This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Spatiotemporal Detection and Analysis of Urban Villages in Mega City Regions of China Using High-Resolution Remotely Sensed Imagery

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

***Abstract*—Due to the rapid urbanization of China, many villages in the urban fringe are enveloped by ever-expanding cities and become so-called urban villages (UVs) with substandard living conditions. Despite physical similarities to informal settlements in other countries (e.g., slums in India), UVs have access to basic public services, and more importantly, villagers own the land legitimately. The resulting socio-economic impact on urban development attracts increasing interest. However, the identification of UVs in previous studies relies on fieldwork, leading to late and incomplete analyses. In this paper, we present three scene-based methods for detecting UVs using high-resolution remotely sensed imagery based on a novel multi-index scene model and two popular scene models, i.e., bag-of-visual-words and supervised latent Dirichlet allocation. In the experiments, our index-based approach produced Kappa values around 0.82 and outperformed conventional models both quantitatively and visually. Moreover, we performed multitemporal classification to evaluate the transferability of training samples across multitemporal images with respect to three methods, and the index-based approach yielded best results again. Finally, using the detection results, we conducted a systematic spatiotemporal analysis of UVs in Shenzhen and Wuhan, two mega cities of China. At the city level, we observe the decline of UVs in urban areas over the recent years. At the block level, we characterize UVs quantitatively from physical and geometrical perspectives and investigate the relationships between UVs and other geographic features. In both levels, the comparison between UVs in Shenzhen and Wuhan is made, and the variations within and across cities are revealed.**

***Index Terms*—China, scene-based classification, settlement, spatiotemporal analysis, urbanization, urban village (UV).**

I. Introduction

URBANIZATION in the developing world often leads to the problem of informal settlements (e.g., slums and shanty towns) [1] because of the growth of urban population and unplanned development. As one of the developing countries with a rapid increase in urbanization, China suffers from the wide spread of urban villages (UVs) in recent years. UVs, which are also known as "*chengzhongcun*" or "villages in the city" [2], are a special type of urban settlement resulting from the complicated socio-economic development of China.

In the last 30 years, peri-UVs are progressively enveloped by the expanding cities due to China's rapid urbanization, in which the residential areas of villages are left intact and the farmland is used for urban development [3]. Original villagers still own the residential areas collectively, but they are not allowed to alienate the land. Meanwhile, the migration of large-scale rural workers to cities creates a great demand for affordable housing along with the fast economic growth [4], [5]. Then, villagers build additional dwellings in the land and rent them to migrant workers and the poor. The rent becomes the major source of livelihood of landless villagers, and these areas become the so-called UVs. However, the development of UVs is neither authorized nor planned, resulting in small and crowded substandard buildings, poor sanitary conditions, absent infrastructure, and some social problems including crime and environmental pollution. Many cities have launched the demolition and redevelopment of UVs recently [6], [7]. Therefore, an up-to-date map of UVs is necessary for planners and policymakers; however, it is usually incomplete or unavailable. In fact, information about UVs, such as their changes, is basically collected by fieldwork, which is extremely labor and time intensive.

High-resolution remotely sensed imagery has been acknowledged as an important data source for urban mapping owing to the advantages of objectiveness, low cost, and global coverage. Despite successful cases such as central business districts (CBDs) [8], private gardens [9], and man-made structures [10], no study is implemented for detecting UVs. In contrast to these urban land cover/land use types, the mapping of UVs is challenging indeed because unplanned development leads to complex spectral and spatiotemporal patterns. For instance, various materials used in the construction of informal buildings result in a large variance of spectral reflectance.

On the other hand, some studies have been conducted for the identification of other urban settlements, and they mainly use the object-oriented method [11]. Hofmann [12] proposed to create a hierarchical network of objects with multiresolution segmentation. Then, objects of different levels are classified by their physical properties. The refinement of settlement areas is finally conducted by a rule-based classification. The similar method is used in [13]–[15] for detecting informal settlements

Manuscript received July 26, 2014; revised October 27, 2014 and November 24, 2014; accepted December 7, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 41101336 and Grant 91338111, by the Program for New Century Excellent Talents, University of China under Grant NCET-11-0396, and by the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (FANEDD) under Grant 201348.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

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

Digital Object Identifier 10.1109/TGRS.2014.2380779

0196-2892 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

in Voi SE-Kenya, Delhi India, and Casablanca Morocco, respectively. However, the segmentation of high-resolution images, particularly urban imagery, is still challenging, and the results usually need to be adjusted or corrected interactively. Moreover, the human-defined rules and the dependence of training samples on the segmentation make the object-oriented method lack flexibility and transferability.

Recent studies of scene classification, in which a scene is an image block that belongs to some user-defined semantic category and usually contains various objects, have exhibited its potential for the semantic annotation and the identification of complex man-made structures [16], [17]. Compared with the object-oriented method, which mainly focuses on the local properties of one object, the scene-based method has the advantage in describing the relationships between objects as a whole; thus, it is suitable for dealing with complex categories related to various objects. The bag-of-visual-words (BOVW) model has been well explored for land use classification [18]–[20]. Latent Dirichlet allocation (LDA) [21], which is a popular model in text analysis that finds latent topics in a document, has been adapted for semantic annotation of satellite images [22], [23]. Vatsavai *et al.* [24] proposed an unsupervised semantic framework based on LDA to identify nuclear power plants. However, the performance of the scene-based method has not been evaluated for the detection of urban settlements including UVs.

Motivated by above facts, we conducted a systematic study of scene-based methods for detecting UVs, where two well-known models, namely, BOVW and supervised LDA (sLDA) [25], were investigated, and a new index-based model was proposed. These methods are briefed as follows: 1) *BOVW approach*—Spectral and spectral–textural BOVW representations were learned for every scene using two low-level image features, i.e., spectral statistics and Gabor texture [26]. Then, we used two popular classifiers, i.e., support vector machine (SVM) and random forest (RF), to categorize these representations into UVs and non-UVs. 2) *sLDA approach*—Based on the BOVW representation, sLDA, which is a well-behaved variant of LDA, was used to learn the topic representation of scenes and infer the category of them. 3) *Index-based approach*—Because previous studies are mainly based on low-level features that hardly describe high-level categories, we proposed a new scene model based on two semantic indexes, i.e., morphological building index (MBI) [27] and NDVI. MBI is a morphological approach for automatic building extraction from high-resolution images. Scenes were modeled with the indexes and classified by SVM and RF.

Moreover, because of the high correlations between multitemporal images, it is possible to reuse previous training samples for new images. Reuse is of importance due to the high cost of representative training samples. Given the different sources of training samples and images to be classified, some studies [28] propose to perform model adaptation to samples. In fact, transferable rules used in object-oriented approaches have been quantitatively studied [29] and tested on different software packages [30], whereas few scene-based studies focus on the transferability of samples. Thus, we conducted multitemporal classification to evaluate the transferability of training samples with respect to the proposed methods.

TABLE I
DESCRIPTION OF SHENZHEN DATA SET

Acquisition date	Satellite	Spectral range
2003/01/17 2005/12/17 2007/12/10 2010/05/26	QuickBird	4 bands 450–900 nm
2010/11/03 2012/03/25	WorldView-2	8 bands 400–1040 nm

Another main focus of this paper is the spatiotemporal analysis of UVs. As also the result of urbanization in the developing world, urban expansion and informal settlements have received much attention, where remotely sensed data play an important role in the spatial analysis [31]–[35]. As far as UVs are concerned, however, fieldwork remains the only way to identify them in previous studies. This paper fills the gap based on the proposed detection algorithms. We experimented with the multitemporal high-resolution images of Shenzhen and Wuhan, two representative mega cities of China. UVs over the recent years were mapped based on the detection results. Then, using the maps and the indexes, we conducted a detailed spatiotemporal analysis from several aspects including spatiotemporal distributions, physical and geometrical characteristics, and relationships between UVs and other geographic features (i.e., roads, parks, and commercial centers).

II. Study Areas and Data

The urban areas of Shenzhen and Wuhan, two mega cities in China, were chosen for this study, where the former is a young immigrant city, and the latter is an enormous inland city. They both undergo severe problems of UVs in recent decades, but UVs in two cities have different development patterns and appearances because of different economic and cultural characteristics. For instance, buildings in Shenzhen's UVs are mostly over six storeys, whereas buildings less than three storeys high are most common in Wuhan's UVs. In addition, the development of UVs in Shenzhen have been discussed in a few studies [6], [36], [37] based on the fieldwork mainly conducted by the local government; however, no study pays attention to Wuhan's UVs.

A. Shenzhen

Shenzhen, located in Guangdong province, was just a small village on the Pearl River Delta before it became China's first special economic zone (SEZ) in 1979. Since then, Shenzhen has experienced fast economic development and became one of the biggest cities in China. During this period, Shenzhen's population increases from less than 100 000 in 1979 to over 10 million in 2010 accompanied with huge migration all over the country, and UVs rapidly spread across the city. Recently, it is estimated that about half of Shenzhen's population live in UVs [38].

Remotely sensed data from QuickBird and WorldView-2 satellites acquired during 2003–2012 were used (see Table I), which has been radiometrically calibrated. QuickBird data have a spatial resolution of 2.4 m with an image size of 5360 × 4507 pixels,

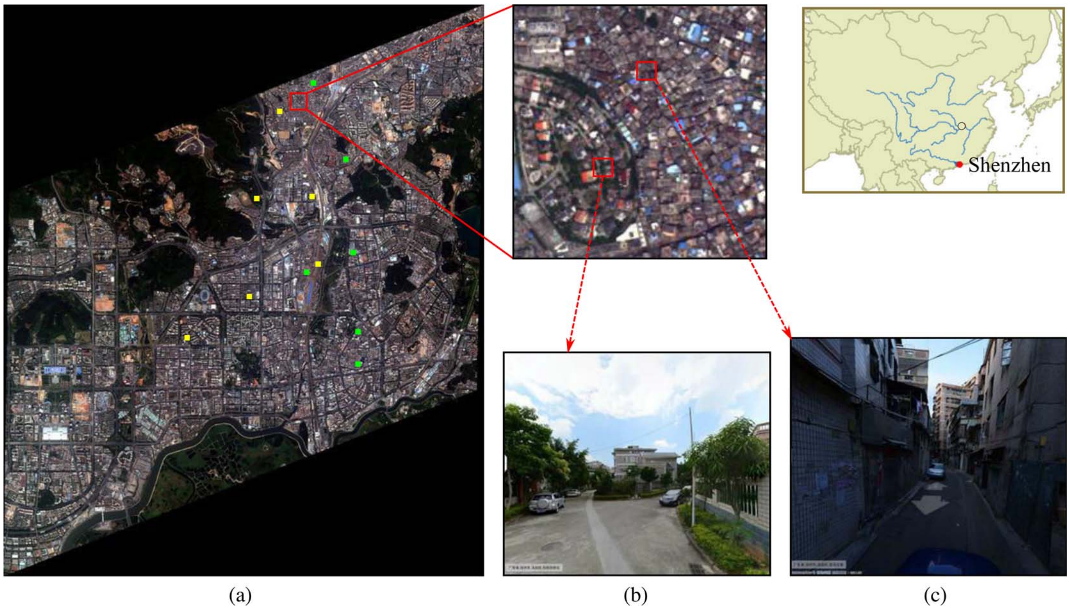


Fig. 1. Study area in Shenzhen. (a) WorldView-2 image acquired on 2012/03/05 and photographs of (b) formal settlements and (c) UVs are shown. (Source of photographs: Tencent Maps).

