Classification of Urban Building Type from High Spatial Resolution Remote Sensing Imagery Using Extended MRS and Soft BP Network

Junfei Xie and Jianhua Zhou

***Abstract*—This study presents a new approach for classification of building type in complex urban scene. The approach consists of two parts: extended multiresolution segmentation (EMRS) and soft classification using BP network (SBP). The technology scheme is referred to here as EMRS-SBP. EMRS is used to guide the design of descriptor. A descriptor is a feature expression or a symbolized algorithm to systematically promote the expressing capability of image features. A classifier can perform far better to discern complex pattern of combining pixels working in an EMRS-based feature space constructed by a number of such descriptors. SBP serves as a classifier model to generate natural clusters of member which refers to here as both pixels and image patches. Class-mark ensured member is denoted as sure member and the rest as unsure (fuzzy) members. The latter can be relabeled through recursive defuzzifying according to the information carried by the gradually increased sure members. By using EMRS-SBP, three building types, i.e., old-fashioned courtyard dwellings, multistorey residential buildings, and high-rise buildings, can be accurately classified from high spatial resolution imagery in a feature space constructed with fifteen descriptors including nine EMRS-based ones. There is evidence that the mean overall accuracy using SBP in the EMRS-based feature space is 19.8% higher than that using the hard classification with BP network in a single resolution segmentation space and meanwhile, the mean kappa statistic value (κ) is 25.1% higher.**

***Index Terms*—Back propagation network, building types, multiresolution segmentation (MRS), soft classification, urban area.**

I. Introduction

BUILDING type (usage information) is one of the key input variables to demographic and socioeconomic models [1]–[3]. Automated extraction of urban building as a single category or as separate types from remotely sensed data is important to many applications, such as map updating, city modeling, urban growth analysis, change monitoring [4], etc. However, it is hard to classify building types in complex urban scene depending only on typical image features [5] due to spectral confusion. Different building types are often similar to each other in spectrum whereas buildings consisting a single type often appear different in spectra. The reason is that most buildings are constructed by limited artificial materials and the selections of the materials, in many cases, do not relate to building types. Furthermore, cooling tower, pavilion, ornamental landscape, etc., on the roofs cause the appearances of building in a high spatial resolution (HSR) image much complex.

A. Related Work

In the middle-to-late 1980s, researchers started to study the methodology of extracting urban buildings from aerial photos [6] and later from medium- or high-resolution satellite images [7]. More recently, finer texture and more accurate boundaries of building can be obtained from HSR imagery and applied to building extraction [8]–[11]. However, it is still difficult to discern building types from HSR imagery by computers because it is hard to find appropriate segmenting scales to completely capture even an individual building from complex patterns of combining pixels [12]. Therefore, most existing studies are either to classify buildings as a single category in terms of land use mapping [5], [7], [8], [13] or to distinguish between building and nonbuilding [9].

Elevation data and building contours have also been applied in classification of building type. Airborne light detection and ranging (LiDAR) is particularly useful to collect the elevation data for expression of building structural characteristics [14]. With a sole use of LiDAR data or a combining use of LiDAR and HSR data, previous difficulties to some applications have been overcome, such as extraction of building object [15]–[17], mapping of building boundary [18], [19], reconstruction of three-dimensional (3-D) building model [20], identification of building roof structure [21], classification of building type [22], etc. However, LiDAR data are costly and rarely available especially in China as compared with HSR imagery.

If accurate building contours are available, accuracy of discerning building types can often be improved. An example of building type classification using contours was reported by Du *et al.* [12] and the typical steps included:

1) acquiring building contours from a geographic information system (GIS),
2) merging pixels within a contour to form an image object,
3) deriving image features of the object, and

Manuscript received August 19, 2016; revised October 12, 2016, December 22, 2016, and February 9, 2017; accepted March 17, 2017. Date of current version August 9, 2017. This work was supported by the National Natural Science Foundation of China under Grant 51278056. *(Corresponding author: Jianhua Zhou.)*

J. Xie is with the Beijing Institute of Landscape Gardening, Beijing 100102, China (e-mail: xiejunfei@126.com).

J. Zhou is with the College of Geographic Science, East China Normal University, Shanghai 200062, China (e-mail: jhzhou@geo.ecnu.edu.cn).

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.2017.2686422

4) classifying all building objects in a given attribute space consisting of these features.

However, its applications are likely to be limited by the outdated data because the update of GIS is usually far behind an actual change of building, particularly in a region of rapid urbanization.

Consequently HSR-image-based monitoring systems for urban buildings are expected to play further crucial roles. Therefore, the testing data used in our scheme are conventional HSR images (e.g., Quickbird and Worldview). This also concerns some advantages of HSR imagery, such as easier to obtain, higher spatial resolution, shorter update cycle, lower cost, fewer requirements in data preprocess, etc., as comparing with those of LiDAR and GIS data. Although there is an obvious fact that majority of people can distinguish a number of building types from HSR imagery through visual interpretation, it is still hard to automatically discern these types by computers. The reason is likely that the existing algorithms of image classification cannot imitate the synthesis behavior of human brain yet. For example, a person may discern different building types by considering some patterns formed by randomly scattering pixels, such as area, shape, shadow, interval, density, etc. Classification accuracy may be improved by enabling a computer to "concern" not only the spectrum and texture features but also their combinative patterns. This study mainly focuses on the latter.

B. Contribution

Extended multiresolution segmentation (EMRS), a component of the EMRS-soft classification using BP network (SBP) scheme, is developed from multiresolution segmentation (MRS) [23] but runs a little further. As interesting objects in an image often have a range of brightness, roughness, and sizes, no single resolution is sufficient to capture all these characteristics. MRS is a bottom-up, region-merging technique that partitions image into image objects on the basis of homogeneity criteria controlled by user-defined scale parameters (SP). EMRS has two extensions to MRS. One is to increase measuring variables to enhance expressions of resolution changes thereby offering unique and often important insights to the differences between building types. The other is to replace definite SP with weighted sum matrices. This releases us from determination of SP which is a necessary but hard work to conventional MRS.

The back-propagation neural network (BPNN) is a widely accepted learning machine model especially for the classification from remote sensing imagery [24]–[26]. In addition, soft classification using BPNN is still in the process of improvement. SBP, the other component of the EMRS-SBP scheme, is characterized by generating a clustering prototype by soft partition and relabeling fuzzy members in the prototype through recursive defuzzifying. The relabeling is evaluated by severe fuzzy measures, thereby reducing the uncertainty.

II. Study Site and Test Data

Fig. 1 shows the study site. The rectangles in the right picture with serial numbers of 1 and 2 are two subsites involved with Worldview 3 images obtained in 2014 with an original spatial resolution of 0.5 m. The images were purchased from a geographic information service institution of the Chinese government. They were preprocessed with geometric correction and vegetation enhancement before selling. The other two with serial numbers of 3 and 4 are other two subsites involved with Quickbird images. They were recently copied from screen images of the Google Earth website with an original spatial resolution of 0.6 m. The image size of each subsite is listed in Table III. In order to ensure the universality of the proposed scheme, these subsites are randomly selected except the concern to include all building types in each site.

III. Method

A. Overview

In EMRS-SBP scheme, EMRS serves to guide the design of descriptor and SBP to generate a more natural classification. Steps for the classification of urban building type are as follows:

1) Smooth image signals by recursive wavelet compression to make image patches for an individual building as homogeneous as possible (see Section III-B).
2) Use three EMRS-based methods to express three kinds of information carrier. The methods are multistructural element morphological operations, multisized neighborhood statistics, and multithreshold segmentation. The carriers are dark details, building and shadow patches, and standard deviation of neighborhood elements (see Section III-C).
3) Construct feature space with EMRS-based descriptors which are usually expressed by the weighted summing of the discrete and hierarchical data obtained in step 2) to express not only spectrum and texture features but also their combinative patterns (see Sections III-C and III-D).
4) Use a soft partition to obtain a prototype of fuzzy cluster and then use a defuzzifying algorithm to relabel unsure members in the prototype (see Section III-E). Fig. 2 shows the entire technological process.

B. Image Preprocessing

The major obstacle to the extraction of building object is spectral confusion. Different building types are often similar to each other in spectrum whereas buildings consisting a single type often appear different in spectra. The band-based classification is likely to be adversely affected. This will cause the extracted building objects to be very broken. Wavelet compression is one of the useful algorithms for image homogenization thereby diminishing the brokenness. Although both wavelet compression and image smoothing can be used for the homogenization, we choose the former because of its two advantages: 1) removing redundant information while taking the homogenization; and 2) remaining fine structure at several specified resolutions. These advantages are useful for subsequent building classification.

We use a recursive process of wavelet compression to gradually attenuate unexpected noise (including spatially distributed noise and radiation pulse). The recursion does not stop until the number of gray scales of a processing image is less than

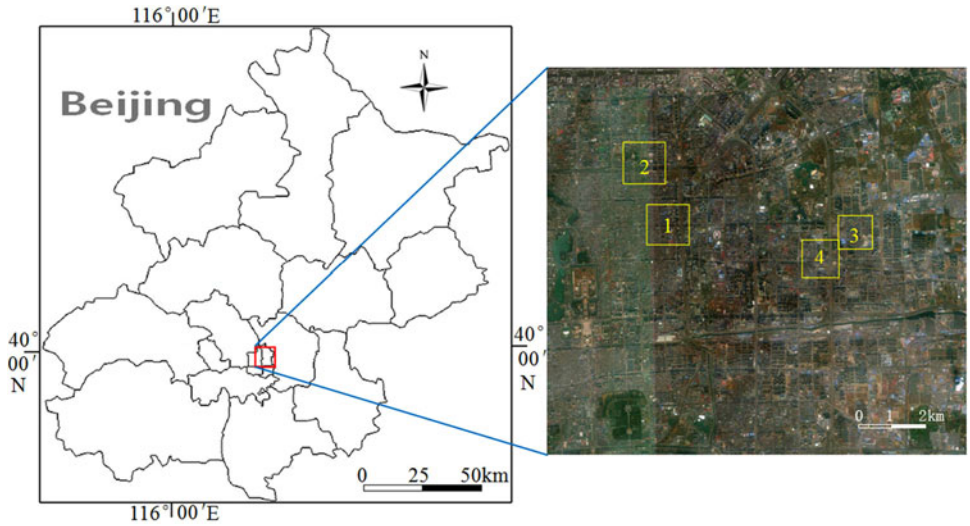

Fig. 1. Study site. The site is in the downtown area of Beijing, the capital of China. The right picture provides a more detail image of the site. The yellow rectangles on it show four subtesting sites and the digitals are the serial number of them, the same as the no. in Table III. The image was downloaded from the Google Earth website.

