feature values are filtered with different spatial constraints, respectively.
- A geometrical index and a vegetation index are used in the post-processing.

The proposed system consists of four blocks: 1) pre-processing, 2) segmentation, 3) filtering of MBI images, and 4) post-processing.

- 1) Pre-processing: This includes calculation of the brightness, MBI, and MSI feature images.
- 2) Segmentation: The proposed method is implemented on an object-based framework because it is able to describe shape attributes and contextual relationship. In addition, the object-based output is very appropriate for construction and update of the building information system. In this study, the bottom-up region merging algorithm [13] is used for segmentation. In an iterative way, image objects are merged into larger ones at each subsequent step. The region merging decision is made with local heterogeneity criteria H , which takes into account spectral, shape, and the MBI features:

$$H = w_{\text{spectral}} \cdot H_{\text{spectral}} + w_{\text{shape}} \cdot H_{\text{shape}} + w_{\text{MBI}} \cdot H_{\text{MBI}} \quad (8)$$

where H_{spectral} , H_{shape} and H_{MBI} are heterogeneity measures of spectral, shape and the MBI components, respectively. w_{spectral} , w_{shape} , and w_{MBI} are their weights. The heterogeneity for the spectral and MBI components is measured by the variance of pixels that constitute an object, while the shape heterogeneity is defined based on the smoothness and compactness of an object. When a possible merge of a pair of image objects is examined, the merge is performed when the heterogeneity H is smaller than a user-defined scale parameter. The segmentation process stops as soon as this condition cannot be met by any possible merge. It should be underlined that over-segmentation is preferred compared to under-segmentation

TABLE II
ACCURACIES OF BUILDING DETECTION FOR IKONOS DATASET

Method	Accuracies			
	OE (%)	CE (%)	OA (%)	Kappa
Original MBI $T_B = 2$	13.4	6.1	90.5	0.811
Original MBI $T_B = 3$	15.5	4.6	90.3	0.806
DMP-SVM	16.0	4.2	90.2	0.805
The Proposed	10.0	1.6	94.3	0.886

since in the case of under-segmentation, buildings are always merged with their adjacent units such as trails and open areas.

- 3) Filtering of MBI: It is difficult to find a balance between commission and omission errors for the original MBI algorithm, since a high threshold T_B results in small commission but large omission errors and a low threshold results in small omission but large commission errors. An effective approach for addressing this problem is to split the MBI image into high-MBI and low-MBI regions and then process them separately with different spatial constraints. The rule of filtering of the MBI image can be expressed as follows:

IF $MBI(x) \geq T_{B_high}$ **AND** $Dist(x) < D_high$,

THEN x is identified as a building structure;

IF $T_{B_low} \leq MBI(x) < T_{B_high}$ **AND** $Dist(x) < D_low$,

THEN x is identified as a building structure,

where T_{B_high} and T_{B_low} represent the high and low thresholds for the MBI image, and D_high and D_low denote the high and low thresholds for the distance to shadows $Dist(x)$. The distance between two objects is calculated by measuring the distance of the smallest enclosing rectangles of the considered objects. Readers can refer to [15] for detailed description about the distance calculation. A large distance threshold is assigned to high MBI regions, and a small distance threshold is assigned to low MBI regions. The weak and strong conditions are imposed on the high