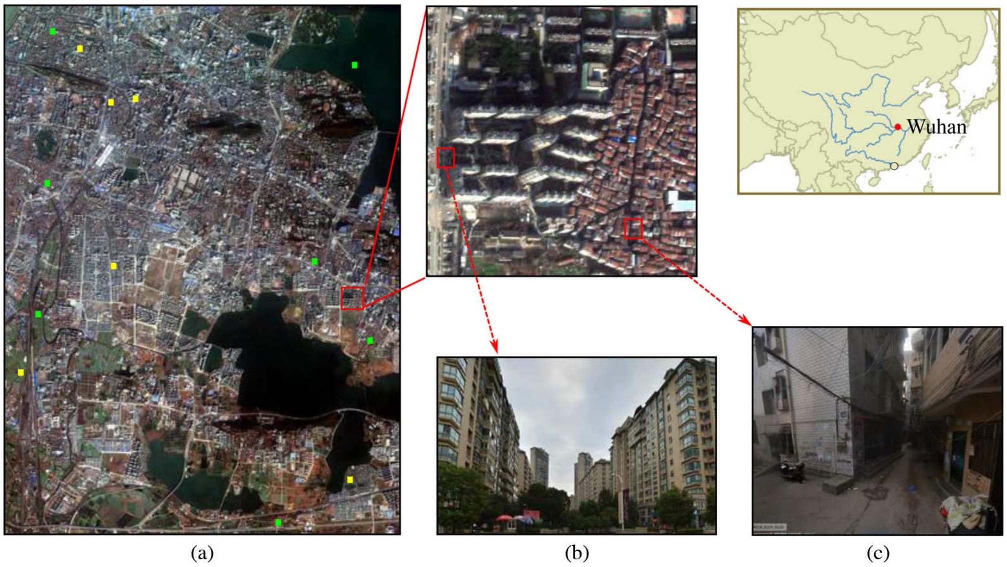


Fig. 2. Study area in Wuhan. (a) GeoEye-1 image acquired on 2009/01/25 and photographs of (b) formal settlements and (c) UVs are shown. (Source of photographs: Tencent Maps).

and WorldView-2 data have a spatial resolution of 2 m with an image size of 6433 × 5409 pixels. All images are between 114°3′E to 114°9′E and 22°30′N to 22°37′N and cover about 91.84 km^2 of SEZ including the Futian CBD (see Fig. 1).

B. Wuhan

Wuhan, the capital of Hubei province, is located in central China. Wuhan is at the confluence of the Han and Yangtze Rivers and is divided by rivers into three towns, i.e., Hankou, Wuchang, and Hanyang. Unlike Shenzhen, Wuhan has been the industrial, commercial, and cultural center of the Central China for centuries despite the slower development than Shenzhen in recent decades. According to the census data, Wuhan's population increases from 6.9 million in 1990 to about 9.8 million in 2010, and, to date, it has exceeded 10 million. Owing to the huge city scale, Wuhan is one of the cities with the most UVs in China.

Remotely sensed data of Wuhan were acquired from the GeoEye-1 satellite on 2009/01/25 and 2012/12/09 with a spatial resolution of 2 m and four spectral bands (blue 450–510 nm, green 510–580 nm, red 655–690 nm, and near infrared 780–20 nm). Both images are within a rectangular bounding box of 114°17′E to 114°22′E and 30°27′N to 30°33′N and have an image size of 5550 × 4156 pixels. They cover about 92.26 km^2 of Wuhan city (see Fig. 2).

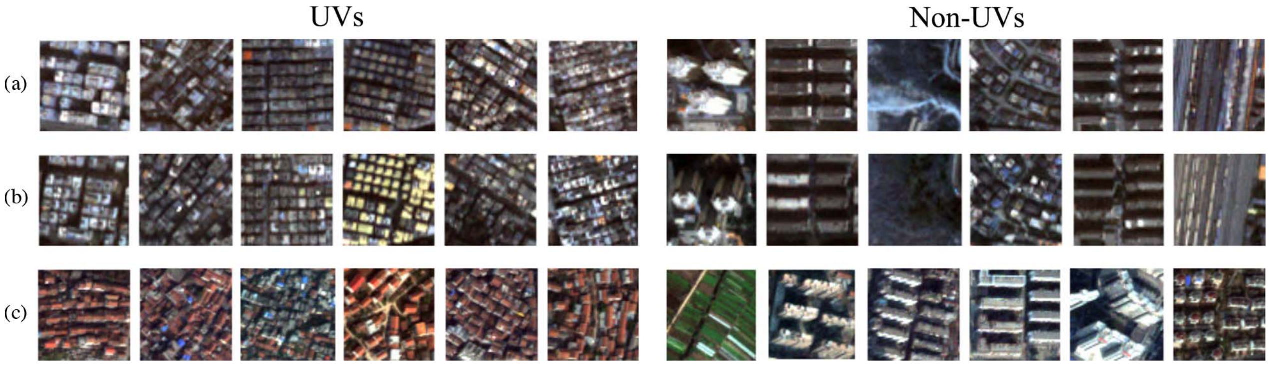


Fig. 3. Example scenes of UVs and non-UVs extracted from the (a) 2003 Shenzhen QuickBird image, the (b) 2007 Shenzhen QuickBird image, and the (c) 2009 Wuhan GeoEye-1 image. The locations of UVs and non-UVs are indicated by the green and yellow points, respectively, in Figs. 1 and 2.

III. Detection of UVs

UVs are a special type of urban settlement in China. They mostly originate from villages. They are surrounded by planned urban areas and are usually adjacent to skyscrapers, highways, and other modern urban infrastructures. Because of the absent urban management, UVs are strongly settled and disorderly developed. Compared with other urban areas, buildings in UVs are much smaller. These buildings are densely distributed in UVs and occupy most spaces. Accordingly, there is little vegetation and public space, which are the fundamental components of planned urban areas. Except for these common features, UVs in different cities usually have different appearance. Many UVs in Wuhan are built with red bricks, which are never used in Shenzhen's UVs, and have red roofs. In addition, UVs in Shenzhen usually have higher buildings and larger building density than that in Wuhan.

Briefly, in the context of high-resolution imagery, observable characteristics that distinguish UVs from formal residential areas are mainly as follows. 1) Most spaces of UVs are occupied by small buildings, leaving little for vegetation, streets, and bare ground. 2) In contrast to formal residential areas, UVs tend to have a disordered layout. Then, proportions of objects of several major classes (i.e., buildings and vegetation) as well as the spatial configuration are the key to detecting UVs.

In this paper, we carry out the detection at the scene level. A scene, which refers to an image block here (see Fig. 3), is a larger semantic entity than the object, hence a better representation of the UV that contains various objects. We chose 120 m × 120 m as the scene size according to the scale of UVs in the real world. For example, the resolution of Wuhan GeoEye-1 image in Fig. 3 is 2 m, and the scene size is 60 × 60 pixels accordingly. Many studies find it useful to regard the scene as a collection of visual words and propose various models for scene classification, such as BOVW and sLDA [18], [19], [22]. We also present a new scene model based on semantic indexes from the object-oriented point of view. These models are described in Section III-A.

The algorithm for detecting UVs is summarized in Fig. 5. First, the large image is partitioned into scenes of size 120 m × 120 m with an overlapping of 60 m. Because the UV does not necessarily occupy most spaces of a scene, the overlapping can decrease the omissions of UVs. Next, the representation of each scene is calculated according to different scene models, in which the proportions and the spatial configuration of objects are implicitly encoded. Finally, these representations are classified into UVs and non-UVs.

A. Scene Representation

1) BOVW Model: The BOVW model stems from the idea in text analysis that a document could be presented by word frequencies without regard to their order. This way, we divided the scene into overlapping patches, i.e., visual words. For each patch, we computed the mean and variance of blue, green, red, and near-infrared bands, which are the four common bands of two data sets. The resulting 8-D feature vector describes spectral information of the patch. Moreover, we used Gabor filters to extract textural features as supplementary information. A 2-D Gabor filter is defined as

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\pi\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)} e^{i(u_0x+v_0y)} \tag{1}$$

where σ_x, σ_y are scale parameters along x and y, and u_0, v_0 are spatial frequencies of the filter, which can be also expressed in polar coordinates as radial frequency f and orientation θ. Three visible bands of the patch were filtered with a set of Gabor filters, where $\sigma_x = \sigma_y = \{4, 6\}$, $f = \{0.006, 0.02, 0.06\}$, and $\theta = \{0, \pi/3, 2\pi/3\}$, resulting in a 54-dimensional feature vector.