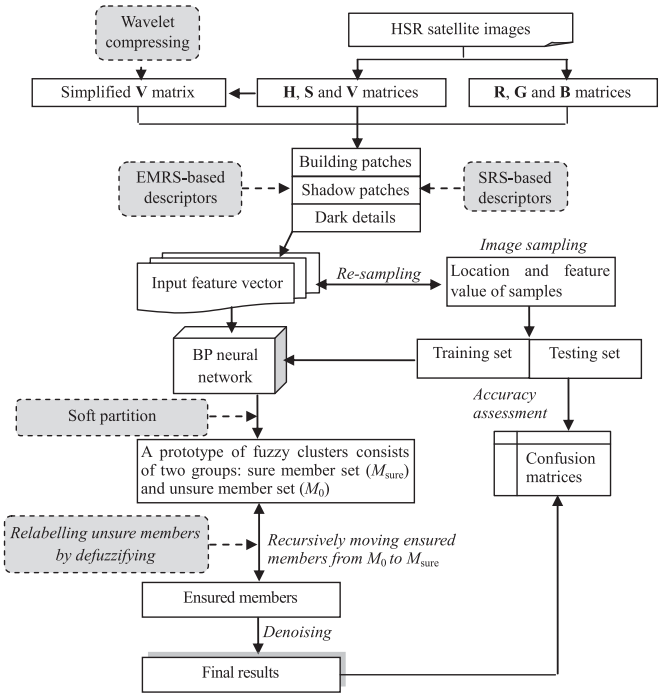

Fig. 2. Flowchart of EMRS-SBP. The main algorithms are represented by the dotted line rectangles.

2^3. A single compression will perform by calling "wdencmp," a MATLAB function. Fig. 3 shows an example of the recursive process. By comparing homogeneity of individual building from a compressed matrix **X** [see Fig. 3(c)] against its original **V** matrix [see Fig. 3(b)] it can be seen that most roofs become more homogeneous, and gray differences between roofs and other impervious surface (e.g., roads and plazas) are enlarged after the compression thereby diminishing the difficulties to extraction of building objects.

C. EMRS

The work to partition pixels into image objects by MRS depends on the homogeneity as specified by SP. However, it is hard to decide a proper SP for a category in complex scenes. There are two problems need to be solved in terms of building type classification. One is how to enable SP variable to adapt to variations of the scene. The other is how to comprehensively interpret the segments associated with the SP(s).

There are two extensions from MRS to EMRS to solve these problems:

1) Instead of using SP, three measure variables, i.e., range of spectral brightness, size of neighborhood, and size of structure element (SE) in regard to variations of radiation threshold, pixel combinative pattern, and member size respectively, are employed to indicate resolution changes.
2) Replacing a simple segmenting matrix with a weighted sum matrix.

The former is segmented with the threshold derived from a specified SP and the latter is a combination of multisegmentations with several thresholds derived from either multivalue ranges of a single feature or a group of features. Then, the latter will serve as either a single descriptor or a component of descriptor to express a kind of mixed model of spectrum or texture feature. The replacement not only enables EMRS to be embedded in the classification but also enriches the information carried by a descriptor. A feature space constructed by these descriptors may be properly complex for identification of seriously confused members whereas the dimension of feature space (the number of features) is unchanged. There is evidence that an EMRS-based descriptor always performs better in expression of complex and mixed pattern of pixels than the SRS-characterized version does where SRS means single resolution segmentation.

As mentioned before, there are three kinds of information carrier in EMRS. Relative analyzing approaches to them and involved descriptors will be introduced in detail as follows.

Fig. 3. Examples of image preprocess. (a) Original image (Worldview 3, 2014). (b) **V** matrix in HSV color model. (c) Compressed image (**X**). The arrows marked with 1, 2, and 3 represent three building types, i.e., the high-rise buildings, multistorey residential buildings, and old-fashioned courtyard dwellings.

1) Dark Details: The length along the north–south direction of a larger dark detail member associated with a building is likely to respond to the height of the building. Therefore, the density of dark details is selected as an indicator for building type classification. Dark details can be extracted through low cap transformation of mathematical morphology. The density of dark details ($\mathbf{D}_d$) was defined by Zhou *et al.* [27] as

$$\mathbf{D}_d(u,v) = \text{count}(\mathbf{B}_d(u,v))/A_{\text{imper}} \tag{1}$$

under the constraints:

$$\mathbf{B}_d = \{\mathbf{B}_d \subseteq \mathbf{I}, \forall b_d | (\mathbf{I} \bullet s - \mathbf{I}) > c \times d_{\text{mean}}\}$$

where $\mathbf{I}$ is a grayscale image; count(•) is a function to count true members in a size-given neighborhood; $\mathbf{B}_d$ denotes a binary image of dark details, and b_{d} is an element in $\mathbf{B}_d$; A_{imper} denotes area of impervious surface in the neighborhood; sign • represents morphological closing and s is its SE; c is a reserved coefficient; d_{mean} denotes the mean gray value of the low hat transform result, and d_{mean} serves as a base number for computation of the segmenting threshold.

In theory, certain size of dark details for a building type can be captured by selecting proper c and s. However, almost all tests for segmentation of building shadow failed when using single resolution of $\mathbf{D}_d$ as defined by (1) because the length and darkness of a shadow member change in different scenes. Therefore, we add EMRS to $\mathbf{D}_d$. That is, enable s to be adjustable in three different sizes to respond to the length changes and enable c to be adjustable also in three different levels to respond to the darkness changes. The newly defined EMRS-based $\mathbf{D}_d$, $\mathbf{D}_{d(\text{all})}$, is

$$\mathbf{D}_{d(\text{all})} = \begin{bmatrix} \mathbf{D}_{d(1,1)} & \mathbf{D}_{d(1,2)} & \mathbf{D}_{d(1,3)} \\ \mathbf{D}_{d(2,1)} & \mathbf{D}_{d(2,2)} & \mathbf{D}_{d(2,3)} \\ \mathbf{D}_{d(3,1)} & \mathbf{D}_{d(3,2)} & \mathbf{D}_{d(3,3)} \end{bmatrix}. \tag{2}$$

$\mathbf{D}_{d(\text{all})}$ is a cell array which consists of nine matrices of dark detail density in responding to variation on shadow length and darkness. The cell in row i and column j of $\mathbf{D}_{d(\text{all})}$ can be calculated by

$$\mathbf{D}_{d(i,j)} = \mathbf{D}_{s(i,j)} + \mathbf{D}_{m(i,j)} + \mathbf{D}_{a(i,j)} \tag{3}$$

where i denotes the serial number of SE and a smaller i indicates a smaller size of SE; j denotes the serial number of gray threshold and a smaller j indicates a lower threshold; $\mathbf{D}_{s(i,j)}$, $\mathbf{D}_{m(i,j)}$, and $\mathbf{D}_{a(i,j)}$ denote the density matrices calculated with a small, middle, and large sizes of neighborhoods, respectively, from $\mathbf{W}_{(i,j)}$ which is a binary image of dark details obtained with the ith SE and then the results are segmented with the jth threshold.

Then, a sum of density matrices as defined by (3) serves as a cell in $\mathbf{D}_{d(\text{all})}$ to more definitely determine whether a pixel is a dark detail or not. Furthermore, the cell row and column numbers in (2) indicate the variation on shadow length and darkness. For example, $\mathbf{D}_{d(2,3)}$ is obtained with a middle size SE and a higher grey threshold. If a pixel has a higher $\mathbf{D}_{d(2,3)}$, the pixel is likely within a middle-size and darker shadow region. Each cell in $\mathbf{D}_{d(\text{all})}$ can be used either individually if the cell has a determined meaning or in combination if the cell needs more constrains from the others. Fig. 4 shows the examples of weighted combination of selected cells from $\mathbf{D}_{d(\text{all})}$.

2) Building or Shadow Patches: Building or shadow patches are obtained through two methods, i.e., multithreshold segmentation and multisize window statistics. The methods enable EMRS to be embedded in the classification. The descriptor $\mathbf{D}_{\text{bu}(t)}$ calculated by using the former method is a combinative matrix of multidensity layers and each layer is derived from a binary image of building or shadow member which is segmented from a compressed gray image $\mathbf{X}$ (see Section III-B) with a gradually strict gray threshold. Another descriptor $\mathbf{D}_{\text{bu}(w)}$ calculated by using the latter method is also a combinative matrix of multidensity layers but these layers are derived from summing the number of neighborhood members for each pixel with a gradually increased neighborhood size and a stable middle gray threshold. $\mathbf{D}_{\text{bu}(t)}$ and $\mathbf{D}_{\text{bu}(w)}$ are described by (4) and (5), respectively, and Fig. 5(a) and (b) shows two examples applying them:

$$\mathbf{D}_{\text{bu}(t)}(u,v) = \left\{ \sum_{T=T_{\text{mean}}}^{T\max} \text{count}(\mathbf{W}_{\text{bu}(t)}(u,v)), \mathbf{W}_{\text{bu}(t)} | \mathbf{X} > T \right\} \tag{4}$$