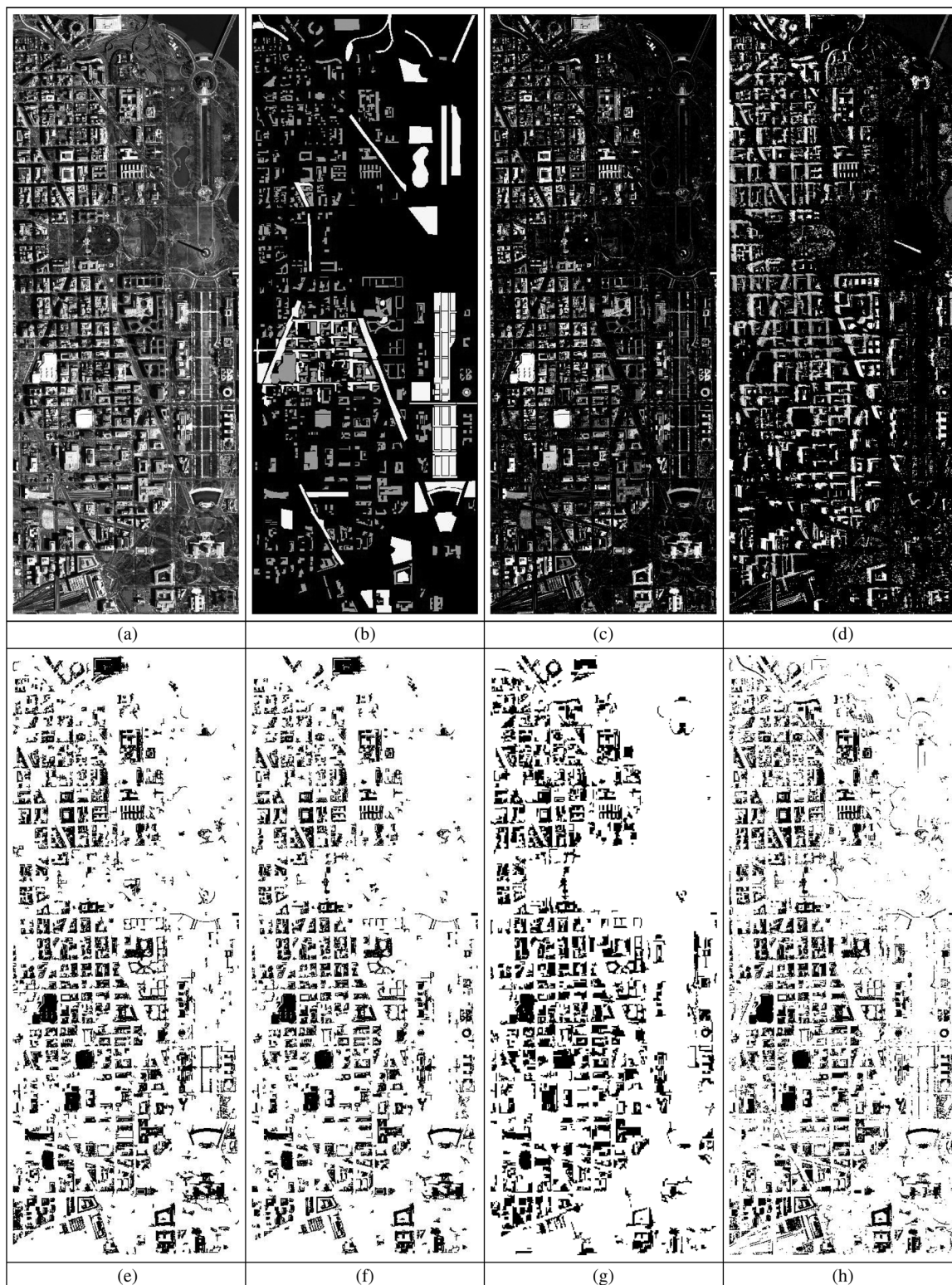


Fig. 4. Results of building extraction for Ikonos DC Mall dataset: (a) and (b) are the RGB image and the ground truth reference of the study area; (c) and (d) are MBI and MSI feature images, respectively; (e) and (f) are generated by the original MBI algorithm with $T_B = 2$ and 3, respectively; (g) is the result of the proposed method; (h) shows the building map produced by the support vector machine (SVM).

and low MBI regions, respectively, in order to reduce the omission errors and at the same time not to increase the commission errors. In this way, the risk of wrongly assigning a low MBI region to a building is reduced.

4) Post-processing: The post-processing strategy of the proposed framework is similar with the original MBI algorithm. The shape attribute of buildings and the NDVI are used to refine the initial result. A geomet-

rical index (GI) is introduced for shape description of buildings:

$$GI = a \cdot \frac{\text{Rectangular Fit}}{\text{Length-Width Ratio}} \quad (9)$$

where a is a coefficient that is used to adjust the range of the GI values ($a = 10$ in this study). The rectangular fit is defined based on the creation of a rectangle with the same area as the considered object. It is calculated by comparing the number of pixels inside the rectangle and the total number of pixels for the considered object [15]. A building structure is expected to have a large GI since it corresponds to a high rectangular fit and a small length-width ratio. In experiments, the structures that have a GI value smaller than a threshold T_G are filtered out in the post-processing. In addition, the NDVI is used to remove bright vegetation by performing a threshold ($NDVI(x) < t1$ and $t1 = 0.15$ in this study). The whole processing flow is presented in Fig. 2.

In Fig. 3, a small example including several building areas are used to present the filtering of MBI. (a) is the example image. (b) and (c) show the MBI and MSI feature images, respectively. (d) shows the result of the filtering, where green and blue represent high and low MBI regions, and magenta represents shadow. (e) is the final result that integrates both high and low MBI regions (orange represents building areas). From the figures, it can be seen that all the high MBI regions are retained in the final result while structures with low MBI values are detected and added into the final result when they are close to shadows. In this manner, the omission errors are reduced and at the same time false alarms are suppressed due to the spatial constraint of shadows.