We quantified all spectral descriptors extracted from training samples with K-means clustering. The cluster centers form a dictionary. Any new spectral descriptor could be quantified by simply assigning the label of the closest cluster center. Then, the spectral representation of a scene is the frequencies of the labeled spectral descriptors, i.e., a vector whose size is equal to the size of the dictionary. Similarly, the Gabor representation was created. Finally, a spectral–textural representation was constructed by stacking the spectral and the supplementary Gabor feature vectors. These low-level features cannot effectively reflect the semantic characteristics (e.g., building size), but they can reflect the spectral and textural differences between built-up areas in the UVs and that in other urban areas. We used the spectral and the spectral–textural feature vectors for classification separately.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

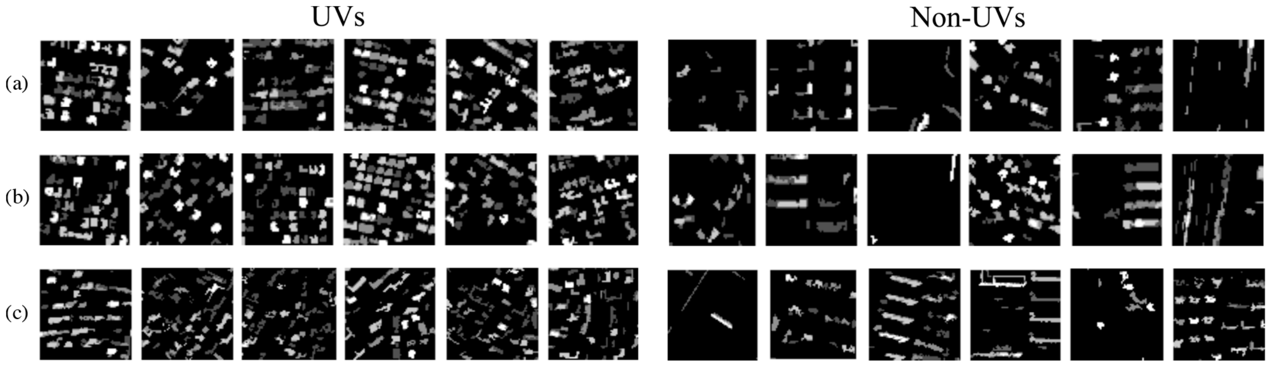


Fig. 4. MBI feature images corresponding to the scenes in Fig. 3.

2) sLDA Model: LDA [21] is a generative probabilistic model for collections of discrete data such as text corpora. It models a document as a mixture of latent topics, and the proportion of topics is an intermediate semantic representation of the document. Since the analogy between documents and scenes has been made based on the BOVW model, given the number of topics K, we can similarly use LDA to represent a scene as a mixture of topics that indicate the proportions of scene types. The generative process of LDA is briefed as follows.

1) For each document d, a topic proportions θ is chosen according to a Dirichlet distribution $\mathrm{Dir}(\alpha)$ over K topics.
2) For each word position in the document, a topic z is chosen from the multinomial distribution $\mathrm{Multi}(\theta)$ first, and then, a word w is generated according to the topic–word distribution β.

The process above shows that LDA is an unsupervised model and that the topics are not specifically learned for classification. Since we are concerned with the category rather than the topics of a scene, this paper uses sLDA, which is a variant of LDA. Compared with LDA, sLDA adds a variable to denote the category of each scene. Based on the variational method described in [25], we could learn a model that fits the categories of known scenes better than the unsupervised version. Then, the learned model can be used to infer the categories of unknown scenes. Similar to the classification in the BOVW approach, the sLDA model was performed on both the spectral and the spectral–textural vectors.

3) Index-Based Model: Conventional approaches use low-level features to describe spectral and textural characteristics of a scene, but a semantic gap exists between low-level features and high-level categories as demonstrated in [23]. To fill the gap, we have employed multiple features and some techniques such as LDA in Section III-A1 and A2. Another way is to use high-level features that straightforwardly indicate semantic categories. They could enable an object-oriented way to model a scene without the segmentation. For example, a connected region associated with large NDVI values can be treated as a vegetation object. Then, complex categories such as the UV can be modeled with multiple high-level features in a human-understandable manner.

TABLE II
COMPOSITION OF THE MBI REPRESENTATION THAT CHARACTERIZES BUILDINGS IN A SCENE. HISTOGRAMS USE EQUAL-WIDTH BINS ACCORDING TO THE RANGE OF CHARACTERISTICS EXCEPT THE AREA HISTOGRAM IN WHICH THE BREAKPOINTS ARE EMPIRICALLY CHOSEN

Characteristic of buildings	Range	Length of the histogram
Area	$(0, +\infty)$	9
Area ratio	$(0, 1]$	10
Aspect ratio	$(0, 1]$	10
Orientation	$[0, \pi)$	4
Number of buildings		1
MBI representation		34

Considering that buildings are the primary land cover in UVs and other settlements, we use the MBI [27], which is a high-level feature for building extraction. The calculation of MBI is briefed below. First, a brightness image is obtained by keeping the maximum reflectance value across visible bands for each pixel. Next, a linear structural element (SE) of size s is constructed, and the top-hat by construction of the brightness image is computed with the SE at multiple directions. Finally, MBI of scale s is formulated as

$$\mathrm{MBI}(s) = \frac{\sum_d (\mathrm{TH}(d, s + \Delta s) - \mathrm{TH}(d, s))}{N_d} \tag{2}$$

where N_d is the number of directions, and $\mathrm{TH}(d, s)$ represents the top-hat by construction with the SE of size s at direction d.

According to the common building size in UVs, we computed the MBI feature of the whole image at scales 5 and 7 with $\Delta s = 2$ and direction $d = \{0, \pi/4, \pi/2, 3\pi/4\}$. Fig. 4 presents MBI feature images of scenes in Fig. 3, which were computed at scale 5. We treat each connected component $b_i (i = 1, 2, \ldots, N)$ in the MBI feature images as a building object and found the minimum enclosing rectangle B_i. The following characteristics of each object were computed: the area of b_i, the area ratio of b_i to B_i, the aspect ratio of B_i, and the orientation of B_i. Because the spatial characteristics (e.g., size and arrangement) of buildings are different in UVs and non-UVs, the statistics of these characteristics can be used to distinguish them. Then, for each scene, a 34-dimensional feature vector was constructed by stacking the histograms of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

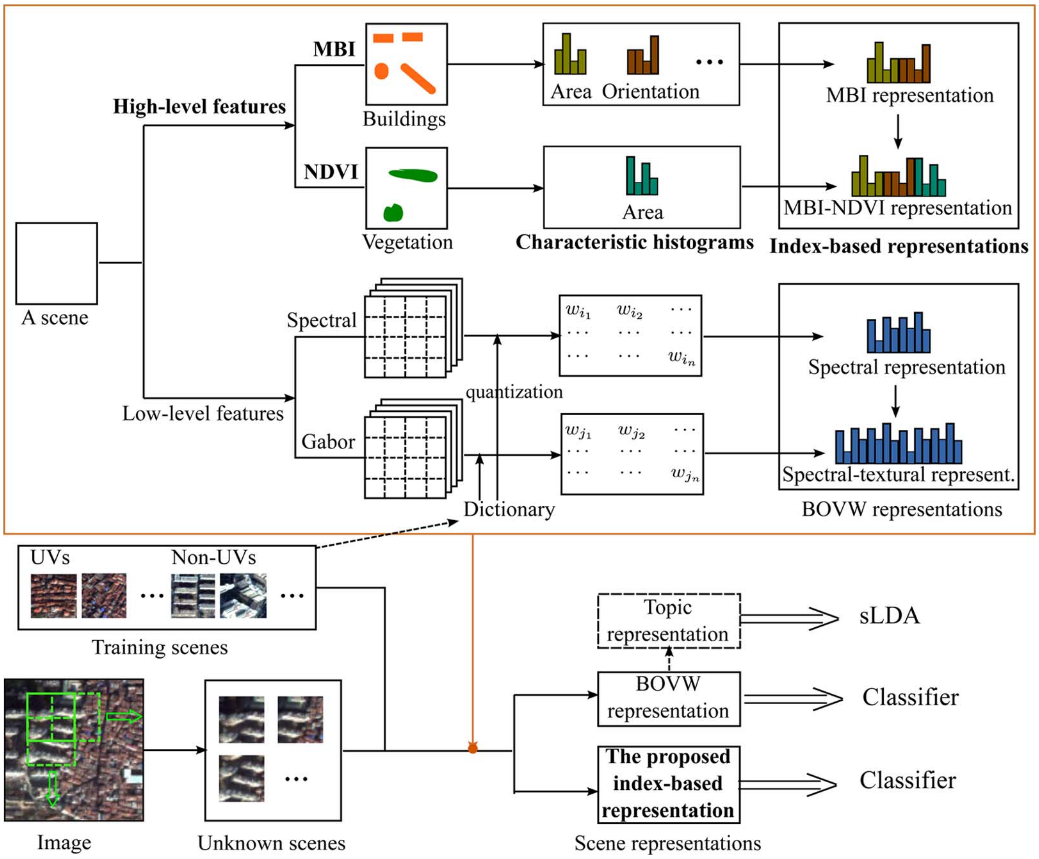


Fig. 5. Flowchart of the proposed algorithm for detecting UVs.

these characteristics (see Table II), which describes both the proportions and the spatial arrangement of buildings in a scene. Finally, we obtained a 68-D feature by combining the MBI feature vectors of two scales.

Apart from buildings, vegetation is another primary land cover in urban settlements. Despite the small proportion of vegetation in UVs, green coverage of formal residential areas is usually high. NDVI, which is a widely used vegetation index, was computed and transformed to binary form with an empirical threshold. We computed areas of connected components in the NDVI feature image. Other characteristics (e.g., orientation) were not considered since vegetation generally has no recognizable spatial patterns. Then, a ten-bin histogram of the area feature was computed for each scene and added to the MBI representation as supplementary information, resulting in a 78-D MBI-NDVI feature vector.

B. Experiments, Results, and Comparison

1) Experimental Settings: In the sLDA approach, as illustrated in Fig. 5, the spectral and spectral–textural vectors were classified by the sLDA model with the topic representation implicitly learned, which could be viewed as a classifier with the number of topics as the parameter. In the BOVW and the index-based approaches, we used two popular classifiers, i.e., SVM with the RBF kernel and RF, to classify the learned representations. In the experiments, all the parameters of classifiers were determined using cross validation.