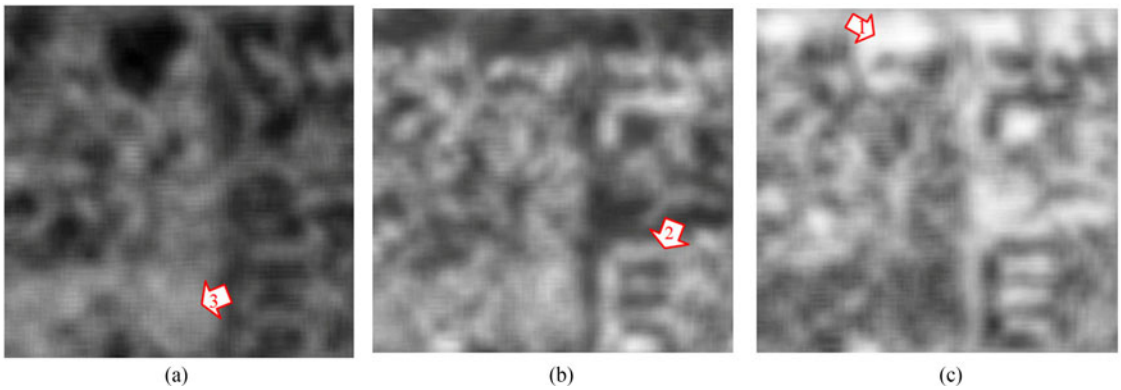

Fig. 4. Three typical weighted combinations of element selected from $\mathbf{D}_{d(\mathrm{all})}$. (a) $\mathbf{D}_{d(s)}$ is a descriptor for extraction of small-size shadow where $\mathbf{D}_{d(s)} = 0.7 \times \mathbf{D}_{d(1,1)} - 0.7 \times \mathbf{D}_{d(1,2)} - 0.7 \times \mathbf{D}_{d(1,3)}$. The highlighting members in it often provide a better agreement with shadows of old-fashioned courtyard dwellings (e.g., the members pointed by arrow 3). (b) $\mathbf{D}_{d(m)}$ is another descriptor for extraction of middle-size shadow where $\mathbf{D}_{d(m)} = 0.7 \times \mathbf{D}_{d(2,1)} + 0.7 \times \mathbf{D}_{d(2,2)} - 0.3 \times \mathbf{D}_{d(3,2)} - 0.7 \times \mathbf{D}_{d(s)}$. The highlighting members in it often agree with shadows of multistorey residential building (pointed by arrow 2). (c) $\mathbf{D}_{d(a)}$ is also a descriptor for extraction of large-size shadow where $\mathbf{D}_{d(a)} = 0.5 \times \mathbf{D}_{d(3,1)} + 0.5 \times \mathbf{D}_{d(3,2)} + 0.5 \times \mathbf{D}_{d(3,3)} - 0.7 \times \mathbf{D}_{d(m)}$. The highlighting members in it often agree with shadows of high-rise building (pointed by arrow 1). The original image is the same as that of Fig. 3(a).

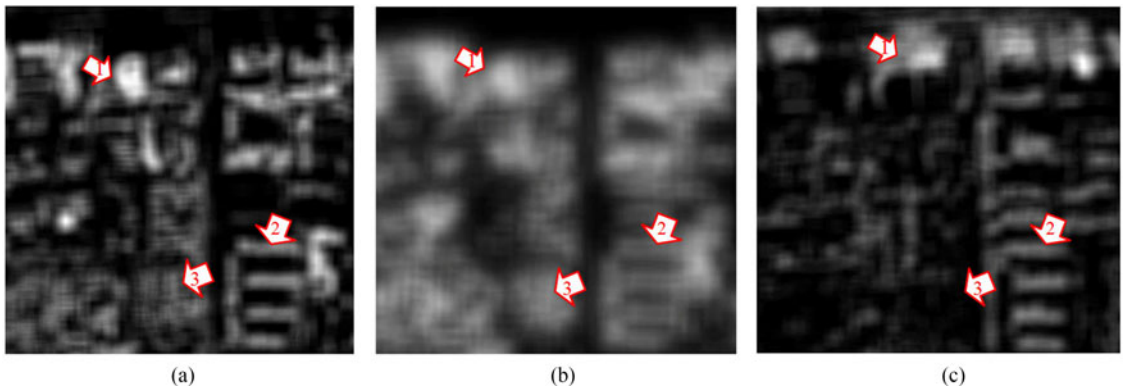

Fig. 5. Examples of forming building or shadow patches through three EMRS-based descriptors. The original image is the same as Fig. 3(a). (a) The brightness from light to dark ($\mathbf{D}_{\mathrm{bu}(t)}$ from high to low) provides a better agreement with building height from tall (pointed by arrow 1) to middle (pointed by arrow 2) and then to low (pointed by arrow 3). (b) Higher values of $\mathbf{D}_{\mathrm{bu}(w)}$ are fairly certain indicators for members of high-rise building. (c) The brightness from light to dark ($\mathbf{D}_{\mathrm{sa}(t)}$ from high to low) provides a better agreement with building height from tall to low and all the shadow patches have been moved to locations of corresponding building. Therefore, a combined use of $\mathbf{D}_{\mathrm{bu}(t)}$, $\mathbf{D}_{\mathrm{bu}(w)}$, and $\mathbf{D}_{\mathrm{sa}(t)}$ is helpful to separate these building types.

where $\mathbf{D}_{\mathrm{bu}(t)}(u, v)$ denotes the *u*th row and *v*th column element of $\mathbf{D}_{\mathrm{bu}(t)}$; $\mathbf{W}_{\mathrm{bu}(t)}$ denotes a binary image of building patch segmented from $\mathbf{X}$. The threshold T increases from T_{mean} (the mean brightness of $\mathbf{X}$) to $T_{\max}$ (the maximum T decided by the increment and the iteration number):

$$\mathbf{D}_{\mathrm{bu}(w)}(u, v) = \left\{ \sum_{w=w_{\mathrm{mean}}}^{w\,\max} \mathrm{count}(\mathbf{W}_{\mathrm{bu(w)}(u,v)})_{w\times w}, \mathbf{W}_{\mathrm{bu(w)}} | \mathbf{X} > T_{\mathrm{mean}} \right\} \tag{5}$$

where $\mathbf{D}_{\mathrm{bu}(w)}(u, v)$ denotes the *u*th row and *v*th column element of $\mathbf{D}_{\mathrm{bu}(w)}$; $\mathbf{W}_{\mathrm{bu}(w)}$ denotes the binary image of building patch segmented from $\mathbf{X}$ with T_{mean}. w (the size of neighborhood for computation of the density) increases from w_{mean} (with a default of 11) to $w_{\max}$ (the maximum w decided by the iteration number). $\mathrm{count}(\cdot)_{w\times w}$ means to count sure members in w-by-w moving window.

Size of building shadow is also a meaningful indicator to building height. By using a method almost the same as that of obtaining $\mathbf{D}_{\mathrm{bu}(t)}$, a new descriptor $\mathbf{D}_{\mathrm{sa}(t)}$ can be generated. $\mathbf{D}_{\mathrm{sa}(t)}$ is a combinative matrix of multidensity layers as described by (6) and each layer is derived from a binary image of shadow member. Experiments indicate that shadow members can be extracted quite completely through the segmentation from an inversing $\mathbf{X}$ matrix with a gradually strict gray threshold:

$$\mathbf{D}_{\mathrm{sa}(t)}(u, v) = \sum_{T_{\mathrm{sa}}=T_{\mathrm{mean}}}^{T\,\max} \mathrm{count}(\mathbf{W}_{\mathrm{sa}(t)}(u, v)), \quad \mathbf{W}_{\mathrm{sa}(t)} | !\mathbf{X} > T_{\mathrm{sa}}. \tag{6}$$

The method of calculating $\mathbf{D}_{\mathrm{sa}(t)}$ is almost the same as that for $\mathbf{D}_{\mathrm{bu}(t)}$ only replacing $\mathbf{X}$ with its inverse (sign ! means inversing) and T with T_{sa}. $\mathbf{W}_{\mathrm{sa}(t)}$ denotes a binary image of shadow member. A shadow patch has additionally been moved to the location of corresponding building by an adaptive computation according to the south–north length of the shadow patch. Fig. 5(c) shows an example of $\mathbf{D}_{\mathrm{sa}(t)}$.

Fig. 6. Examples of EMRS-based standard deviation descriptor. (a) A standard deviation matrix derived from component $\mathbf{V}$ of Fig. 3(a) with a 3-by-3 moving window. (b) A sum matrix derived from (a). (c) One of the weighted combinations of $\mathbf{S}_{\mathrm{sum}(i)}$ for indication of members of low-rise dwelling.

3) Neighborhood Deviation: When an image gradually zooms out on a screen, homogeneity between pixels appears early at the region of low-rise dwelling (old-fashioned courtyard dwelling), later the region of multistorey residential building, and finally the region of high-rise building. The standard deviation of grayscale in a size-given neighborhood of a pixel can serve as an indicator to express the homogeneity. Therefore, the standard deviation from a size gradually increased neighborhood can simulate the zooming-out process. $\mathbf{D}_{\mathrm{MR}}$ is such an EMRS-based descriptor for the simulation.

As setting the initial, increment, and iteration number with defaults of 3, 8, and 5, respectively, the neighborhood sizes are 3-by-3, 11-by-11, 19-by-19, 27-by-27, 35-by-35, and 43-by-43 in turn. Five matrices of standard deviation, $\mathbf{S}_{\mathrm{td}(1)}$, $\mathbf{S}_{\mathrm{td}(2)}$, . . ., and $\mathbf{S}_{\mathrm{td}(5)}$ agreeing with these neighborhood sizes, are derived from a grey scale image. Fig. 6(a) shows an example of $\mathbf{S}_{\mathrm{td}(1)}$ calculated from the brightness component of Fig. 3(a). It can be seen that "higher $\mathbf{S}_{\mathrm{td}(1)}$" is a quite certain indicator to the outlines of high-rise building (pointed by arrow 1). Another meaningful phenomenon is that the regions having dense distribution of middle $\mathbf{S}_{\mathrm{td}(1)}$ elements fairly agree with the members of low-rise dwelling (pointed by arrow 3). In order to express grayscale and distribution in the same time, replace $\mathbf{S}_{\mathrm{td}(i)}$ $(i = 1, 2 \ldots 5)$ with $\mathbf{S}_{\mathrm{sum}(i)}$ to indicate the changes of homogeneity between the building types as

$$\mathbf{S}_{\mathrm{sum}(i)} = \mathbf{D}_{s(i)} + \mathbf{D}_{m(i)} + \mathbf{D}_{a(i)} \; (i = 1, 2 \ldots 5) \tag{7}$$

where $\mathbf{D}_{s(i)}$, $\mathbf{D}_{m(i)}$, and $\mathbf{D}_{a(i)}$ denote the density matrices of qualified element derived with 11-by-11, 22-by-22, and 33-by-33 neighborhoods, respectively, where the qualified elements meet the condition that their neighborhood standard deviations are higher than mean $\mathbf{S}_{\mathrm{td}(i)}$.