IV. EXPERIMENTS AND ANALYSIS

A. Ikonos DC Mall Dataset

This experiment was conducted on an Ikonos image with three visible bands of 1-m spatial resolution over the Washington DC Mall. The study image and a manually delineated ground truth map are shown in Fig. 4(a) and (b). Numbers of samples for buildings and backgrounds are 116,159 and 117,836 pixels, respectively. The backgrounds include bright vegetation, soil, roads, open areas, and trails.

Parameters of this experiment are listed in Table I. Due to the unavailability of the near infrared band, the NDVI is not used in this experiment. Quantitative statistics of the results are compared in Table II, where omission errors (OE) and commission errors (CE) are used to evaluate the accuracy of building detection, and overall accuracy (OA) and Kappa coefficient are used to assess the accuracy of discrimination between buildings and backgrounds. The first comment to Table II is that the proposed method significantly suppresses both omission and commission errors of the original MBI algorithm. Compared to the original algorithm, the proposed method reduces the omission errors by 3.4% and 5.5%, respectively, for $T_B = 2$ and 3, and at the same time the commission errors decrease by 4.5% and 3.0%, respectively. Furthermore, the overall accuracy increases

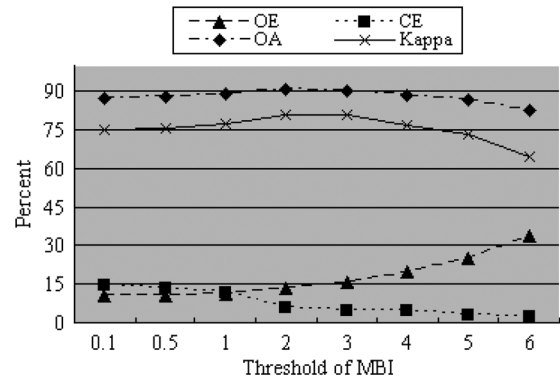


Fig. 5. Relationship between the accuracies of building detection and the MBI thresholds (T_B) of the original algorithm (Ikonos DC Mall dataset).

from 90.5% ($T_B = 2$) and 90.3% ($T_B = 3$) to 94.3%, and the Kappa coefficient increases from 0.811 and 0.806 to 0.886, which shows that the proposed system is more effective for discrimination between buildings and backgrounds. Considering that a large number of samples (233,995 pixels) in the reference image were used for accuracy assessment, the improvements achieved by the proposed method can be regarded as promising. The extracted building maps for $T_B = 2$, $T_B = 3$ of the original MBI algorithm and the proposed method are displayed in Fig. 4(e), (f) and (g), respectively, for a visual inspection.

The statistical accuracies of the original MBI algorithm with different binarization threshold values (T_B) are shown in Fig. 5. By observing the curves of OE and CE, it can be found that a small value of T_B corresponds to a small omission error but a large commission error, and a large T_B is related to a small commission error but a large omission error. From the figure, it can be seen that $T_B = 2$ achieves the highest overall accuracy (90.5%), which can be viewed as a balance between omission and commission errors. Comparing the accuracies in Fig. 5 and Table II, it can be stated that the dual-threshold filtering is more effective for building extraction from the MBI feature image compared to the traditional simple threshold algorithm. The better performance of the dual-threshold filtering should be attributed to introduction of the morphological shadow index (MSI) that provides complementary information to the MBI.

In order to further evaluate the effectiveness of the proposed method, a supervised machine learning approach is carried out for comparison. Specifically, the multiscale and multidirectional DMPs for the white and black top-hat, which are used to produce the MBI and MSI, are concatenated with the spectral bands and then interpreted by a support vector machine (SVM) [4]. In this experiment, 200 building pixels and 300 background pixels were randomly selected from the ground truth for training of the DMP-SVM. Its building detection result is shown in Fig. 4(h) and accuracies are reported in Table II. The supervised approach (i.e., DMP-SVM) gave comparable result with the original MBI algorithm, but its accuracies were significantly lower than the proposed method. With the supervised approach, a large number of training samples are needed to precisely model the features of buildings. In addition, the process of machine learning is always time-consuming. Consequently, the proposed building detection