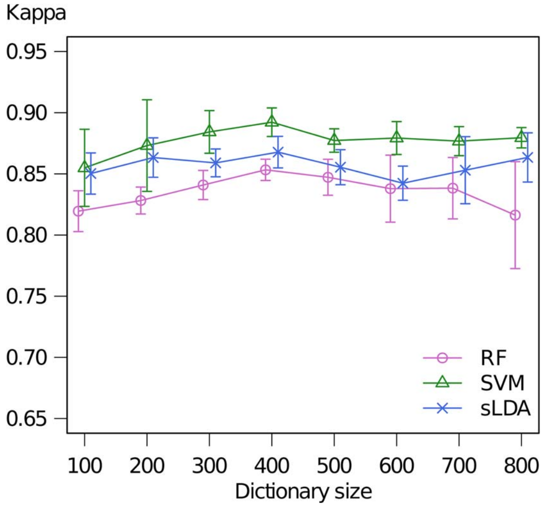


Fig. 6. Classification performance of SVM, RF, and sLDA under different dictionary sizes on the spectral representation of 2003 QuickBird scenes.

For each image in Shenzhen and Wuhan data sets, 12 training samples, i.e., six UVs and six non-UVs, were selected. Fig. 3 shows training samples of the 2003 and 2007 Shenzhen QuickBird images and the 2009 Wuhan GeoEye-1 image. Within each data set, training samples are at the same locations [see Fig. 3(a) and (b)]. Then, we partitioned each image into overlapped scenes and selected all scenes of which more than 90% are occupied by UVs as positive testing samples, whose

TABLE III
CLASSIFICATION ACCURACIES OF THE BOVW APPROACH

		Shenzhen						Wuhan	
	classifier	2003 QB	2005 QB	2007 QB	2010 QB	2010 WV-2	2012 WV-2	2009 GE-1	2012 GE-1
spectral	SVM	**0.89±0.01**	0.82±0.02	0.74±0.01	0.57±0.18	0.64±0.25	0.84±0.01	0.48±0.12	0.37±0.06
	RF	0.85±0.01	**0.84±0.02**	**0.75±0.02**	**0.82±0.02**	0.63±0.06	0.82±0.01	**0.66±0.04**	0.53±0.06
spectral-textural	SVM	0.85±0.03	0.76±0.04	0.71±0.04	0.66±0.06	**0.76±0.01**	0.79±0.02	0.55±0.03	0.59±0.07
	RF	0.83±0.02	**0.84±0.02**	0.75±0.03	0.80±0.03	0.69±0.04	**0.85±0.01**	**0.66±0.04**	**0.77±0.03**

TABLE IV
CLASSIFICATION ACCURACIES OF THE sLDA APPROACH

		Shenzhen						Wuhan	
	topic number	2003 QB	2005 QB	2007 QB	2010 QB	2010 WV-2	2012 WV-2	2009 GE-1	2012 GE-1
spectral	10	0.86±0.02	0.80±0.04	0.71±0.03	0.52±0.20	0.69±0.05	0.81±0.01	0.51±0.04	0.24±0.05
	15	**0.87±0.01**	0.81±0.02	0.72±0.02	0.49±0.18	0.70±0.02	**0.82±0.01**	**0.55±0.04**	0.27±0.01
	20	0.87±0.01	**0.82±0.01**	**0.73±0.02**	0.49±0.18	0.71±0.03	0.82±0.02	0.54±0.02	0.28±0.02
	25	0.85±0.01	0.78±0.02	0.70±0.01	0.60±0.17	**0.72±0.02**	0.81±0.02	0.53±0.03	**0.40±0.04**
	30	0.86±0.03	0.78±0.03	0.70±0.02	**0.70±0.12**	0.72±0.02	0.81±0.0	0.50±0.04	0.40±0.072
spectral-textural	10	0.87±0.01	0.81±0.02	0.72±0.02	0.49±0.17	0.71±0.02	0.81±0.01	0.54±0.04	0.27±0.03
	15	0.86±0.01	0.81±0.01	0.72±0.02	0.57±0.20	0.70±0.01	0.82±0.02	0.53±0.02	0.27±0.01
	20	0.83±0.03	0.77±0.02	0.70±0.04	0.63±0.19	0.69±0.03	0.80±0.02	0.48±0.05	0.37±0.05
	25	0.85±0.02	0.78±0.02	0.70±0.02	0.68±0.11	0.71±0.02	0.81±0.03	0.50±0.06	0.40±0.04
	30	0.84±0.03	0.78±0.02	0.70±0.03	0.58±0.20	0.72±0.02	0.82±0.02	0.51±0.05	0.40±0.06

numbers are about 190 and 90 for images in Shenzhen and Wuhan data sets, respectively. Given the larger proportion of non-UVs than UVs, for every image, we randomly chose 600 negative testing samples from scenes that have no overlapping with UVs. Kappa was used for accuracy assessment since the unbalanced number of positive and negative testing samples.

The construction of BOVW representation, which is used in BOVW and sLDA approaches, depends on the dictionary size. We learned dictionaries with size $K = \{100, 200, \ldots, 800\}$ using K-means and repeated ten times with each size. Then, for each size, feature vectors of scenes were generated and classified ten times, and the average classification accuracy and standard deviation was used as the final classification result. In the range 100–800, as illustrated in Fig. 6, a significant relation between the dictionary size and the classification accuracy was not found. There is a reasonable explanation that visual words that can distinguish UVs from non-UVs are a small part of these dictionaries. Therefore, we give only the best accuracies and omit corresponding dictionary sizes for simplicity in the following sections.

2) Results: The results produced by the BOVW approach are presented in Table III. Except for 2009 GeoEye-1, the optimal Kappa values of all images exceeded 0.75—an encouraging accuracy. Two representations produced comparable accuracies for most images, and for the exceptions 2010 WorldView-2 and 2012 GeoEye-1, the spectral–textural representation produced substantial improvements of 0.12 and 0.24, respectively, owing to the supplementary Gabor textural feature. In terms of classifiers, RF behaved better than SVM in most cases and cooperated better with the textural feature. After the textural feature was added, RF yielded slightly lower accuracies for two images, whereas SVM produced four lower Kappa values with the maximum decrease of 0.06.

We applied the sLDA model to spectral and spectral–textural vectors with the number of topics ranging from 10 to 30, and the results are shown in Table IV. The accuracies of the Shenzhen data set are acceptable, and the Kappa values of the Wuhan data set were all less than 0.6 where even the additional textural feature did not improve the results. The spectral representation achieved the highest accuracy for every image with a slight edge less than 0.02 over the spectral–textural representation. Moreover, Table IV indicates that 10–15 topics are enough to obtain Kappa values larger than 0.8, i.e., a good distinction between UVs and non-UVs, if such an accuracy is reachable. Although some results seemed to improve by further increasing the number of topics, such as the accuracies of the 2010 QuickBird and 2012 GeoEye-1 images, they were too poor or variable to be reliable.

The results produced by index-based models are shown in Table V. With vegetation information, MBI-NDVI gave better results than MBI on the Shenzhen data set, whereas such an improvement was not observed in the 2009 GeoEye-1 image of the Wuhan data set. It does not imply that the NDVI feature should be responsible for the inconsistent performance, because the accuracy also depends on the classifier. From the point of view of RF, the results were indeed more or less improved by the additional NDVI feature in all experiments, and the worst Kappa value of the MBI and MBI-NDVI representations still reached 0.75. On the other hand, the accuracy produced by SVM was sensitive to the NDVI feature, such as the large improvement of the 2007 QuickBird image and the surprising decline of the 2009 GeoEye-1 image.

TABLE V
CLASSIFICATION ACCURACIES OF THE INDEX-BASED APPROACH

		Shenzhen						Wuhan	
	classifier	2003 QB	2005 QB	2007 QB	2010 QB	2010 WV-2	2012 WV-2	2009 GE-1	2012 GE-1
MBI	SVM	0.66	0.74	0.61	0.75	0.86	0.83	**0.80**	0.68
	RF	0.75	0.79	0.77	0.76	**0.87**	0.90	0.76	0.78
MBI-NDVI	SVM	0.86	**0.91**	**0.93**	**0.79**	**0.87**	0.82	0.47	**0.80**
	RF	**0.88**	0.83	0.87	0.77	0.86	**0.91**	0.75	**0.80**

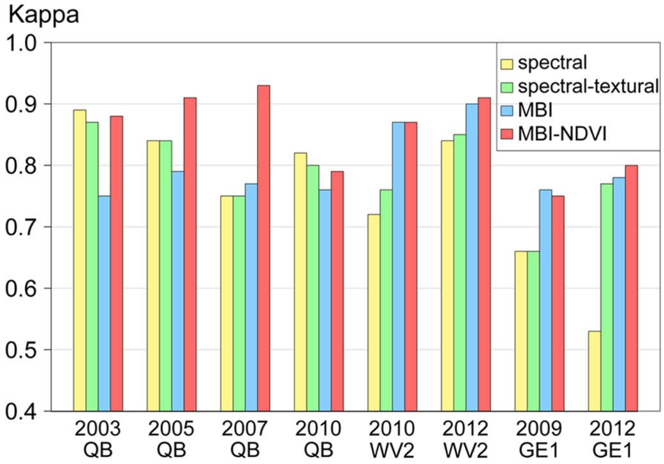


Fig. 7. Comparison of the best accuracies produced by different scene representations.