Fig. 6(b) shows an example of $\mathbf{S}_{\mathrm{sum}(1)}$. Three EMRS-based descriptors $\mathbf{D}_{\mathrm{MR}(s)}$, $\mathbf{D}_{\mathrm{MR}(m)}$, and $\mathbf{D}_{\mathrm{MR}(a)}$ are derived from the weighted combinations of $\mathbf{S}_{\mathrm{sum}(i)}$ $(i = 1, 2 \ldots 5)$ as

$$\mathbf{D}_{\mathrm{MR}(s)} = 0.7 \times \mathbf{S}_{\mathrm{sum}(1)} - 0.3 \times \mathbf{S}_{\mathrm{sum}(2)} - 0.2 \times \mathbf{S}_{\mathrm{sum}(3)} \tag{8}$$

$$\mathbf{D}_{\mathrm{MR}(m)} = 0.5 \times \mathbf{S}_{\mathrm{sum}(2)} + 0.3 \times \mathbf{S}_{\mathrm{sum}(3)} - 0.5 \times \mathbf{D}_{\mathrm{MR}(s)} \tag{9}$$

$$\mathbf{D}_{\mathrm{MR}(a)} = 0.7 \times \mathbf{S}_{\mathrm{sum}(4)} + 0.7 \times \mathbf{S}_{\mathrm{sum}(5)} - 0.4 \times \mathbf{D}_{\mathrm{MR}(m)} - 0.4 \times \mathbf{D}_{\mathrm{MR}(s)}. \tag{10}$$

Fig. 6(c) shows an example of $\mathbf{D}_{\mathrm{MR}(s)}$. It can be seen that the elements with higher $\mathbf{D}_{\mathrm{MR}(s)}$ quite fairly agree with the members of low-rise dwelling (pointed by arrow 3). Experiments indicate that these constant weights in (8)–(10) need no adjustment as the image changes among these testing images.

D. Descriptors for Building Type Classification

The feature space used here consists of 15 descriptors, $\mathbf{D}_{d(a)}$, $\mathbf{D}_{d(m)}$, $\mathbf{D}_{d(s)}$, $\mathbf{D}_{\mathrm{bu}(t)}$, $\mathbf{D}_{\mathrm{bu}(w)}$, $\mathbf{D}_{\mathrm{sa}(t)}$, $\mathbf{D}_{\mathrm{MR}(a)}$, $\mathbf{D}_{\mathrm{MR}(m)}$, $\mathbf{D}_{\mathrm{MR}(s)}$, $\mathbf{M}_{\mathrm{SV}}$, $\mathbf{M}_{\mathrm{VI}}$, $\mathbf{A}_{\mathrm{bu}}$, $\mathbf{A}_{\mathrm{sa}}$, $\mathbf{S}_{\mathrm{ra}}$, and $\mathbf{D}_{\mathrm{bu}(m)}$. There is evidence that the space is effective in building type classification using HSR imagery solely. A descriptor is a grayscale image and can be regarded as a two-dimensional matrix. The first nine are EMRS-based ones which are introduced in Section III-C. The rest is described as follows:

1) $\mathbf{M}_{\mathrm{SV}}$ denotes the matrix of NDSV median where NDSV is the normalized difference between saturation and brightness values as defined by (11) [27]. NDSV can be used to capture shadow and vegetation members because both of them have higher NDSV values:

$$\mathrm{NDSV} = (\mathbf{S} - \mathbf{V}) \,/\, (\mathbf{S} + \mathbf{V}) \tag{11}$$

where $\mathbf{S}$ and $\mathbf{V}$ denote the saturation and brightness components in HSV color model.

2) $\mathbf{M}_{\mathrm{VI}}$ is the matrix of NDVI median where NDVI is the normalized difference vegetation index [28].
3) $\mathbf{A}_{\mathrm{bu}}$ denotes the matrix of weighted building area as defined by (12). The value of an element in $\mathbf{A}_{\mathrm{bu}}$ is the weighted area of a patch where the element locates. The computation of the weighted area includes two steps: a) marking smaller patches with their average area to save computational time, and b) marking each remaining by its weighted area to avoid possible mistakes from unexpected

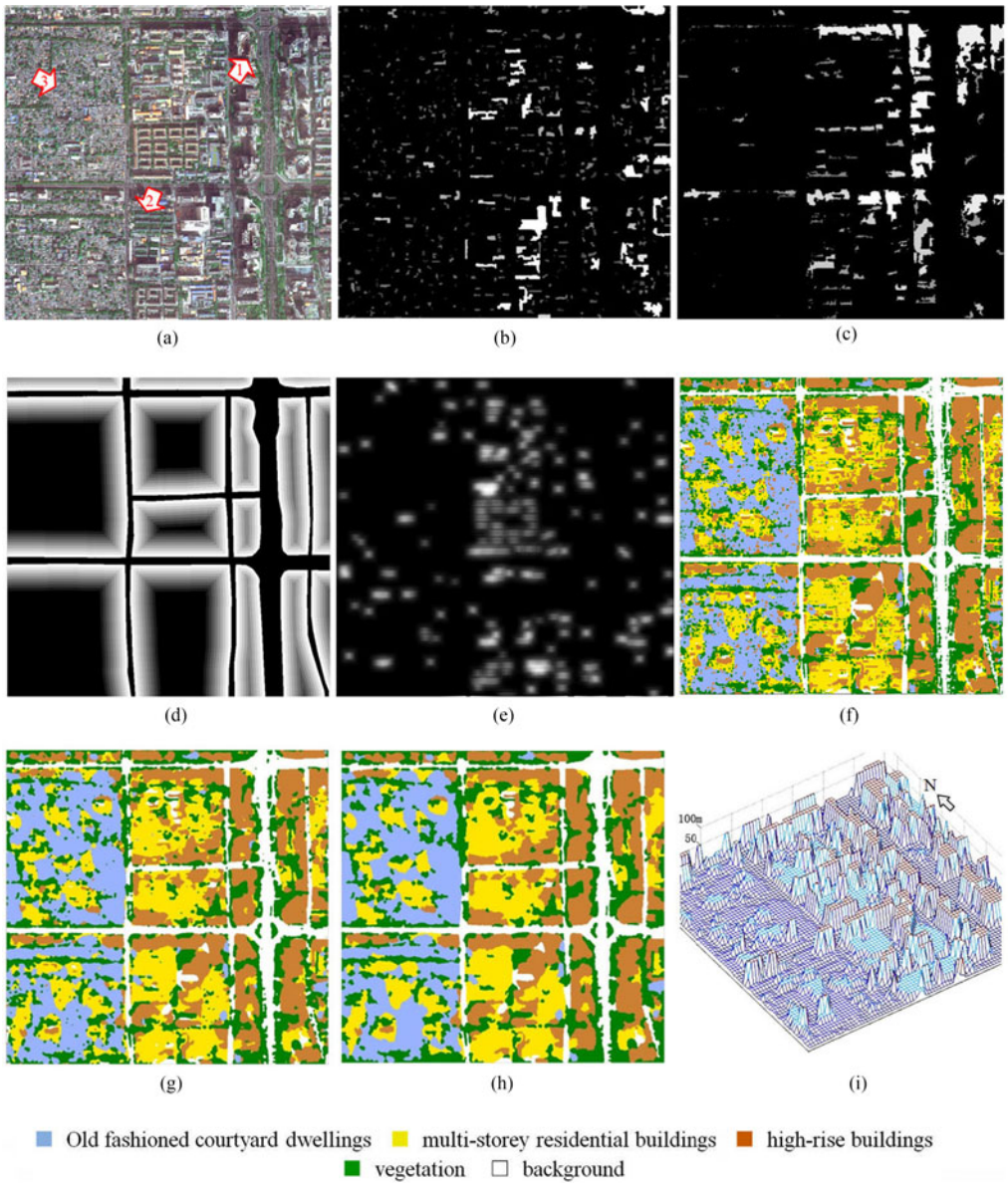

Fig. 7. SRS descriptors and comparison between HBP and SBP classification. (a) Original Worldview 3 image. (b) $\mathbf{A}_{\mathrm{bu}}$. (c) $\mathbf{A}_{\mathrm{sa}}$. (d) $\mathbf{S}_{\mathrm{ra}}$. (e) $\mathbf{D}_{\mathrm{mb}}$. (f) Output by HBP. (g) Output by SBP. (h) Final results simplified from (g). (i) 3D view.

adhesion between patches:

$$\mathbf{A}_{\mathrm{bu}}(u,v) = \begin{cases} 0.07 \times A_{\mathrm{mean}} & (A_{(i)} < 0.1 \times A_{\mathrm{mean}}) \\ 0.18 \times A_{\mathrm{mean}} & (0.1 \times A_{\mathrm{mean}} \le A_{(i)} < 0.25 \times A_{\mathrm{mean}}) \\ c_a \times \mathrm{abs}(1 - P_{a(i)}) \times A_{(i)} & (A_{(i)} \ge 0.25 \times A_{\mathrm{mean}}) \end{cases} \tag{12}$$

where $\mathbf{A}_{\mathrm{bu}}(u,v)$ denotes the *u*th row and *v*th column element of $\mathbf{A}_{\mathrm{bu}}$ and locates in patch *i*. c_a is a factor for area adjustment. $P_{a(i)}$ denotes the weight for area adjustment of patch *i* where $P_{a(i)} = a \times \exp(b \times C_{\mathrm{cox}}(u,v))$ where *a* and *b* are two experimental coefficients with defaults of 0.06 and 2.77, respectively. The defaults are derived from real experimental data through the least square fitting and have proved to be good in universality. $C_{\mathrm{cox}}(i)$ is the shape complex coefficient of patch *i* serving as an indicator to assess the adhesion between patches. $C_{\mathrm{cox}(i)} = L_{\mathrm{adj}(i)}/A_i$ where A_i is the area of patch i; $L_{\mathrm{adj}(i)}$ is the corrected perimeter, and $L_{\mathrm{adj}(i)} = L_{(i)}/(1-d)$ where $L_{(i)}$ is the real perimeter where $d = (l-h)/(l+h)$ where *l* and *h* are the patch length and width. Fig. 7(b) shows an example of $\mathbf{A}_{\mathrm{bu}}$ where higher brightness indicates larger weighted area.