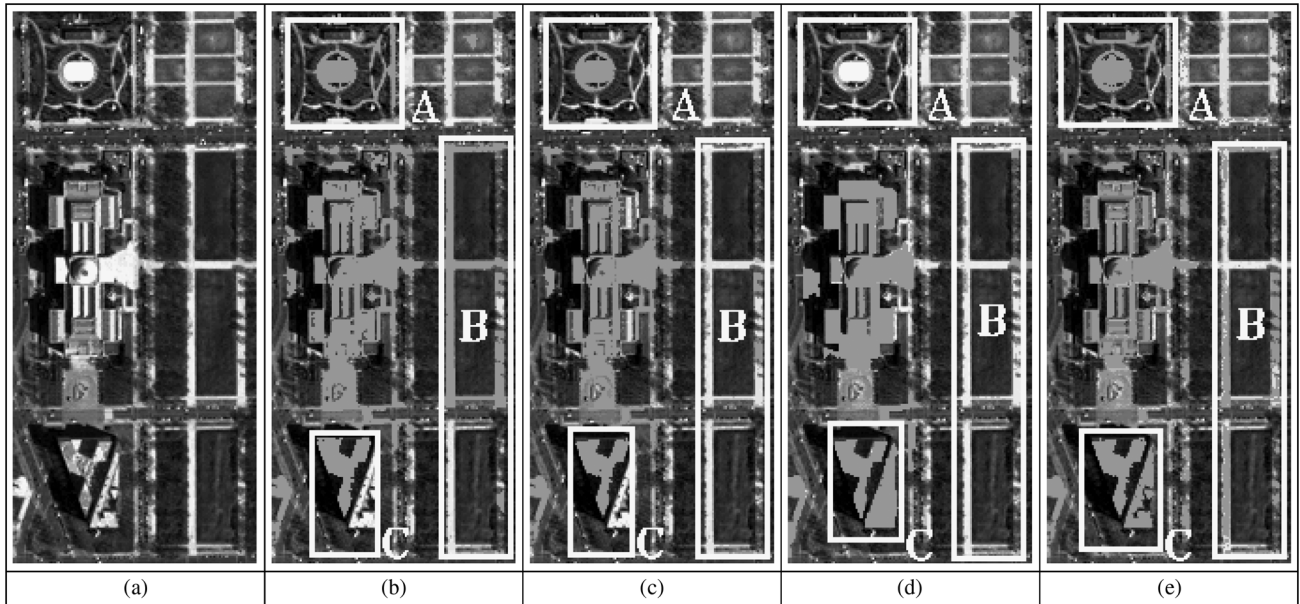


Fig. 6. Building detection results of a subset image (a) for $T_B = 2$ (b), $T_B = 3$ (c), the proposed method (d), and the DMP-SVM (e). The extracted buildings are highlighted in orange and overlaid on the original image.

method is a more suitable option for building extraction, especially for the large amount of urban high-resolution imagery.

The building detection results of different algorithms in a subset image are presented in Fig. 6. The challenge for this subset is to accurately extract buildings and at the same time suppress the trails (region B) and the open area (region A). By comparing the results in (b) and (c), it can be seen that a small value of T_B leads to more commission errors as the trails in region B are wrongly identified as buildings for $T_B = 2$ but are removed for $T_B = 3$. However, in Fig. 6(c), an open area in region A is incorrectly labelled as a building due to its high MBI value. In addition, a heterogeneous roof in the region C is omitted. The proposed method shown in Fig. 6(d) overcomes the above problems. The commission errors in region A and B are removed by the spatial constraint of shadows. Furthermore, the omission error of the heterogeneous roof in region C is avoided. The reason is that the original MBI feature calculated by the DMP is sensitive to the locally spectral variations of the heterogeneous roofs, but the object-based processing is able to smooth the local heterogeneity and focus on the whole structures of interest by courtesy of the segmentation. The supervised approach (i.e., DMP-SVM) overcomes the omission in region C but it gives poor results in regions A and B.

B. WorldView-2 Hangzhou Dataset

A WorldView-2 satellite image over Hangzhou, east of China, is used in this experiment for a further evaluation of the proposed building detection method. The WorldView-2 satellite, launched on October 8, 2009, is able to provide eight multispectral bands with 2-m spatial resolution. The five visible bands (coastal, blue, green, yellow, and red) are used to generate the brightness image, and the first near infrared channel and the red channel are used to compute the NDVI. The study image and the ground truth reference are shown in Fig. 7(a) and (b), and the

MBI and MSI feature images are displayed in Fig. 7(c) and (d). Numbers of samples for buildings and backgrounds are 49,184 and 43,479 pixels, respectively. The backgrounds include roads, bare soil, playgrounds, and bright vegetation.

Parameters of this experiment are listed in Table III. Quantitative statistics of the results are compared in Table IV. Once again, it is found that a small MBI threshold (T_B) corresponds to small omission but large commission errors, and a large value leads to small commission but large omission errors. As depicted in Fig. 8, it is difficult to find a simple threshold to simultaneously reduce omission and commission errors. However, the proposed method effectively addresses this problem by introducing the spatial constraint of shadows with a dual-threshold approach. Compared to the original MBI algorithm with $T_B = 0.5$, the proposed method gives a similar omission error (16.1%) and a substantially lower commission error (0.8%). Comparing the original MBI algorithm and the proposed method, the overall accuracy increases from 87.5% ($T_B = 0.5$) and 69.8% ($T_B = 2$) to 91.1%, and the Kappa coefficient increases from 0.751 and 0.415 to 0.823, respectively. The building detection results for the original MBI algorithm are presented in Fig. 7(e) and (f), respectively, for $T_B = 0.5$ and $T_B = 2$.