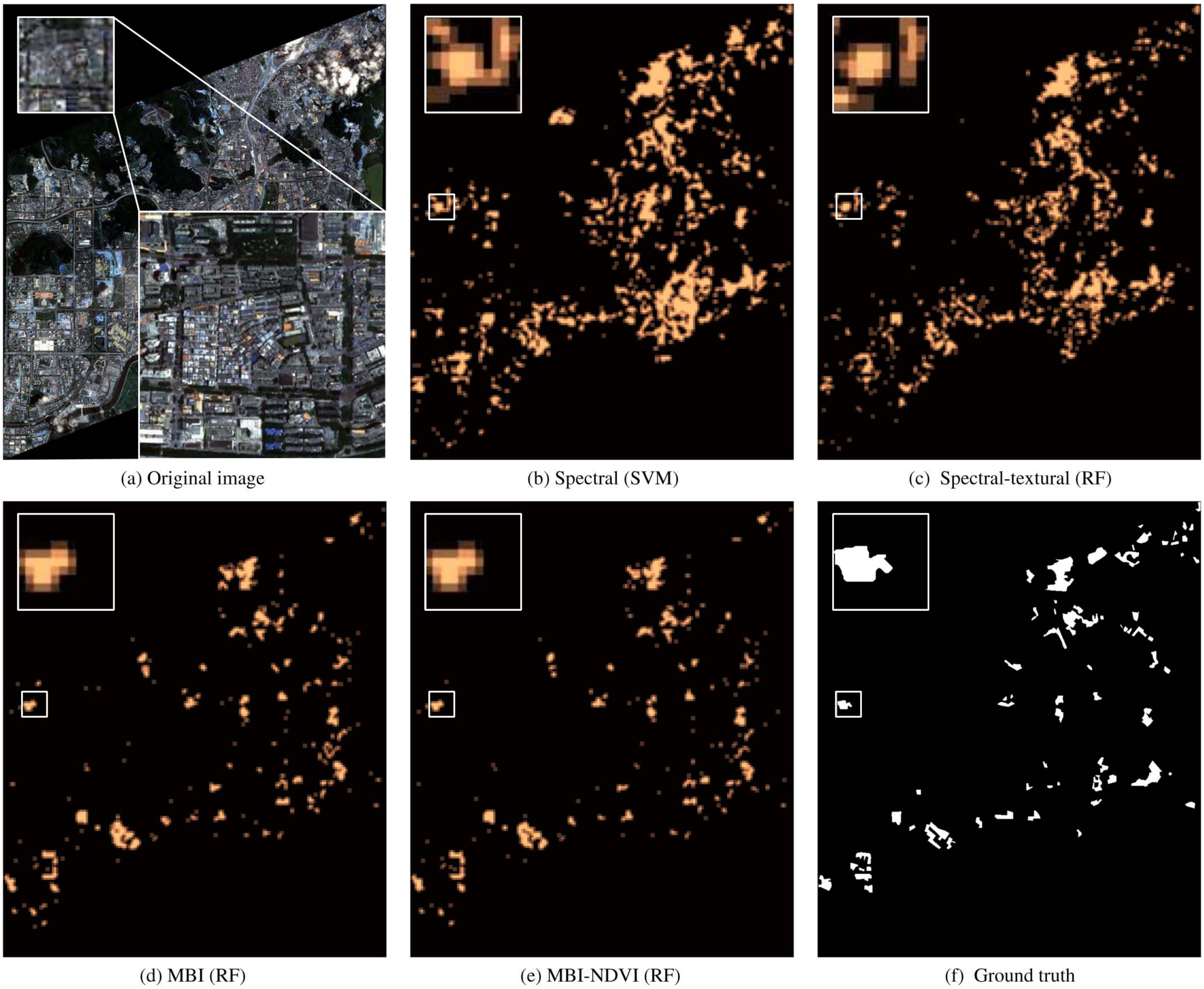
As shown in Tables III–V, the BOVW approach significantly outperformed the sLDA approach, and the index-based approach significantly outperformed the BOVW approach. Moreover, except for the topic representation implicitly learned in the sLDA model, there are four representations explicitly constructed in the proposed approaches: spectral and spectral–textural representations based on two low-level features, namely, MBI and MBI-NDVI representations based on two high-level features. For a comparison of them, we present their best Kappa values for every image in Fig. 7. MBI-NDVI yielded best results with an average of 0.82. MBI also performed well—even the worst Kappa value was larger than 0.75, and the spectral–textural representation had a similar performance for all images except 2009 GeoEye-1. Moreover, as supplementary information, both the Gabor texture and the NDVI feature improved the results significantly.

Despite the randomness of testing samples, the same accuracies of different approaches do not imply the same classification maps because of the small proportion of testing samples in all scenes. Therefore, a visual assessment is necessary. In Fig. 8, we present classification maps of the 2010 QuickBird image, for which various approaches produced comparable Kappa values (see Fig. 7). On the classification map, each place is covered by at most four scenes because the overlapping of two adjacent scenes is half of the scene size (see Fig. 5). The more scenes covering the place classified as the UV, the higher the possibility that the place belongs to the UV, and the lighter the place displayed on the map. Moreover, the scale of the classification map (i.e., the minimum differentiable unit) is half of the scene size because of the overlapping, and the polygons on the map can be viewed as the approximation of ground truth under this scale. Clearly, maps produced by the index-based approach are close to the reference map, and others seem noisy due to the numerous commissions of UVs, which is also illustrated by the comparison of enlarged rectangle areas. Moreover, the omissions in maps were rare, ensuring the practical use of these methods.

Finally, the computation times are shown in Table VI, which were obtained on a 3.07-GHz computer with 6-G RAM. The textural representation is computationally more expensive than the spectral or the index representations due to the high dimensionality of textural features. The MBI representation is most efficient in contrast with spectral, spectral–textural, and MBI-NDVI representations. However, at the cost of a slight increase in computational cost, the MBI-NDVI representation obtained significantly better results than MBI representation by adding vegetation features. As for classifiers, sLDA consumed more time than SVM and RF since it is a generative model, and the classification time is negligible compared with that for constructing scene representations.

3) Effect of the Scene Size on the Result: The scene size in previous experiments was 120 m, and the results look good. To better understand the relationships between results and the scene size, we carried out experiments with five scene sizes: 100, 110, 120, 130, and 140 m, where training scenes remained the same locations as previous ones despite different sizes. The best accuracies of four scene models are shown in Fig. 9. For Shenzhen images [see Fig. 9(a) and (b)], the best accuracy of each representation was basically achieved at 120 m; for Wuhan images, it was basically achieved at 100 m. To further illustrate the difference between the results, classification maps of the 2009 Wuhan GeoEye-1 image under different scene sizes are visualized in Fig. 10, which were produced by the MBI representation and SVM. As shown in Fig. 9(c), the first three maps have better accuracy than the other two maps. These three maps have no significant difference visually despite different scales, although the first map looks a little noisy. In the enlarged rectangle areas, the UV at the bottom can be detected with 100–120-m scenes, whereas larger scenes omitted it due to its small size.

4) Effect of Training Samples on the Result: In previous experiments, six representative fixed training samples per class were used for classification. Carefully selected samples can guarantee the usability of results, whereas random samples can make a fair comparison between different methods. Therefore, different numbers of random training samples were evaluated in our experiments. Particularly, we first selected about 15 training samples per class for each image, which included the


(a) Original image

(b) Spectral (SVM)

(c) Spectral-textural (RF)

(d) MBI (RF)

(e) MBI-NDVI (RF)

(f) Ground truth

Fig. 8. Comparison of classification maps of the 2010 Shenzhen QuickBird image.

TABLE VI
COMPUTATION TIMES IN SECONDS. (a) TOTAL TIME FOR CONSTRUCTING SCENE REPRESENTATIONS. THE DICTIONARY SIZE OF BOVW REPRESENTATIONS IS 400. (b) TOTAL TIME FOR CLASSIFYING ALL SCENES (INCLUDING TRAINING TIME), WHICH IS EVALUATED ON THE SPECTRAL REPRESENTATION

Representation	QB	WV-2	GE-1
Spectral	1102	1438	960
Textural	1845	2472	1690
MBI	602	995	453
NDVI	73	134	41

(a)

Classifier	QB	WV-2	GE-1
SVM	1	1	1
RF	3	3	3
sLDA	29	27	30

(b)

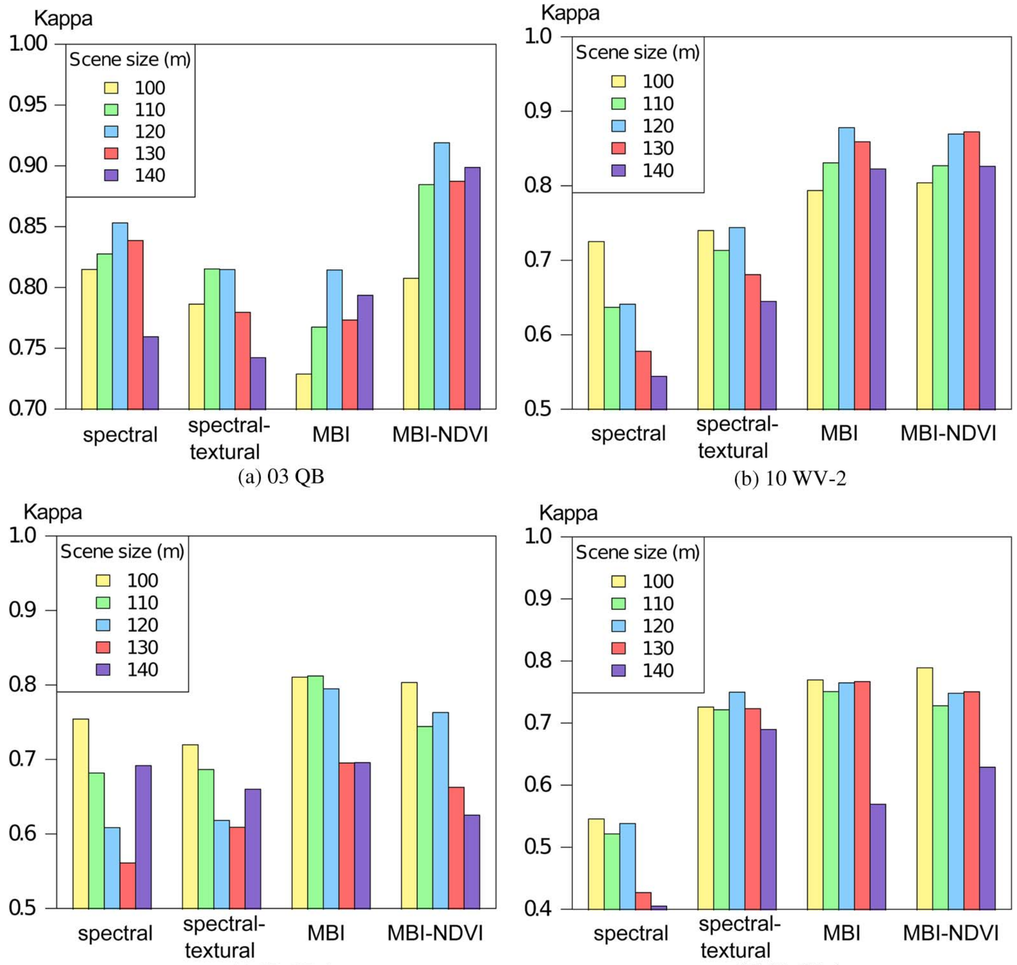
six existing samples. Then, 6, 9, and 12 samples per class were randomly chosen from all training samples and used for classification, which was repeated 15 times. Fig. 11 shows the best average accuracy together with the standard deviation for each representation, which was obtained from the results produced by different dictionary sizes and classifiers.