4) $\mathbf{A}_{\mathrm{sa}}$ denotes the matrix of weighted building shadow area. The computation method for it is almost the same as that for $\mathbf{A}_{\mathrm{bu}}$. Fig. 7(c) shows an example of $\mathbf{A}_{\mathrm{sa}}$ where higher brightness indicates larger weighted area.
5) $\mathbf{S}_{\mathrm{ra}}$ denotes a matrix of road buffer region. The value of an element in $\mathbf{S}_{\mathrm{ra}}$ is the distance between a road and the element. $\mathbf{S}_{\mathrm{ra}}$ is derived from $\mathbf{W}_{\mathrm{ra}}$ through morphological dilating with a size gradually increased SE where $\mathbf{W}_{\mathrm{ra}}$ is a binary image of road derived from a GIS. The dilating

process is as follows:

$$\mathbf{S}_{\mathrm{ra}}\left(\mathbf{W}_k\right) = 1 - 0.1 \times k \left(k = 1, 2, \ldots 9\right) \tag{13}$$

where $\mathbf{S}_{\mathrm{ra}}\left(\mathbf{W}_k\right)$ means to select $\mathbf{W}'_k$ s true members from $\mathbf{S}_{\mathrm{ra}}$; $\mathbf{W}_k$ is the binary image of newly extended region in loop k, and $\mathbf{W}_k = \mathbf{W}_{\mathrm{ra}} \oplus \mathrm{s}_k \cap ! \mathbf{W}_{[1,k-1]} \cap ! \mathbf{W}_{\mathrm{ra}}$ where sign $\oplus$ represents morphological dilating and $\mathrm{s_k}$ is the SE at loop k.

Fig. 7(d) shows an example of $\mathbf{S}_{\mathrm{sa}}$ where higher brightness represents shorter distance to road.

6) $\mathbf{D}_{\mathrm{bu}(m)}$ denotes the density of multistorey residential building which is designed to separate this type from the others. There are two characters making the former different: 1) each building in this type being far narrower in south–north direction, and 2) several patches being densely arranged as similar to each other in terms of size and shape. Such narrow patches can be extracted by the operation of morphological opening as defined by

$$\mathbf{W}_{\mathrm{bu}(m)} = (\mathbf{W}_{\mathrm{bu}} \circ \mathrm{s}_s) \cap ! (\mathbf{W}_{\mathrm{bu}} \circ \mathrm{s}_m) \tag{14}$$

where $\mathbf{W}_{\mathrm{bu}}$ and $\mathbf{W}_{\mathrm{bu}(m)}$ denote the building binary images for all buildings and for only multistorey residential buildings separately. Sign °is morphological opening while s_s and s_m denote small and middle size SEs to filter out noise and the multistorey residential building, respectively.

$\mathbf{D}_{\mathrm{bu}(m)}$ can be computed by (15). Fig. 7(e) shows an example of $\mathbf{D}_{\mathrm{bu}(m)}$ where higher brightness indicates higher density of multistorey residential buildings:

$$\mathbf{D}_{\mathrm{bu}(m)}(u, v) = \sum_{w=w_S}^{w_L} \mathrm{count}(\mathbf{W}_{\mathrm{bu}(m)}(u, v))_{w \times w} \tag{15}$$

where w_L and w_S denote the minimum and maximum of w.

E. Soft Classification With BP Network

The building type classification is still hard work due to poor homogeneity containing a type although the classification conducts in an EMRS space. To solve this problem, a soft classification approach, referred to here as SBP, is also addressed. SBP includes the technologies of soft partition and defuzzifying.

The soft partition is reached by setting output mode with continuous variable to adapt to the gradual transformation between membership and nonmembership. This enables to generate a natural prototype of fuzzy clusters. Output vector $\mathbf{A}$ is a m-by-n matrix as defined by (16). $\mathbf{A}_{\mathrm{idea}}$ is an idea $\mathbf{A}$ as defined by (17). Each column of $\mathbf{A}$ is the mode value vector of a pixel. Of all components of a column, the higher the mode value of a component, the higher the probability that the pixel belongs to the class indicated by the row number of the component. In traditional hard classification with BP network (HBP), the row number of the largest component in a column serves as the class mark. However, the reliability of the result is often challenged by reality:

Columns 1 through 7

0.0117	0.0042	0.0032	0.0245	0.0149	0.0327	0.0053
0.0085	0.0633	0.0583	0.4046	0.0751	0.5695	0.5754
0.9948	0.9585	0.0362	0.0059	0.0684	0.0035	0.0063
0.1079	0.0305	0.9840	0.7562	0.9655	0.6336	0.7962
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fig. 8. Example of output matrix $\mathbf{A}$.

$$\mathbf{A} = \begin{vmatrix} a_{1,1} & .. & a_{1,j} & .. & a_{1,n} \\ .. & & .. & & .. \\ a_{i,1} & .. & a_{i,j} & .. & a_{i,n} \\ .. & & .. & & .. \\ a_{m,1} & .. & a_{m,j} & .. & a_{m,n} \end{vmatrix} \text{ Referring to } |a_{i,j}|_{\substack{i=1:m \\ j=1:n}} \tag{16}$$

$$\mathbf{A}_{\mathrm{idea}} = \begin{bmatrix} 1 & 0 & 0 & .. & 0 \\ 0 & 1 & 0 & .. & 0 \\ : & : & : & .. & : \\ 0 & 0 & 0 & .. & 1 \end{bmatrix} \tag{17}$$

Fig. 8 shows several columns of a real $\mathbf{A}$ for the first seven pixels. It can be seen that the third component (the highlighted) of the first column (for the first pixel) is the largest and very close to 1; therefore, the pixel can be safely marked with 3. However, to the sixth pixel, the largest and sublarge components are 0.6336 and 0.5696; that is, the probabilities that the pixel belongs to class 2 and class 4 are very close. Similar conclusions can also be derived from other five pixels. In many cases, it is unsafe to determine class marks according to the maximum membership solely.

In an entire process of SBP classification, an original $\mathbf{A}$ serves as a prototype of natural cluster and then unsure members in $\mathbf{A}$ are relabeled through a systematic defuzzifying. Steps for the defuzzifying are as follows:

1) Add specified weights to each element in $\mathbf{A}$ to enlarge the differences between components of a column. If a membership is higher than a given threshold T_A, set the weight with 2, else if less than half of T_A, set with 0.5 and else set with 1 (not processing). For example, when defining $T_A = 0.9$ the renewed membership vector of column 1 in Fig. 8 should be [0.085, 0.00425, 1.9896, 0.05395, 0.0000]$^{\mathrm{T}}$. Denote the renewed $\mathbf{A}$ as $\mathbf{U}$ as defined by (18) which is the weighted membership matrix.
2) The initial class label matrix $\mathbf{L}_0$ is derived from $\mathbf{U}$ according to the rule of maximum membership. $\mathbf{L}_0$ is a 1-by-n matrix as defined by (19). Each element in $\mathbf{L}_0$ is an initial class mark for a pixel.
3) Then $\mathbf{L}_0$ will be converted into $\mathbf{L}_1$ as defined by (20) which is a r-by-c matrix where r and c are the row and column numbers of the processing image where n equals r-by-c.
4) By relabeling an element with the dominant label in its neighborhood, generate a new r-by-c label matrix $\mathbf{L}_2$ as defined by (21). The so-called dominant label represents

the class having the highest proportion of elements in the neighborhood. If the proportion of class *i* is the highest and higher than T_p, the pixel is relabeled with *i*, otherwise with zero. A zero-labeled member (fuzzy member) can be relabeled later according to the labels and density of the sure members around this fuzzy.

5) Within a loop of the recursive defuzzifying, parts of the fuzzy members adjacent to sure ones are relabeled first and then these serve as new sure members to help relabel the rest fuzzy. The recursion does not stop until all the fuzzy are relabeled. Denote the updated matrix as $\mathbf{L}_3$ as defined by (22).
6) Find out noise members from $\mathbf{L}_3$, relabel them with zero, and then recursively relabel them with the same process at the last step. The relabeling process can be described by (24) and $\mathbf{L}_{\text{final}}$ is the final qualified label matrix:

$$\mathbf{U} = |u_{i,j}|_{\substack{i=1:m \\ j=1:n}} \text{ and } u_{i,j} = \begin{cases} 2 \times a_{i,j} & (a_{i,j} > T_{\text{A}}) \\ a_{i,j} & (0.5 \times T_{\text{A}} \le a_{i,j} \le T_{\text{A}}) \\ 0.5 \times a_{i,j} & (a_{i,j} < 0.5 \times T_{\text{A}}) \end{cases} \tag{18}$$

$$\mathbf{L}_0 = |l_j|_{j=1:n} \text{ and } l_j = i, i|u_{i,j} = \max(|u_{i,j}|_{\substack{i=1:m \\ j=j}}). \tag{19}$$

Equation (19) states that mark pixel *j* with *i* if component *i* in column *j* of **U** is the largest where max(•) is the maximum function:

$$\mathbf{L}_1 = |l_{u,v}|_{\substack{u=1:r \\ v=1:c}} \quad (u = \text{floor}(j/c); v = j - u \times c) \tag{20}$$

where floor(•) is the round down function:

$$\mathbf{L}_2 = |l_{u,v}|_{\substack{u=1:r \\ v=1:c}} \text{ and } l_{u,v} = \begin{cases} i, & i|A_i = \max(A_1, A_2 ... A_m) \,\&\, A_i \ge T_{\text{area}} \\ 0, & i| \quad (\max(A_i) < T_{\text{area}} \end{cases} \tag{21}$$

where A_i denotes the weighted density of *i*-labeled element in the *w*-by-*w* neighborhood of pixel (*u,v*). T_{area} denotes an area threshold with a default of $0.3 \times w^2$:

$$\mathbf{L}_3 = \{M_{\text{sure}}^{(1)}, ... M_{\text{sure}}^{(i)}, ... M_{\text{sure}}^{(m)}\} \tag{22}$$

and

$$M_{\text{sure}}^{(i)} = \{M_{\text{sure}}^{(k-1,i)} + \sum_{k=1}^{q} (M_{\text{sure}}^{(k,i)} \oplus \text{s}) \cap \forall M_0^{(k)}, \; M_0^{(k)} | \text{area}(\forall M_0) < k \times N\}. \tag{23}$$