The supervised machine learning algorithm is also implemented for comparison. In this experiment, two kinds of spatial features are fed into the SVM classifier, respectively. The first is the black and white top-hat DMPs that are used to generate the MBI and MSI. The second is a recently developed multi-scale urban complexity index (MUCI) [14], which is built on the 3D wavelet transform and can be used to discriminate different urban structures such as roads, buildings, trees and grass. In this experiment, 50 building pixels and 300 background pixels were selected from the ground truth for training of SVM. The extracted buildings for the DMP-SVM and MUCI-SVM are displayed in Fig. 7(h) and (i), and their accuracies are reported in

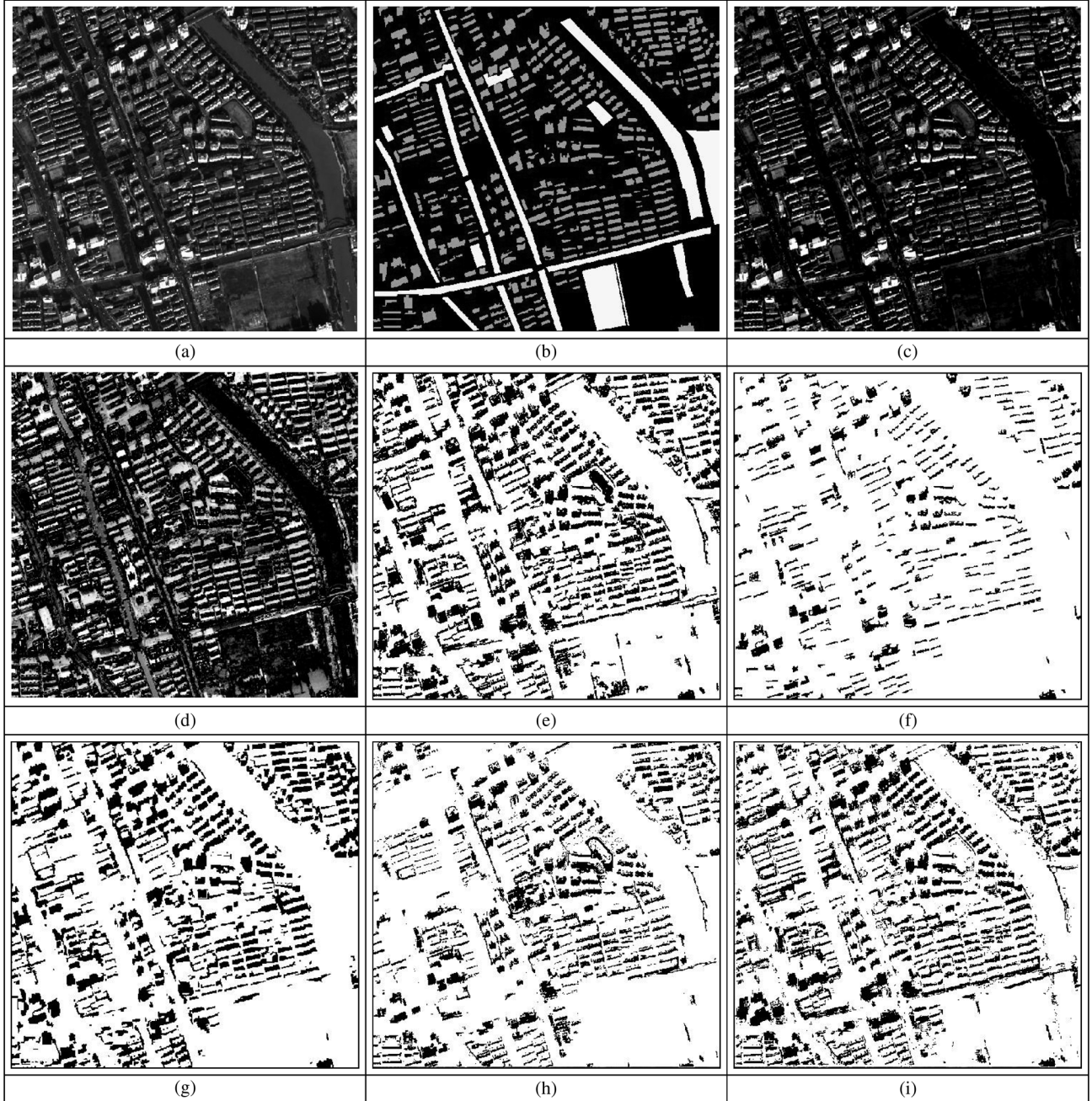


Fig. 7. Results of building extraction for WorldView-2 Hangzhou dataset: (a) and (b) are the RGB image and the ground truth reference; (c) and (d) show the MBI and MSI feature images, respectively; (e) and (f) are generated by the original MBI algorithm with $T_B = 0.5$ and 2, respectively; (g) is the result of the proposed method; (h) and (i) present the building maps extracted by DMP-SVM and MUCI-SVM, respectively.