Compared with the previous results, the randomness of training samples decreased the accuracies for almost all cases and led to a larger accuracy variation on index-based results than BOVW-based results. The number of training samples also had a larger influence on the index-based results: Its increase significantly decreased the accuracy variation and improved the average accuracy. Overall, BOVW models are more dependent on the quality than the number of training samples, and, by contrast, index-based models need either representative or enough training samples. Moreover, Fig. 11 indicates that index-based models basically outperformed BOVW models when training samples were adequate in quality (see yellow bars) or in quantity (see red bars).

C. Multitemporal Classification

Here, we assess the transferability of training samples across multitemporal images with respect to different scene models,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) 03 QB

(b) 10 WV-2

(c) 09 GE-1

(d) 12 GE-1

Fig. 9. Best accuracies produced by four scene representations under different scene sizes.

which is not only an important task of pattern recognition but also an essential step of the automated procedure for monitoring UVs. For this aim, images were reclassified with previous training samples from the same sensor. Particularly, we classified the 2005, 2007, and 2010 QuickBird images using training samples of the 2003 QuickBird, the 2012 WorldView-2 image using training samples of 2010 WorldView-2, and the 2012 GeoEye-1 image using training samples of 2009 GeoEye-1.

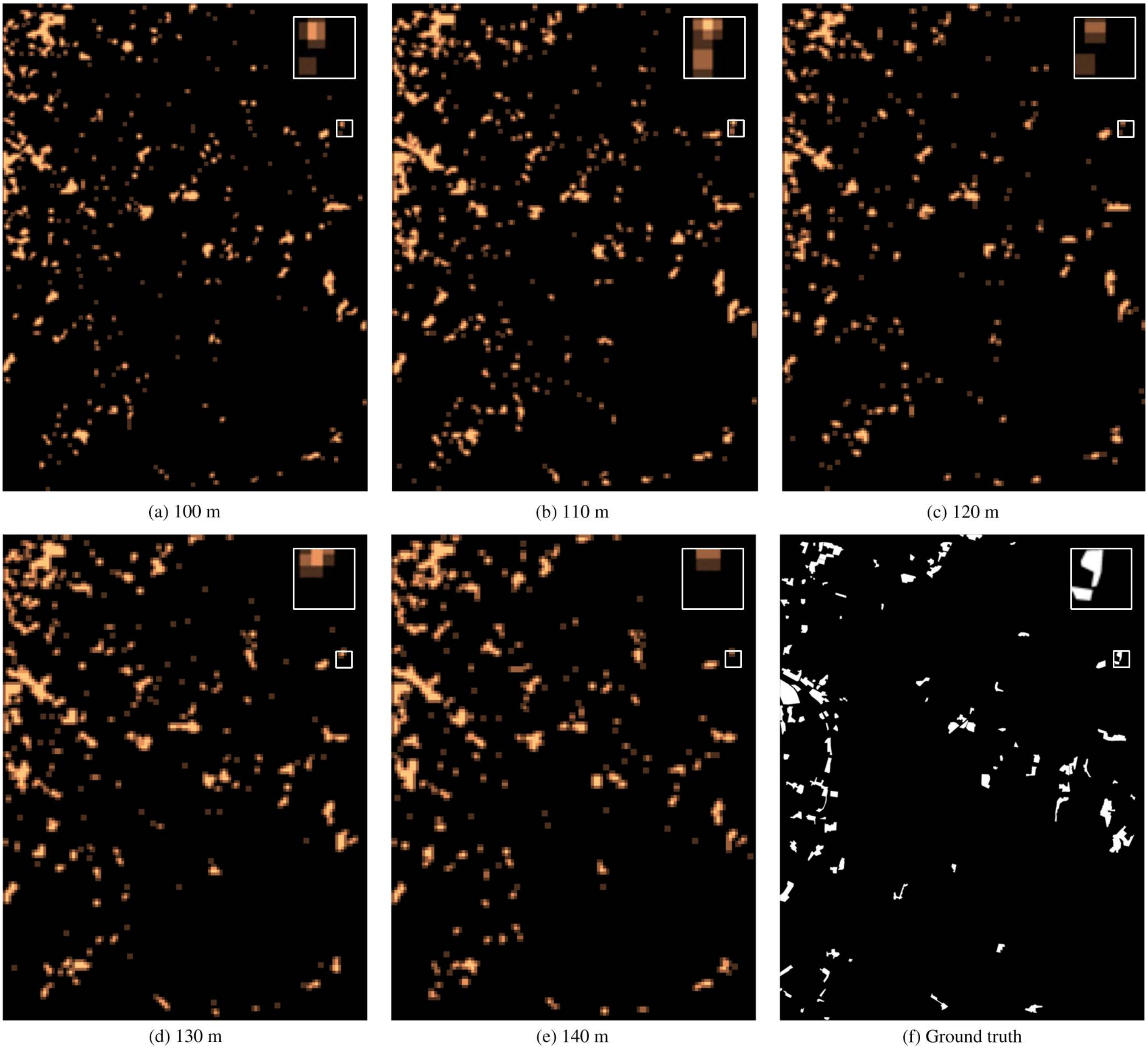
All three models were evaluated. Because the topic representation was implicitly learned inside the sLDA model, we regarded sLDA model as a classifier based on the BOVW representation. Then, SVM, RF, and sLDA were used to classify the BOVW representations; SVM and RF were used to classify the index-based representations. We found that SVM and RF cooperated best with the BOVW model and the index-based model, respectively, and the best accuracies are presented in Table VII. The accuracies produced by the BOVW model were all below 0.8, and they were still image dependent as the original results in Table III. The textural feature to some extent improved accuracies and alleviated the dependence on images. By contrast, the index-based approach consistently yielded results comparable to that in Table V, where MBI-NDVI still outperformed MBI.

D. Discussions

UVs have different shapes and sizes across cities. The scene size used for detecting UVs in different areas should be first determined by the common size of UVs. Fig. 9 shows the scene sizes achieving the best accuracies of 120 and 100 m in Shenzhen and Wuhan, respectively, which reflects a larger average size of UVs in Shenzhen than that in Wuhan. In fact, the results were robust to the scene size quantitatively and visually within a certain range. As indicated in Figs. 9 and 10, the size in 110–130 m and 100–120 m is appropriate for the UV detection in Shenzhen and Wuhan, respectively. The desired scale of the UV map is another factor to be considered for selecting scene size. The map of finer scale tends to be noisy, whereas a coarser map may omit small objects. Therefore, the postprocessing of UVs may need maps of multiple scales.

In the experiments, scene-based methods have proven effective for the detection of UVs. In particular, the index-based approach yielded satisfactory results in both accuracy assessment and visual inspection. The success is mainly attributed to the use of high-level features, which enable a straightforward way to model complex scenes. As far as UVs are concerned, we used MBI and NDVI to model a scene as the proportions and the spatial configuration of building and vegetation objects.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.


(a) 100 m (b) 110 m (c) 120 m

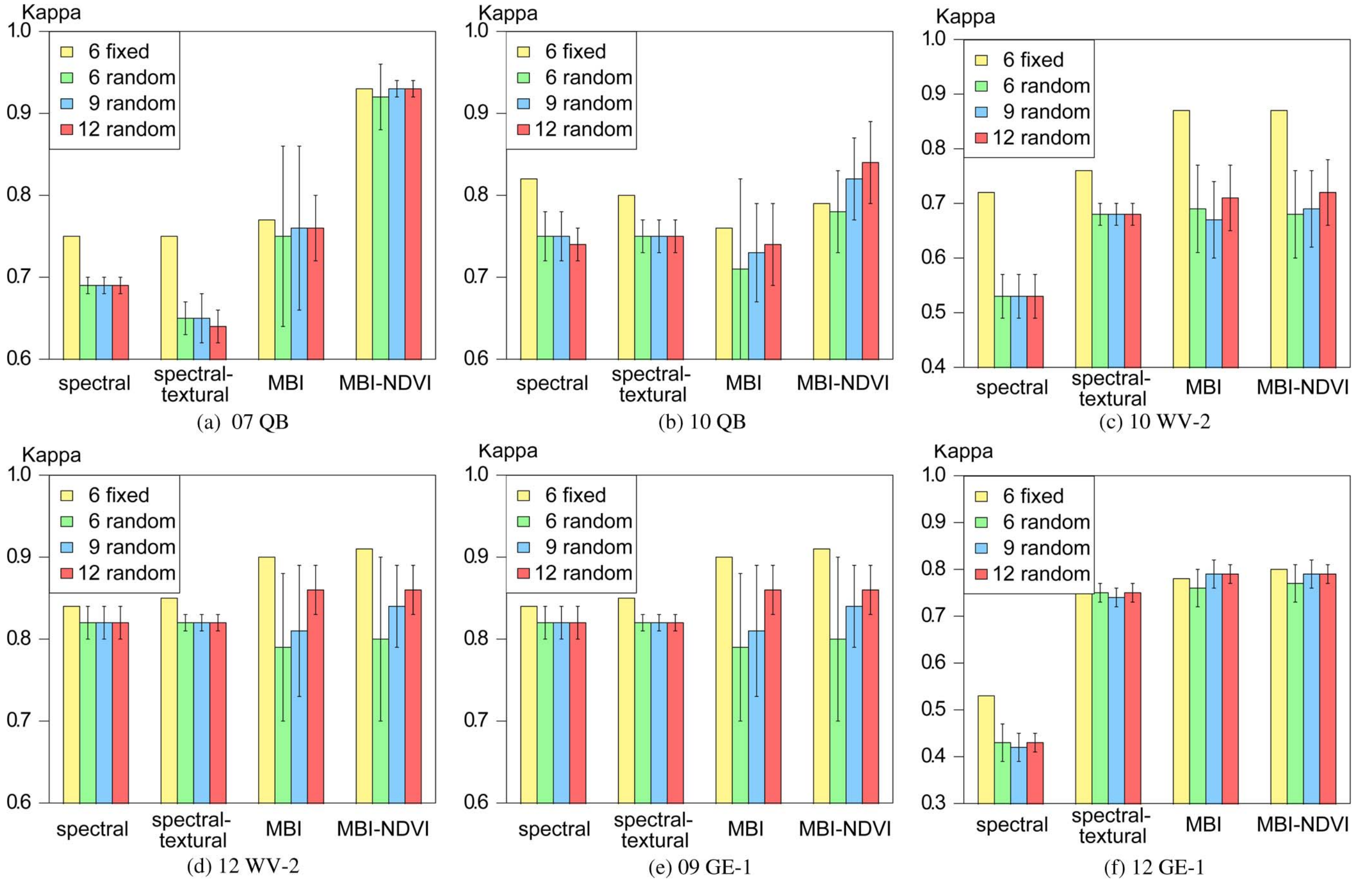
(d) 130 m (e) 140 m (f) Ground truth

Fig. 10. Classification maps of the 2009 Wuhan GeoEye-1 image under different scene sizes.