Equation (22) states that $\mathbf{L}_3$ consists of all ensured members belonging to *m* classes. Equation (23) shows the recursive process of updating $M_{\text{sure}}^{(i)}$ which is the subset of ensured elements of class *i*. *k* is the loop variable ($k = 1,2 \ldots q$); $M_{\text{sure}}^{(k,i)}$ is $M_{\text{sure}}^{(i)}$ at loop *k* (especially, $M_{\text{sure}}^{(k-1,i)}$ equates the original $M_{\text{sure}}^{(i)}$ when $k = 1$). **s** denotes a size-given SE. Area(•) is the patch area function. $M_0^{(k)}$ denotes the set of zero-labeled member at loop *k* while the limitation on it means that the area of $M_0^{(k)}$ should not larger than *k*-by-*N* where *N* is the base number of area threshold. The limitation promises zero-labeled elements being evenly divided into all adjacent classes, rather than all assigned to the class encountered first. At loop *k*, the members in renewed $M_{\text{sure}}^{(k,i)}$ serve as the sure members of class *i* to absorb remaining zero-labeled members. An iterative process of traversing *m* classes is embedded in loop *k*:

$$\mathbf{L}_{\text{final}} = \left\{ \sum_{k=1}^{q} \sum_{i=1}^{n} \left[\mathbf{L}_3^{(k)} (\mathbf{W}_0^{(k-1)} \cap \mathbf{W}_{\text{uni}}^{(k-1)}) = i, \mathbf{W}_0^{(k)} (\mathbf{W}_0^{(k-1)} \cap \mathbf{W}_{\text{uni}}^{(k-1)}) = 0 \right] \right\}$$
$$\text{and } \mathbf{W}_{\text{uni}} = \{(\mathbf{W}_i \bullet \mathbf{s}_2) \cup (\mathbf{W}_i \oplus \mathbf{s}_3)\} \tag{24}$$

where the *i*th loop (the interior loop) traverses each class and conducts a recursive absorption at loop *k* (the external loop). $\mathbf{W}_0$ and $\mathbf{W}_i$ denote the binary images of the zero-labeled and the *i*-labeled members, respectively. $\mathbf{L}_3^{(k)}$ denotes the *k*th renewed $\mathbf{L}_3$.$\mathbf{W}_0^{(k)}$ and $\mathbf{W}_0^{(k-1)}$ denote $\mathbf{W}_0$ at loop *k* and just before loop *k*, respectively. $\mathbf{W}_{\text{uni}}$ denotes the union of two sets derived from adjacent analyses. The former ($\mathbf{W}_i \bullet \mathbf{s}_2$) means morphologically closing $\mathbf{W}_i$ with $\mathbf{s}_2$ where $\mathbf{s}_2$ is a size-given SE. This operation is for a deleting patch surrounded by a far larger patch in $\mathbf{W}_i$ to be absorbed by the larger. The latter ($\mathbf{W}_i \oplus \mathbf{s}_3$) means to dilate $\mathbf{W}_i$ with $\mathbf{s}_3$ where $\mathbf{s}_3$ is another size-given SE. The expression, $\mathbf{L}_3^{(k)}(\bullet) = i$, means to relabel some of members in $\mathbf{L}_3^{(k)}$ with *i* if these members conform to the conditions in the parentheses. Similar to this, the expression, $\mathbf{W}_0^{(k)}(\bullet) = 0$, means to remove these relabeled members from $\mathbf{W}_0$.

Find more information about the process of recursive defuzzifying from another literature provided by our team recently [29]. The literature introduced a defuzzifying method similar to that of this study but the soft partition before the defuzzifying was conducted by combined binary support vector machines instead of SBP.

In the examples shown in Figs. 7 and 9, graphs presented in Figs. 7(f) and 9(d) show the classification results using HBP. Graph presented in Fig. 7(g) displays the results using SBP where an originally unsure pixel is relabeled with the dominant label in the pixel's neighborhood. Graphs presented in Figs. 7(h) and 9(e) show $\mathbf{L}_{\text{final}}$ derived from $\mathbf{L}_3$ after removing noises. It can be seen that SBP performs better than HBP in terms of resisting noise and generating more complete patches. Graphs presented in Figs. 7(i) and 9(f) show 3-D views from $\mathbf{L}_{\text{final}}$ by setting the old-fashioned courtyard dwellings, multistorey residential buildings, and high-rise building with their mean heights of 4, 18, and 60, respectively.

IV. Results and Discussion

The significance of the study (see Sections IV-A and IV-B), accuracy assessments (see Section IV-C), computation com-

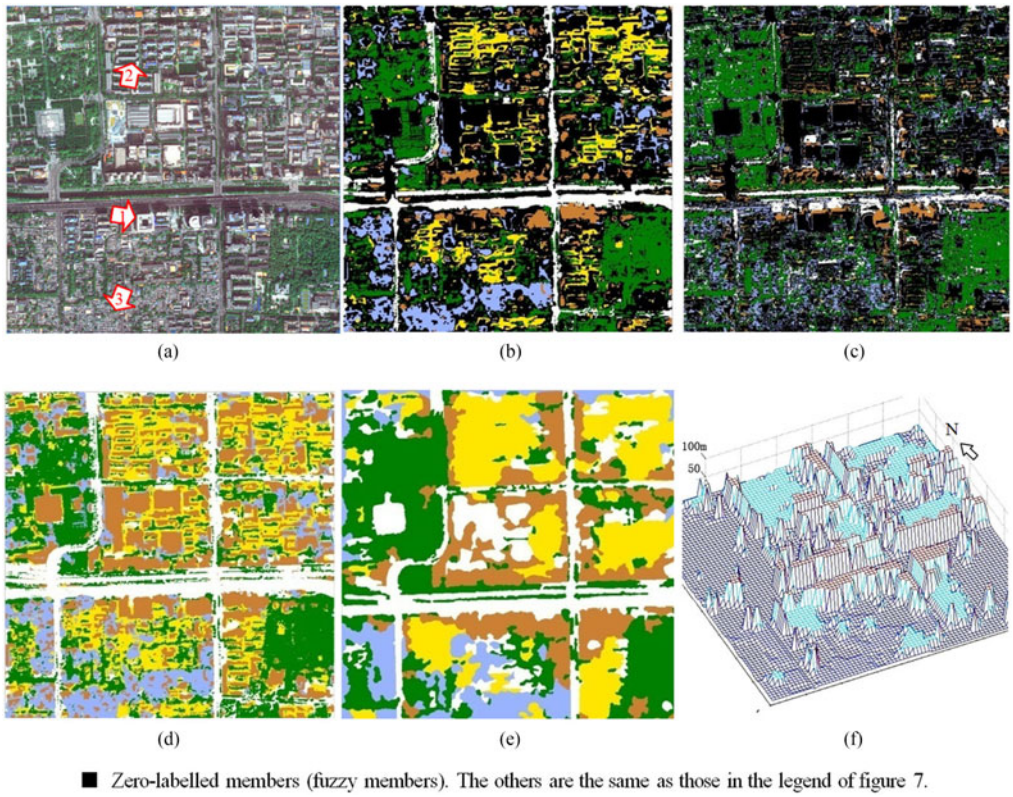

Fig. 9. Comparison of performance in regard to using soft and hard classifiers in single-resolution and multiresolution feature spaces. (a) Original Worldview 3 image. (b) $\mathbf{L}_1$ (classified in EMRS space). (c) $\mathbf{L}_1$ (classified in SRS space). (d) Classification results using EMRS and HBP. (e) Classification results using EMRS and SBP. (f) 3D view.

plexity (see Section IV-D), and sensitivity for parameter variations (see Section IV-E) will be discussed as follows.

A. *Performance of EMRS Against SRS*

When replacing single resolution descriptor with its EMRS-based one, there is a significant increase of carrying information. Take $\mathbf{D}_{\mathrm{sa}(t)}$, a shadow involved descriptor as defined by (6) and displayed by Fig. 5(c), as an example. The value of an element in $\mathbf{D}_{\mathrm{sa}(t)}$ is the weighted sum of percentages of shadow member calculated in several size-given neighborhood of the element. Thereby, the higher the value, the higher the probability the element belongs to a larger area of shadow patch and therefore, the higher the probability the element indicates a higher building. In general, an EMRS-based descriptor is able to express the mixed patterns of spectrum and texture features associated with various resolutions. Some classes could not be distinguished in an SRS space which consists of single resolution descriptors only but now possible to be distinguished in an EMRS space which replaces some of the SRS-characterized descriptors with the EMRS-based ones. Graphs presented in Fig. 9(b) and (c) show a comparison of classification in both of the spaces. It can be seen that in the initial stage of soft partition, the proportion of the black members (fuzzy members) as classified in the SRS space is far larger than that in the EMRS space. The sure building members softly partitioned in the SRS space are so few and so fragmented that they can hardly be connected into building patches. In contrast to this, most patches of the three building types are preliminarily formed as the partition running in the EMRS space. This benefits from these EMRS-based descriptors with which the majority of members of a building type scattered in several layers associated with various resolutions can be captured separately and then integrated into the patches. Consequently, classification accuracy can significantly be improved as classification running in an EMRS space (see the data in Table III).

B. *Significance of the Recursive Defuzzifying in SBP*

The membership and nonmembership between separate building types often converse gradually rather than flatly due to changeable building structure and complex scene even the classification running in an EMRS space. Consequently, there are considerable fuzzy members still remaining in the initial output of classification.

SBP is developed to solve this problem. By setting the output mode with continuous values and taking logistic curve as the activation function, all building types can more naturally be separated. The fuzzy members in the output can be relabeled by recursive defuzzifying. Take Fig. 9 as an example. Using HBP, three building types are heavily confused and many of the original fuzzy members are flatly and mistakenly classified for high-rise buildings [the brown in Fig. 9(d)]. But using SBP, these fuzzy members are marked with zero [the black in Fig. 9(b)] at first, and then more reasonable labels for the black will gradually be clear depending on the information carried by the ensured

members around the black as the recursive defuzzifying going. Fig. 9(f) shows final results after the defuzzifying and denoizing. Comparisons of overall accuracy (OA) using SBP against HBP from four testing images are listed in Table III.

C. Accuracy Assessment

This section provides accuracy assessments using following approaches for comparison:

1) supervised classification in either an EMRS space or an SRS space.
2) supervised classification by either SBP or HBP.