TABLE III
PARAMETERS OF THE AUTOMATIC BUILDING EXTRACTION IN WORLDVIEW-2 HANGZHOU DATASET

	MBI	MSI	Dist	GI	NDVI
High MBI regions	$T_{B_high} = 2$	$T_S = 0.5$	$D_high = 35$	$T_G = 1.1$	$t1 = 0.15$
Low MBI regions	$T_{B_low} = 0.5$	$T_S = 0.5$	$D_low = 5$	$T_G = 1.1$	$t1 = 0.15$

Table IV. In this experiment, the DMP-SVM does not give satisfactory result ($OA = 79.2\%$, Kappa coefficient = 0.592), in particular, its omission error is as high as 36.7%. On the other hand, the MUCI-SVM achieves acceptable results since it outperforms all the simple-threshold MBI algorithms.

In Fig. 9, a subset image of the WorldView-2 dataset is used for a detailed comparison. The challenge for building extrac-

tion from this subset is to accurately identify the buildings in the upper half of the image and at the same time suppress the noise such as roads and bright soil in the bottom half of the image. As observed in Fig. 9(b) and (c), the original algorithm cannot simultaneously overcome both omission and commission errors. The proposed method is able to accurately delineate buildings and suppress the noise such as roads and soil. In this

TABLE IV
ACCURACIES OF BUILDING DETECTION FOR WORLDVIEW-2 DATASET

Method	Accuracies			
	OE (%)	CE (%)	OA (%)	Kappa
Original MBI $T_B = 0.5$	15.6	8.5	87.5	0.751
Original MBI $T_B = 2$	56.7	0.4	69.8	0.415
DMP-SVM	36.7	3.7	79.2	0.592
MUCI-SVM	18.9	4.6	87.9	0.760
The Proposed	16.1	0.8	91.1	0.823

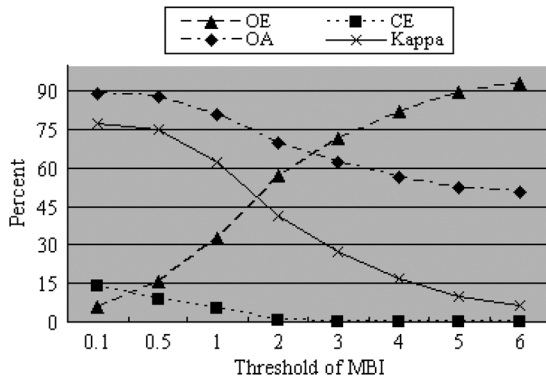


Fig. 8. Relationship between the accuracies of building detection and the MBI thresholds (T_B) of the original algorithm (WorldView-2 Hangzhou dataset).

subset image, the supervised approaches do not show satisfactory results as some buildings are omitted in the DMP-SVM (e) and the noise of roads is not removed in both DMP-SVM and MUCI-SVM ((e) and (f)).

C. Parameter Analysis

In this subsection, suggestions for the parameter setting are given and the sensitivity of some critical parameters is analyzed.

- 1) Parameters of the dual-threshold filtering (T_{B_high} , T_{B_low} for the MBI, and D_high , D_low for the distance): According to our previous work [10] and the current study, the suitable threshold of MBI ranges from 0.5 to 3 for different types of images such as GeoEye-1, Ikonos, and WorldView-2. Therefore, the suggested values for T_{B_low} and T_{B_high} are within the range of [0.5, 3]. In addition, experiments show that a large threshold of MBI is appropriate for an image that has large local contrast (and vice versa) because MBI is calculated by measuring the local contrast. On the other hand, the distance parameters (D_high and D_low) should be adjusted in different experimental environments according to the spatial characteristics of images such as resolution, building sizes, etc.
- 2) Threshold of the Morphological Shadow Index (T_S): The suggested value for T_S (threshold of MSI) is equal to T_{B_low} since in general the parameters of shadows can be considered similar with their adjacent buildings.
- 3) Threshold of NDVI ($t1$): NDVI is not a critical parameter for building extraction from urban areas. Its threshold was set to $t1 = 0.15$ in this study. It should be noted that satisfactory results were obtained in the Ikonos DC Mall experiment although the NDVI feature was not used in

this dataset because of unavailability of the near infrared channel.