MBI acted as a fundamental feature due to the important role of buildings in urban settlements, which is verified by the fact that the individual use of MBI with RF as the classifier produced Kappa values with an average of 0.8. NDVI further increased the accuracy significantly. Because of the large diversity of vegetation in UVs, however, the NDVI feature may fail to improve results, e.g., 2009 GeoEye-1 image. In this case, a robust classifier such as RF would be helpful in eliminating the impact of noise in the feature. Moreover, the power of semantic indexes MBI and NDVI is also exhibited in the pixel-based classification of high-resolution images [39].

On the other hand, only a few results produced by conventional models, mostly the BOVW model, were comparable to the index-based results. Although low-level features reliably reflect color or textural characteristics of local patches, they cannot describe and distinguish land covers (e.g., buildings and vegetation) as accurately as semantic indexes (e.g., MBI and NDVI) because different objects often have similar characteristics in high-resolution images. Fortunately, the combination of multiple features could improve the performance to some extent (see Table III). However, the real problem, which should be chiefly responsible for the vulnerable results, is we can hardly adapt these general-purpose models, i.e., BOVW and sLDA, to a specific task because the features they depend on do not contain semantic information we can understand. Thus, we cannot make full use of the prior knowledge. For example, we could not encode the spatial relationship of buildings, which would be useful for reducing the high commissions (see Fig. 8), in the BOVW model since we had no idea about which words are related to buildings.

Multitemporal classification further demonstrates the advantage of high-level features over low-level features in an



(a) 07 QB (b) 10 QB (c) 10 WV-2

(d) 12 WV-2 (e) 09 GE-1 (f) 12 GE-1

Fig. 11. Best average accuracy together with the standard deviation produced by four representations with different numbers of random training samples.

TABLE VII
MULTITEMPORAL CLASSIFICATION ACCURACIES. FIRST ROW INDICATES THE SOURCE OF TRAINING SAMPLES

		2003 QB			2010 WV-2	2009 GE-1
Representation	classifier	2005 QB	2007 QB	2010 QB	2012 WV-2	2012 GE-1
spectral	SVM	0.19±0.08	0.76±0.01	0.75±0.01	0.54±0.02	0.57±0.13
spectral-textural	SVM	0.60±0.10	0.75±0.03	0.75±0.02	0.66±0.04	0.62±0.06
MBI	RF	0.86	0.88	0.77	0.87	0.75
MBI-NDVI	RF	**0.88**	**0.93**	**0.80**	**0.88**	**0.80**

important aspect, i.e., transferability, because the transferability of samples and models largely relies on that of features. By our definition, high-level features intrinsically imply transferability since they refer to things at the semantic level. They usually are low dimensional and understandable, such as MBI and NDVI. Although the capacity of high-level features is often limited by their generalization, the indexes used in this paper have no such problems, where MBI has been tested in original study [27] as well as our experiments (see Fig. 4), and NDVI is formulated based on the biophysical characteristics. By comparison, low-level features are usually high dimensional and image dependent. Because of the inevitable differences between multitemporal images caused by the illumination, sensor angle, and shadows (see Fig. 3), low-level features are subject to these nonsemantic changes, leading to an unpredictable impact on results and weakening the transferability of samples and models. Table VII shows that the textural feature was more robust than the spectral feature to these image differences.

In addition to the feature and the model, the classifier is another important part of the proposed method and also determines the accuracy. As two popular classifiers, SVM and RF performed comparably, but RF was more robust than SVM to the noisy Gabor or NDVI feature. Compared with SVM and RF, the performance of sLDA was poor. By nature, sLDA/LDA is a generative probabilistic model learning the topics of scenes, and a drawback of generative models is they need an assumption about the data distribution. When the model does not fit the data well, the accuracy is usually lower than that of discriminative classifiers such as SVM or RF.

Finally, the scene-based method may be faced with the problem of mixed scenes when processing large images of complicated urban landscapes. Many studies [19], [22], [40] just ignore the scenes mixed with multiple classes when performing the scene-based annotation or classification of large images, because these scenes are beyond any model assuming the label of a scene is unique. In this paper, the problem of mixed scenes

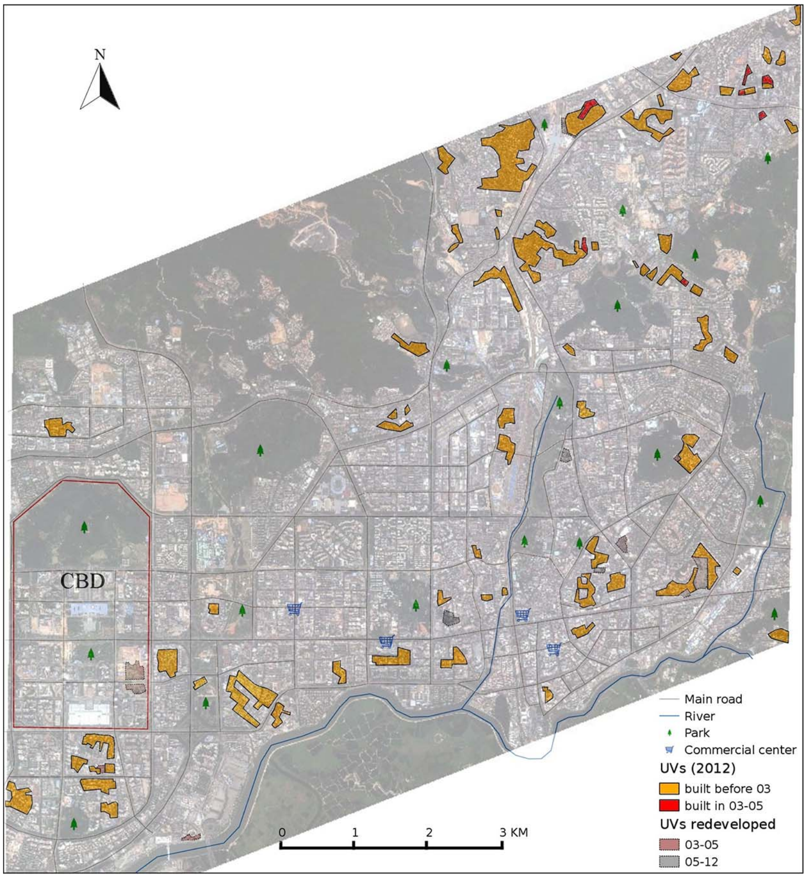


Fig. 12. UV map of study area in Shenzhen.

is actually minor because there are only two classes and such scenes are few. On the other hand, whether a mixed scene is classified as the UV is determined by the proportion of UVs in the scene. Because the index-based and the BOVW models both use histograms as the final representation and histograms reflect proportions of different classes in the scene, scenes with similar class proportions have similar final representations. Since the positive training samples are totally occupied by UVs, the possibility that a mixed scene is classified as the UV decreases with the proportion of the UV in the scene decreasing, which explains why the interior of UVs tend to be lighter, whereas boundaries of UVs and small UVs tend to be darker (see Fig. 8).

IV. Spatiotemporal Analysis of UVs

A practical analysis of UVs requires accurate maps. Because the classification maps produced by the MBI-NDVI approach are close to the ground truth (see Fig. 8), we just need to make some minor modifications to them, such as the removal of false alarms and the refinement of boundaries. Then, the postprocessing maps of UVs over the recent years in Shenzhen and Wuhan are presented in Figs. 12 and 13. Layers of road networks, water bodies, and some facilities in 2012 are also overlaid on the maps. The source of these layers includes the acquired satellite images listed in Section II and online maps (e.g., Google Earth and OpenStreetMap). These information layers are used as ancillary data for the subsequent analysis of UVs, and they have been validated by fieldwork. All figures and statistics in the following analysis are derived from the classification maps and the extracted feature images, and they can be used for urban planning and policy making. It should be noted that an administrative UV is often divided into several disjoint parts in recent decades by the urban development (e.g., major roads shown in Figs. 12 and 13) and the redevelopment led by the government. Because of the weak relations between these parts, we call every isolated region in the maps as a UV for the sake of simplicity in the following analysis.

UVs are a mirror of China's urbanization, and researchers from various fields have investigated the relationships between UVs and urbanization, migration, and housing market from the social or economic perspective [41]–[45]. Nonetheless, the spatiotemporal data of UVs used in these studies all come from the fieldwork, which usually spends much time and restricts the research to several local areas or a single city. In fact, a systematic geographic analysis of UVs, which would be useful for both researchers and policymakers, has been lacking. In Section III, we have shown the ability of remote sensing data to detect UVs in an objective and large-scale manner. Now, using the maps and semantic indexes (i.e., MBI and NDVI), we focus on the following questions:

1) How UVs distribute spatially and temporally?
2) What parameters characterize UVs?
3) What relationships between UVs and other geographic features could be found?
4) Are there similarities and differences between UVs across cities?