We do not compare EMRS with MRS because the latter is a special case of the former, namely the case of the number of SP(s) being equal to one. By using the latter, a member set for a class is involved with a specified SP but by using the former the set is involved with a group of SP(s) and these SP(s) can adaptively be decided.

The accuracy for an approach is assessed by checking if there is a reasonable spatial agreement between the classification outputs and the visual interpretation results. In order to ensure if the variation of classification accuracy was caused only by the variation of reference item, the number of descriptors for each compared feature space is the same. The EMRS space consists of $\mathbf{D}_{d(a)}$, $\mathbf{D}_{d(m)}$, $\mathbf{D}_{d(s)}$, $\mathbf{D}_{\mathrm{bu}(t)}$, $\mathbf{D}_{\mathrm{bu}(w)}$, $\mathbf{D}_{\mathrm{sa}(t)}$, $\mathbf{D}_{\mathrm{MR}(a)}$, $\mathbf{D}_{\mathrm{MR}(m)}$, $\mathbf{D}_{\mathrm{MR}(s)}$, $\mathbf{M}_{\mathrm{SV}}$, $\mathbf{M}_{\mathrm{VI}}$, $\mathbf{A}_{\mathrm{bu}}$, $\mathbf{A}_{\mathrm{sa}}$, $\mathbf{S}_{\mathrm{ra}}$, and $\mathbf{D}_{\mathrm{bu}(m)}$ (see Sections III-B and III-C) total fifteen descriptors while the SRS space replaces the first nine EMRS-based descriptors with their single components. For example, replace $\mathbf{D}_{d(a)}$ with $\mathbf{D}_{d(3,1)}$ as defined by (2). The samples for training and testing were acquired through visual interpretation and image point sampling. For each test image, the average number of samples for each class is 57, half for the training and the other for the testing. Both of the SBP and HBP networks use single hidden layer with 32 nodes.

The measures for accuracy assessment are the OA and the mean kappa statistic value (κ). Figs. 7, 9, and 10 display four examples for the assessments. Tables I and II provide two confusion matrices for two comparison tests both by SBP but one in EMRS space and the other in SRS space. Table III provides assessing statements for all four examples.

The data in Table III indicate that the mean OA by HBP in EMRS space is 17.2% higher than that in SRS space and the mean κ is 21.8% higher. As running both in EMRS space, the mean OA by SBP is 2.6% higher than that by HBP and the mean κ is 3.3% higher. That is, the mean OA by SBP in EMRS space is 19.8% higher than that by the traditional HBP classifier in the common SRS space and the mean κ is 25.1% higher.

D. Computational Complexity

Computational complexity is an important metric which can be used to evaluate the practicability of a new algorithm. By comparing the computation of descriptor with EMRS against SRS and comparing the classification with SBP against HBP, it can be seen that there is a huge number of extra neighborhood operations being added to EMRS and SBP. The extra addings are used to obtain the multilayer densities of neighborhood elements which are involved with certain conditions and/or resolutions.

We replace the computation for density of gray elements with a calculation for mean value of binary elements to reduce time consumption. The binary matrix for the latter is segmented from a gray matrix according to given conditions. The mean value of binary elements in a neighborhood is approximately equal to the density of gray elements in the same neighborhood under the same segmenting conditions. Big O notation can serve as a measure to assess how the new algorithm responds to changes in input size. By taking function $f(x)$ to denote the computation for the mean value of binary elements, the exponent of x is 1. That is, $f(x) = \mathrm{O}(x^1)$. $f(x)$ should increase linearly as x increases. Experiments also prove that the execution time is steadily and linearly growing as the input size increases. Fig. 11 shows an example of the relation between input size A (image area) and execution time t and the testing image is the same as that shown in Fig. 10(d).

Meanwhile, a comparison of time consumption with the study algorithm against conventional algorithm is listed in Table IV. Four images for the comparison are the same as those for accuracy assessment. The data in the table indicate that the average consuming time for descriptor computation is 11.140 and 8.789 s using EMRS and SRS, respectively, and the former is 22.769% higher than the latter; the average consuming time for classification is 14.628 and 14.160 s using SBP and HBP, respectively, and the former is only 3.330% higher than the latter. That is,

TABLE I
CONFUSION MATRIX FOR THE EXAMPLE IN FIG. 7(A) CLASSIFIED BY SBP IN EMRS SPACE

	$\mathrm{B}_{\mathrm{high}}$	$\mathrm{B}_{\mathrm{mid}}$	$\mathrm{B}_{\mathrm{low}}$	Ve	Bg	Σ_{row}	UA
$\mathrm{B}_{\mathrm{high}}$	58	8	0	1	1	68	0.935
$\mathrm{B}_{\mathrm{mid}}$	3	49	0	0	0	52	0.817
$\mathrm{B}_{\mathrm{low}}$	0	2	32	1	1	36	1
Ve	1	0	0	42	1	44	0.955
Bg	0	1	0	0	69	70	0.958
Σ_{col}	62	60	32	44	72	270	
$\kappa = 0.906$					OA = 0.926		

$\mathrm{B}_{\mathrm{high}}$, $\mathrm{B}_{\mathrm{mid}}$, and $\mathrm{B}_{\mathrm{low}}$ represent high-rise buildings, multistorey residential buildings, and old-fashioned courtyard dwellings, respectively. Ve and Bg represent vegetation and background, respectively. UA denotes user accuracy. Σ_{row} and Σ_{col} represent sums of row and column value, respectively.

TABLE II
CONFUSION MATRIX FOR THE EXAMPLE IN FIG. 7(A) CLASSIFIED BY SBP IN SRS SPACE

	$\mathrm{B}_{\mathrm{high}}$	$\mathrm{B}_{\mathrm{mid}}$	$\mathrm{B}_{\mathrm{low}}$	Ve	Bg	Σ_{row}	UA
$\mathrm{B}_{\mathrm{high}}$	59	17	1	0	3	80	0.952
$\mathrm{B}_{\mathrm{mid}}$	2	30	0	0	1	33	0.5
$\mathrm{B}_{\mathrm{low}}$	0	2	29	1	0	32	0.906
Ve	1	1	2	42	2	48	0.955
Bg	0	10	0	1	66	77	0.917
Σ_{col}	62	60	32	44	72	270	
$\kappa = 0.793$					OA = 0.837		

TABLE III
SUMMARY OF ACCURACY ASSESSMENT FOR ALL TEST IMAGES

No	Image size	N_{Fig}	In EMRS space						In SRS space					
			By HBP		By SBP		By SBP after simplification		By HBP		By SBP		By SBP after simplification	
			OA	κ	OA	κ	OA	κ	OA	κ	OA	κ	OA	κ
1	757 × 800	7(a)	0.882	0.840	0.922	0.887	0.926	0.906	0.811	0.76	0.855	0.817	0.837	0.793
2	797 × 800	9(a)	0.913	0.891	0.927	0.908	0.927	0.908	0.648	0.558	0.700	0.625	0.732	0.663
3	718 × 711	10(a)	0.888	0.856	0.898	0.869	0.905	0.877	0.675	0.573	0.747	0.670	0.770	0.701
4	702 × 739	10(d)	0.848	0.806	0.889	0.859	0.903	0.876	0.709	0.629	0.799	0.743	0.789	0.729
	Mean		0.883	0.848	0.909	0.881	0.915	0.892	0.711	0.630	0.775	0.714	0.782	0.722

No and N_{Fig} denote the serial number of test and the figure number of test image, respectively.

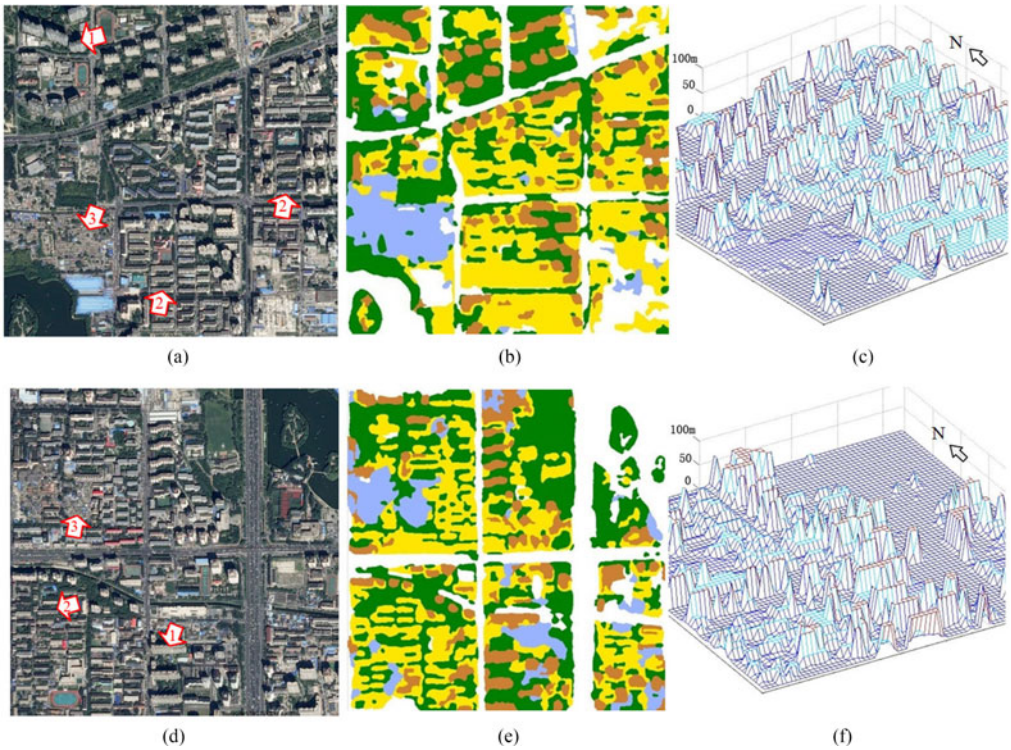

Fig. 10. Third and fourth examples for accuracy assessments. (a) Original Quickbird image. (b) Final results. (c) 3-D view. (d) Original Quickbird image. (e) Final results. (f) 3-D view.

the consuming time does not significantly increase as the extra neighborhood operations being added to the new algorithms.