- 4) Threshold of Geometrical Index (T_G): The suggested value of GI is within the range [0.5, 1.1]. Small values of GI correspond to elongated and narrow structures that are more likely to be roads. In order to analyze the parameter sensitivity of T_G , the building detection accuracies with a series of parameter values are reported in Table V. From the table, it can be seen that high accuracies (over 91% for the OA and larger than 0.82 for the Kappa coefficient) were achieved for $T_G = 0.5, 0.7, 0.9,$ and 1.1 . When T_G increases gradually from 1.3, the omission errors (OE) become larger since more building objects are removed. It should be underlined that the thresholds of GI are stable as the shape characteristics of buildings are relatively stable for different image scenes.
- 5) Scale parameter for the segmentation: The heterogeneity threshold for the segmentation should be tuned in different experiments (it was set to 15 in this study). In order to discuss the impact of the segmentation on the final building detection, the performances for different segmentation scale parameters are compared in Table VI.

The table shows that the scale parameters of 15 and 20 achieve the best results. In addition, the accuracies for scales 10 and 25 are also satisfactory since they outperform the original MBI algorithm and the SVM-based methods (see Table IV).

V. CONCLUSION

In this study, a building detection system is proposed based on a recently developed morphological building index (MBI). The original algorithm performs a simple threshold on the MBI feature image and hence leads to commission and omission errors. As pointed out in [10], the commission errors are mainly related to roads, soil, and open areas due to their similar spectral and structural characteristics with buildings; the omission errors correspond to heterogeneous roofs as they have low MBI values and are then removed in the binarization. The contribution of this study lies in construction of the morphological shadow index (MSI), and joint use of the MBI and MSI in an object-based framework for building extraction. MBI and MSI are built on white and black top-hat transform used to highlight the bright and dark structures in an image, respectively. MSI is able to automatically indicate presence of shadows that are subsequently used as spatial constraint of buildings. Some parameters, such as MBI/MSI thresholds and the segmentation scale, should be set in the processing chain by users.

The proposed method was validated on two high-resolution datasets: the Ikonos DC Mall image and the WorldView-2 Hangzhou image. The experimental results revealed that the proposed method was able to simultaneously reduce the commission and omission errors. This conclusion is supported by both visual inspection and accuracy assessment. In the Ikonos dataset, the overall accuracy (OA) of building extraction increased from 90% to 94%, and in the WorldView-2 dataset, the OA increased from 70% and 88% to 91%. The reasons why the proposed system significantly improved the original MBI algorithm are summarized as follows:

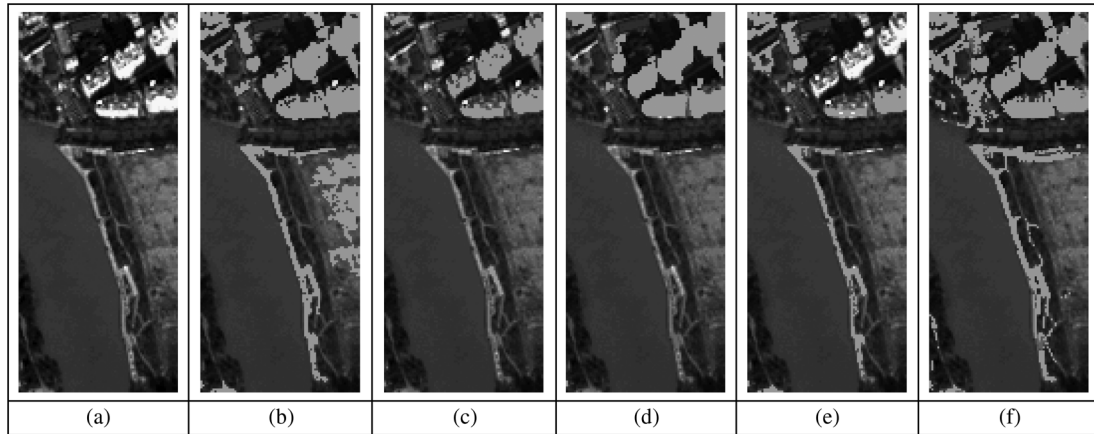


Fig. 9. Building detection results of a subset image (a) of WorldView-2 dataset for $T_B = 0.5$ (b), $T_B = 2$ (c), the proposed method (d), the DMP-SVM (e), and the MUCI-SVM (f). The extracted buildings are highlighted in orange and overlaid on the original image.

TABLE V
SENSITIVITY ANALYSIS FOR THE PARAMETER OF GI (WORLDVIEW-2 DATASET)

Value of T_G	Accuracies			
	OE (%)	CE (%)	OA (%)	Kappa
0.5	13.9	1.0	92.2	0.844
0.7	13.5	1.3	92.2	0.845
0.9	14.5	1.3	91.7	0.835
1.1	16.1	0.8	91.1	0.823
1.3	19.1	1.3	89.3	0.787
1.5	23.0	1.3	87.3	0.748
1.7	27.2	1.3	85.0	0.705
1.9	33.8	1.5	81.6	0.638
2.1	37.5	1.6	79.6	0.600

TABLE VI
SENSITIVITY ANALYSIS FOR THE SCALE PARAMETER OF SEGMENTATION
(WORLDVIEW-2 DATASET)

Scale Parameter	Accuracies			
	OE (%)	CE (%)	OA (%)	Kappa
10	21.5	1.7	87.9	0.760
15	16.1	0.8	91.1	0.823
20	14.6	2.3	91.1	0.824
25	17.9	1.3	89.9	0.800

- 1) Introduction of MSI provided spatial constraint for building extraction and hence removed the false alarms such as roads, soil, and open areas;
- 2) The object-based processing smoothed local variation and suppressed omission errors from the heterogeneous roofs. Moreover, the dual-threshold approach reduced omission errors from the dark roofs.

The proposed method was also compared with the supervised machine learning algorithms. In experiments, the differential morphological profiles (DMPs) [2] and a recently developed multiscale urban complexity index (MUCI) [14] were extracted from the test images, respectively, and then fed into the support vector machines (SVMs) for building detection. The experimental results showed that the proposed method outperformed the DMP-SVM and MUCI-SVM in terms of statistical accuracies and visual inspection. This phenomenon is promising since

the proposed method is able to achieve satisfactory results for building detection without collection of training samples and process of supervised learning.

REFERENCES

- [1] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 260–264, Apr. 2007.
- [2] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
- [3] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
- [4] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
- [5] Y. O. Ouma, T. G. Ngigi, and R. Tateishi, "On the optimization and selection of wavelet texture for feature extraction from high-resolution satellite imagery with application towards urban-tree delineation," *Int. J. Remote Sens.*, vol. 27, no. 1, pp. 73–104, Jan. 2006.
- [6] S. W. Myint, N. S. N. Lam, and J. Tyler, "Wavelets for urban spatial feature discrimination: Comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 7, pp. 803–812, Jul. 2004.
- [7] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE J. Sel. Topics. Appl. Earth Observ. Remote Sens. (JSTARS)*, vol. 1, no. 3, pp. 180–192, Sep. 2008.
- [8] M. Pesaresi and A. Gerhardinger, "Improved textural built-up presence index for automatic recognition of human settlements in arid regions with scattered vegetation," *IEEE J. Sel. Topics. Appl. Earth Observ. Remote Sens. (JSTARS)*, vol. 4, no. 1, pp. 16–26, Mar. 2011.
- [9] P. Gamba, M. Pesaresi, K. Molch, A. Gerhardinger, and G. Lisini, "Anisotropic rotation-invariant built-up presence index: Application to SAR data," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS'08)*, Boston, MA, Jul. 6–11, 2008, vol. 5, p. V-338-V.
- [10] 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, Jul. 2011.
- [11] P. Soille and H. Talbot, "Directional morphological filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1313–1329, Nov. 2001, 2001.
- [12] P. Soille, *Morphological Image Analysis: Principle and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

- [13] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, "A comparison of three image-object methods for the multiscale analysis of landscape structure," *ISPRS J. Photogramm. Remote Sens.*, vol. 57, no. 5, pp. 327–345, Apr. 2003.
- [14] X. Huang and L. Zhang, "A multiscale urban complexity index based on 3D wavelet transform for spectral-spatial feature extraction and classification: An evaluation on the 8-channel WorldView-2 imagery," *Int. J. Remote Sens.*, 2011, DOI:10.1080/01431161.2011.614287.
- [15] Definiens Developer 7, Reference Book. Definiens AG, Munich, Germany, 2007.
- [16] 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. 14, pp. 2196–2206, Jan. 2005.



Xin Huang received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIES-MARS), Wuhan University, China, in 2009.

He is currently an Associate Professor at the LIES-MARS, Wuhan University. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition and remote sensing applications. He has published more than 20 peer-reviewed articles in international journals

such as *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, *IEEE Geoscience and Remote Sensing Letters*, *Photogrammetric Engineering and Remote Sensing*, and *International Journal of Remote Sensing*.

Dr. Huang has served as a reviewer for most of the international journals for remote sensing. He was the recipient of the Top-10 Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the best paper in image analysis and interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS).



Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, ChangSha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as the head of the Remote Sensing Division. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education, China. He has more than 200 research papers and five patents. He is now Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing and artificial intelligence.

Dr. Zhang regularly serves as a Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR), Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and the Geoinformatics Symposia. He also serves as an Associate Editor of *International Journal of Ambient Computing and Intelligence* (IJACI), *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*. He is a Fellow of IEE, Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme, and Executive Member for the China Society of Image and Graphics, among others.