E. Sensitivity for Parameter Variation

In order to automatically set parameters in the proposed scheme, a parameter is usually divided into two parts: a base number and two experimental variables. The base number is often an average and can be derived adaptively from image data. Two variables are the increment and the iteration number and need to be set experimentally. However, the experimental setting is not difficult because a 10–15% deviation is verified acceptable. Take $\mathbf{D}_d$ as defined by (1) as an example. In order to calculate $\mathbf{D}_d$, a threshold series needs to be specified for the multisegmentation of a gray matrix derived from a low hat transform. The series is set with c by d_{mean} where d_{mean} is the mean

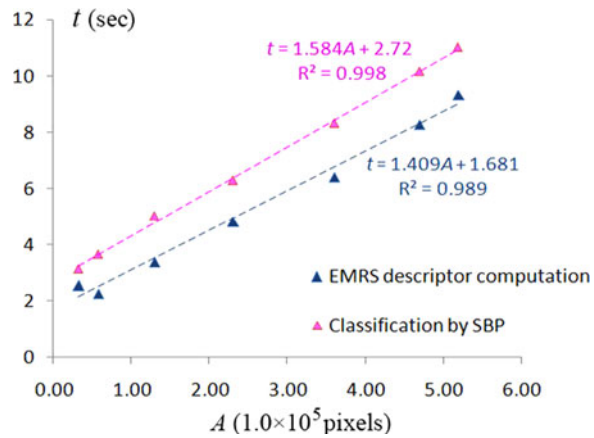

Fig. 11. Relation between input size A (image area) and execution time t.

TABLE IV
ASSESSMENT OF TIME CONSUMPTION

No	N_{Fig}	Descriptor computation			Classification by BP		
		t_{EMRS} (s)	t_{SRS} (s)	$\mathrm{d}t_1$ (%)	t_{SBP} (s)	t_{HBP} (s)	$\mathrm{d}t_2$ (%)
1	7(a)	10.370	7.978	29.991	16.772	16.289	2.965
2	9(a)	13.201	10.584	19.824	16.360	15.830	3.348
3	10(a)	11.341	8.571	24.425	12.719	12.282	3.558
4	10(d)	9.646	8.022	16.836	12.662	12.240	3.448
Mean		11.140	8.789	22.769	14.628	14.160	3.330

value of the gray matrix. d_{mean} also serves as a base number (a central threshold) to avoid the series deviating from actual situations seriously. On the other hand, c changes in a range around 1. In this example, set the increment and the iteration number with 0.3 and 3, respectively. The range of c is $\{0.7,1,1.3\}$. Experiments indicate that the c range is often stable and needs no adjustment for all four testing images resulting from taking d_{mean} as an approximate center value of the threshold series. That is, classification accuracy seems almost insensitive to a not too big deviation in setting c.

V. CONCLUSION

A literature review of current practices in building type classification indicates that this study has several meaningful innovations:

1) Propose the concept of EMRS. EMRS uses the multistructural element morphological operations, multisized neighborhood statistics, and multithreshold segmentation to get the information scattered in different layers (e.g., the layer of pixel, neighborhood, image object, etc.) and to generate meaningful segments.
2) Integrate the segments into a number of weighted sum matrices which serve as either a descriptor or a component of descriptor in a feature space, therefore, enabling EMRS to be embedded in the classification and enabling the members of a single building type as scattered in several layers of different resolutions to be captured entirely.
3) Develop a defuzzifying method to recursively relabel fuzzy members in a clustering prototype derived from soft partition depending on the information carried by the gradually increased sure members around the fuzzy.
4) The extra neighborhood operations almost do not increase the computational consumption due to having the mean value of binary element as a replacement for the density of gray element.
5) Classification accuracy is usually not sensitive to the deviation in setting parameters because the center values of these parameters are adaptively derived from image data.

The main contribution of this study is the EMRS-SBP scheme explored for building type classification in complex urban scene using general HSR imagery solely. The excellent classification accuracy and good universality indicate that the scheme, in many cases, has potential for identification of other objects, especially for those with spectral confusion and in complex scene.

REFERENCES

[1] Z. Y. Lu, J. Im, L. J. Quackenbush, and K. Halligan, "Population estimation based on multi-sensor data fusion," *Int. J. Remote Sens.*, vol. 31, no. 21, pp. 5587–5604, Nov. 2010.

[2] A. Troy and J. M. Grove, "Property values, parks, and crime: A hedonic analysis in Baltimore, MD," *Lands Urban Plan.*, vol. 87, no. 3, pp. 233–245, Aug. 2008.

[3] S. S. Wu, L. Wang, and X. Qiu, "Incorporating GIS building data and census housing statistics for sub-block population estimation," *Prof. Geographer*, vol. 60, no. 1, pp. 121–135, Feb. 2008.

[4] K. Khoshelham, C. Nardinocchi, E. Frontoni, A. Mancini, and P. Zingaretti, "Performance evaluation of automated approaches to building detection in multi-source aerial data," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 65, no. 2010, pp. 123–133, Jan. 2010.

[5] F. Rottensteiner *et al.*, "The ISPRS benchmark on urban object classification and 3D building reconstruction," in *Proc. ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Aug. 2012, pp. 293–298.

[6] A. Huertas and R. Nevatia, "Detecting buildings in aerial images," *Comput. Vis. Graph. Image Process.*, vol. 41, no. 2, pp. 131–152, Feb. 1988.

[7] C. Lin and R. Nevatia, "Building detection and description from a single intensity image," *Comput. Vis. Image Understanding*, vol. 72, no. 2, pp. 101–121, Nov. 1998.

[8] A. O. Ok, "Automated detection of buildings from single VHR multi spectral images using shadow information and graph cuts," *ISPRS Int. J. Photogram. Remote Sens.*, vol. 86, no. 12, pp. 21–40, Sep. 2013.

[9] S. W. Myint, P. Gober, A. Brazela, S. Grossman-Clarkeb, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, May 2011.

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

[11] M. Bouziani, K. Goïta, and D. C. He, "Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geo database and prior knowledge," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 143–153, Nov. 2010.

[12] S. H. Du, F. L. Zhang, and X. Y. Zhang, "Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 105, pp. 107–119, Apr. 2015.

[13] H. Mayer, "Object extraction in photogrammetric computer vision," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 63, no. 2, pp. 213–222, Mar. 2008.

[14] K. Kraus and N. Pfeifer, "Determination of terrain models in wooded areas with airborne laser scanner data," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 53, no. 4, pp. 193–203, Aug. 1998.

[15] J. Niemeyer, F. Rottensteiner, and U. Uwe Soergel, "Contextual classification of LiDAR data and building object detection," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 87, pp. 152–165, Jan. 2014.

[16] M. Awrangjeb, M. Ravanbakhsh, and C. Fraser, "Automatic detection of residential buildings using LiDAR data and multi spectral imagery," *ISPRS Int. J. Photogramm. Remote Sens.*, vol. 65, no. 5, pp. 457–467, Sep. 2010.

[17] X. Meng, L. Wang, and N. Currit, "Morphology-based building detection from airborne LiDAR data," *Photogramm. Eng. Remote Sens.*, vol. 75, no. 4, pp. 437–442, Apr. 2009.

[18] A. Sampath and J. Shan, "Building boundary tracing and regularization from airborne LiDAR point clouds," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 7, pp. 805–812, Jul. 2007.

[19] D. Lee, K. Lee, and S. Lee, "Fusing of LiDAR and imagery for reliable building extraction," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 215–226, Feb. 2008.

[20] F. Rottenstainer and C. Briese, "Automatic generation of building models from LiDAR data and the integration of aerial images," in *Proc. ISPRS Working Group III/3 Workshop*, vol. 34, 2003, pp. 174–180.

[21] C. Alexander, S. Smith-Voysey, C. Jarvis, and K. Tansey, "Integrating building footprints and LiDAR elevation data to classify roof structures and visualize buildings," *Comput. Environ. Urban Syst.*, vol. 33, no. 4, pp. 285–292, Jul. 2009.

[22] Z. Y. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from LiDAR remote sensing data," *Lands Urban Plan.*, vol. 130, no. 1, pp. 134–148, Oct. 2014.

[23] M. Baatz and A. Schäpe, "Multi resolution segmentation–An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informations Verarbeitung XII*, J. Strobl, T. Blaschke, and G. Griesebner, Eds. Karlsruhe, Germany: Wichmann Verlag, 2000, pp. 12–23.
[24] L. Y. Li, Y. Chen, T. B. Xu, R. Liu, K. F. Shi, and C. Huang, "Super-resolution mapping of wetland inundation from remote sensing imagery based on integration of back-propagation neural network and genetic algorithm," *Remote Sens. Environ.*, vol. 164, no. 1, pp. 142–154, Jul. 2015.
[25] J. Zeng, H. F. Guo, and Y. M. Hu, "Artificial neural network model for identifying taxi gross emitter from remote sensing data of vehicle emission," *J. Environ. Sci.*, vol. 19, no. 4, pp. 427–431, Apr. 2007.
[26] R. Prasad, A. Pandey, K. P. Singh, V. P. Singh, R. K. Mishra, and D. Singh, "Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions," *Adv. Space Res.*, vol. 50, no. 3, pp. 363–370, Aug. 2012.
[27] J. H. Zhou, Y. F. Zhou, and W. S. Mu, "Mathematic descriptors for identifying plant species: A case study on urban landscape vegetation," *J. Remote Sens.*, vol. 15, no. 3, pp. 524–538, Sep. 2011, [in Chinese].
[28] J. C. Price, "Estimating vegetation amount from visible and near infrared reflectances," *Remote Sens. Environ.*, vol. 41, no. 1, pp. 29–34, Jul. 1992.
[29] J. H. Zhou, J. Qin, K. Gao, and H. B. Leng, "SVM-based soft classification of urban tree species using very high-spatial resolution remote-sensing imagery," *Int. J. Remote Sens.*, vol. 37, no. 11, pp. 2541–2559, May 2016.

Junfei Xie received the Ph.D. degree in atmospheric physics and atmospheric environment from the Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, in 2016.

He is a Senior Engineer in urban ecological benefit evaluation based on geo-informatics. He is a Researcher working in the Beijing Institute of Landscape Architecture, Beijing, China.

Jianhua Zhou received the M.Sc. degree in remote sensing and GIS from the East China Normal University, Shanghai, China, in 1999.

She is an Associate Professor in geo-informatics (cartography/remote sensing/GIS). She is a Teacher and a Researcher working in the Key Laboratory of Geographic Information Science, Ministry of Education of China, Shanghai, China.