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

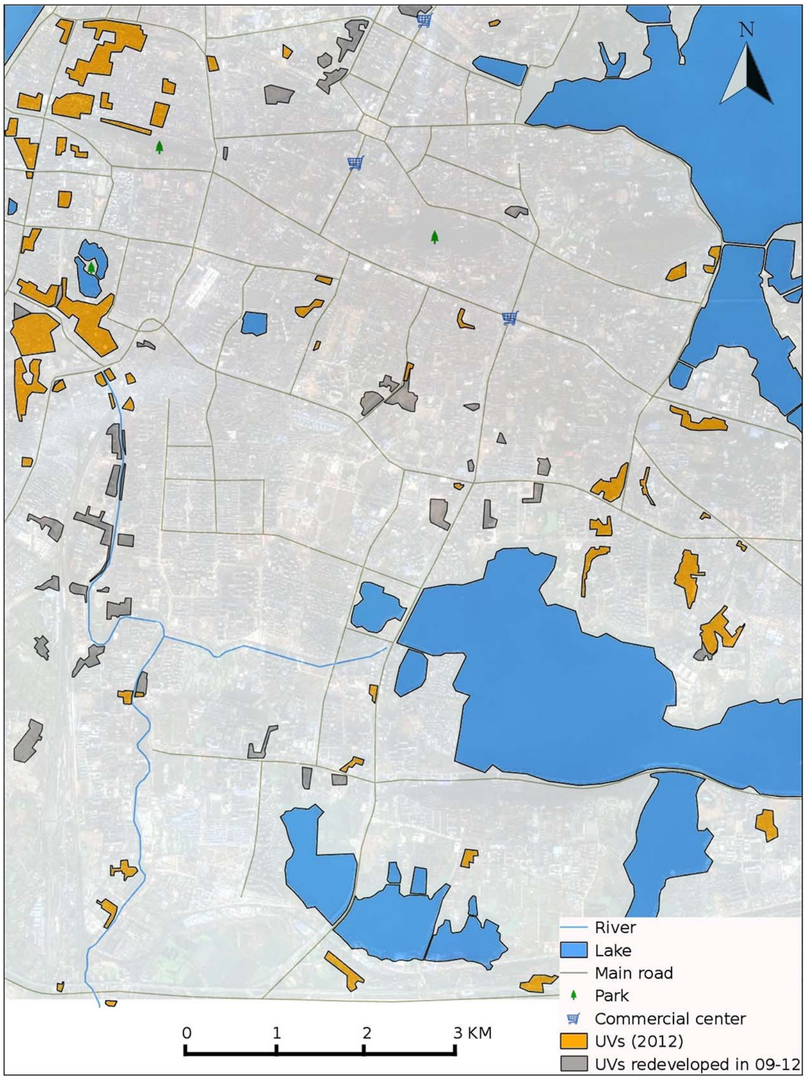


Fig. 13. UV map of study area in Wuhan.

A. Spatiotemporal Distribution

The total area of UVs in the study area of Shenzhen decreases from 348.6 ha (2003) to 328.6 ha (2012). Particularly, there were new UVs built in 2003–2005, which are in the outermost urban areas and far from the CBD (see top right of Fig. 12). At the same time, some UVs near the CBD were demolished. During this period, in fact, the total area of UVs increases by 1.6%. Since 2005, no UVs have been built in the study area, and 7.8% of the UVs have been demolished, including UVs located in the CBD and UVs at the bottom right of Fig. 12 where the road network is dense. In the study area of Wuhan, up to 33.2% of UVs have been demolished in 2009–2012, with the total area decreasing from 332.3 to 222.1 ha. No UVs were built in this period. Fig. 13 shows that the demolished UVs could be divided into three parts: UVs beside the left river, UVs in the top center where a new commercial center is built later, and UVs under the commercial area near the center.

In both cities, there are considerable UVs in 2012, orange regions in Figs. 12 and 13, although the proportion of UVs in the whole study area, with respect to which Shenzhen (3.6% in 2012) is higher than Wuhan (2.4% in 2012), is still low compared with the built-up area. UVs in both cities do not have a significant spatial pattern, but they tend to gather together. Fig. 12 shows UVs mainly locate in the top-right area and the elongated zone along the road beside the river, and Fig. 13 shows UVs largely distribute in the top left area and the right area between two lakes. Clearly, the aggregation of UVs is not spontaneous but is the result of demolition and redevelopment led by the government in recent years.

B. Physical and Geometrical Characteristics

In general, UVs consist of three major land covers: buildings, vegetation, and open spaces. Their coverage is a valuable reference for urban planning. With NDVI indicating vegetation, we can easily compute the vegetation coverage ratio (VCR) of a UV. Nonetheless, the building coverage ratio could not be estimated because the nonorthographic view together with the limited resolution (2/2.4 m) makes it impracticable to accurately distinguish buildings and narrow open spaces (e.g., roads) between buildings. At the same time, it should be noted that buildings in UVs are usually small and disjoint because

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

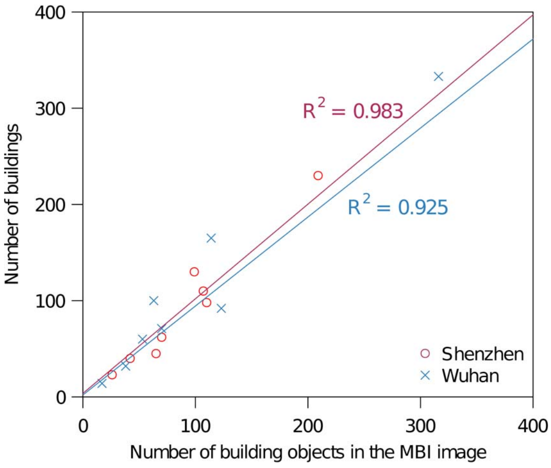


Fig. 14. Scatter plot of the number of buildings in randomly chosen UVs and the number of building objects in corresponding MBI feature images. The upper and the lower lines are fitted regression lines for Shenzhen and Wuhan, respectively.

they are built by individuals and small constructions can make full use of the land. Then, more buildings mean higher building density. Therefore, we used the number of buildings per hectare together with VCR to characterize UVs physically.

Estimating the number of buildings in urban environment via building detection using the optical imagery is challenging [46], [47], and reliable building detection usually needs multisource data such as LiDAR data [48]. At the block level, however, it is reasonable to hypothesize a linear relationship between the number of buildings in a UV and the number of building objects in the corresponding MBI feature image (see Fig. 4), although MBI is not absolutely reliable to locate each single building. To assess the hypothesis, we randomly chose eight UVs in each study area, counted the buildings in these UVs according to images with a higher spatial resolution from Google Earth, and fitted a linear regression model (see Fig. 14). The resulting R^2 were both close to 1, implying a strong linear correlation. The coarse resolution makes MBI incapable of indicating small buildings, resulting in the fitted lines with a slope less than 1. Finally, the numbers of buildings in other UVs were estimated according to the regression equations.

We used MBI and NDVI extracted from the 2012 Shenzhen WorldView-2 and 2009 Wuhan GeoEye-1 images to calculate the two physical parameters, and the results are visualized in Fig. 15. We first conducted a spatial autocorrelation analysis on the two parameters with distances between UVs as the spatial weights. The Moran's *I* of building density and VCR in Shenzhen is −0.0920 and 0.0631, respectively, implying weak spatial dependence. In Wuhan, VCR (0.4761) has a significant positive autocorrelation than the building density (−0.0008). There is also a negative correlation between VCR and the building density in Wuhan (−0.1648) compared with Shenzhen (0.0501). On the other hand, variations of UVs across cities are significantly revealed. Most UVs of Shenzhen have 20–30 buildings per hectare, which is higher than the average density in Wuhan, i.e., 15–25 buildings per hectare [see Fig. 15(a) and (b)]. Shenzhen's UVs also have a higher VCR. 43.5% of UVs in Shenzhen have a VCR larger than 0.05, whereas the proportion of UVs with such a VCR in Wuhan is only 25.3% [see Fig. 15(c) and (d)]. With high building density and high green coverage, there leaves little space for open ground and roads in Shenzhen's UVs where the gap between adjacent buildings is usually just 1 m or less. Considering that the numbers of floors of buildings in Shenzhen are much higher than that in Wuhan, Shenzhen's UVs undoubtedly have a higher floor area ratio and population density than Wuhan's.

In addition to physical parameters, we also computed geometrical parameters (e.g., area and perimeter) of UVs using the 2012 Shenzhen and 2009 Wuhan maps. The area of UVs in Shenzhen and Wuhan ranges from 0.2 to 42 ha and from 0.4 to 30.7 ha, respectively, with an average of 4.8 and 3.6 ha. Compared with planned urban areas, UVs tend to have more complex shapes. To measure the shape complexity quantitatively, we calculated the shape index of UVs. The shape index [49] of a square is 1, and it increases as the shape becomes more irregular. Fig. 16 shows relationships between the area, the shape index, and the building density and reveals the following findings. 1) Large bubbles tend to be at the bottom, i.e., the building density of UVs with a larger area tends to be lower. 2) Most large bubbles are at the right, i.e., large UVs generally have complex shapes. 3) UVs of Shenzhen and Wuhan have a similar distribution of the shape complexity. In both cities, for instance, most of UVs are in the interval 1.0–1.5, and the number of UVs falls as the shape index increases. 4) Moreover, the blue bubbles look below the red ones, which intuitively illustrates the higher density of UVs in Shenzhen.

C. Relationships With Other Facilities

Although most UVs have access to tap water and electricity that are often absent in informal settlements of other countries, many important public services (e.g., sewerage and sanitation facilities) are still lacking. Since UVs are not isolated from but indeed enclosed in urban areas, the accessibility to the public transport system or hospitals, which can be measured by the linear distance between UVs and these facilities, becomes an important factor to be considered for not only the daily life of inhabitants but also the housing market and the further redevelopment toward formal residential areas.

Based on the 2012 Shenzhen and Wuhan maps of UVs, we computed the linear distance from the central location of UVs to the nearest main road, park, and commercial center. The relationships are shown in Fig. 17. Because there are many changes to Wuhan's UVs in 2009–2012, we also took UVs demolished in this period into account for a temporal comparison. Obviously, most UVs are within 1 km from the main road, suggesting the convenient transportation for workers lived in UVs. By contrast, the distance of UVs from the park and the commercial center in both cities has a large variation except that almost all UVs in Shenzhen are within 2 km from the park [see Fig. 17(a)] because Shenzhen is one of the cities with the highest green coverage in China. Furthermore, up to 15 UVs are within the 2-km circle of the park and the commercial center in Shenzhen. By comparison, there were only five such UVs in Wuhan in 2009, and three of them were demolished by 2012, reflecting the strict attitude of the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

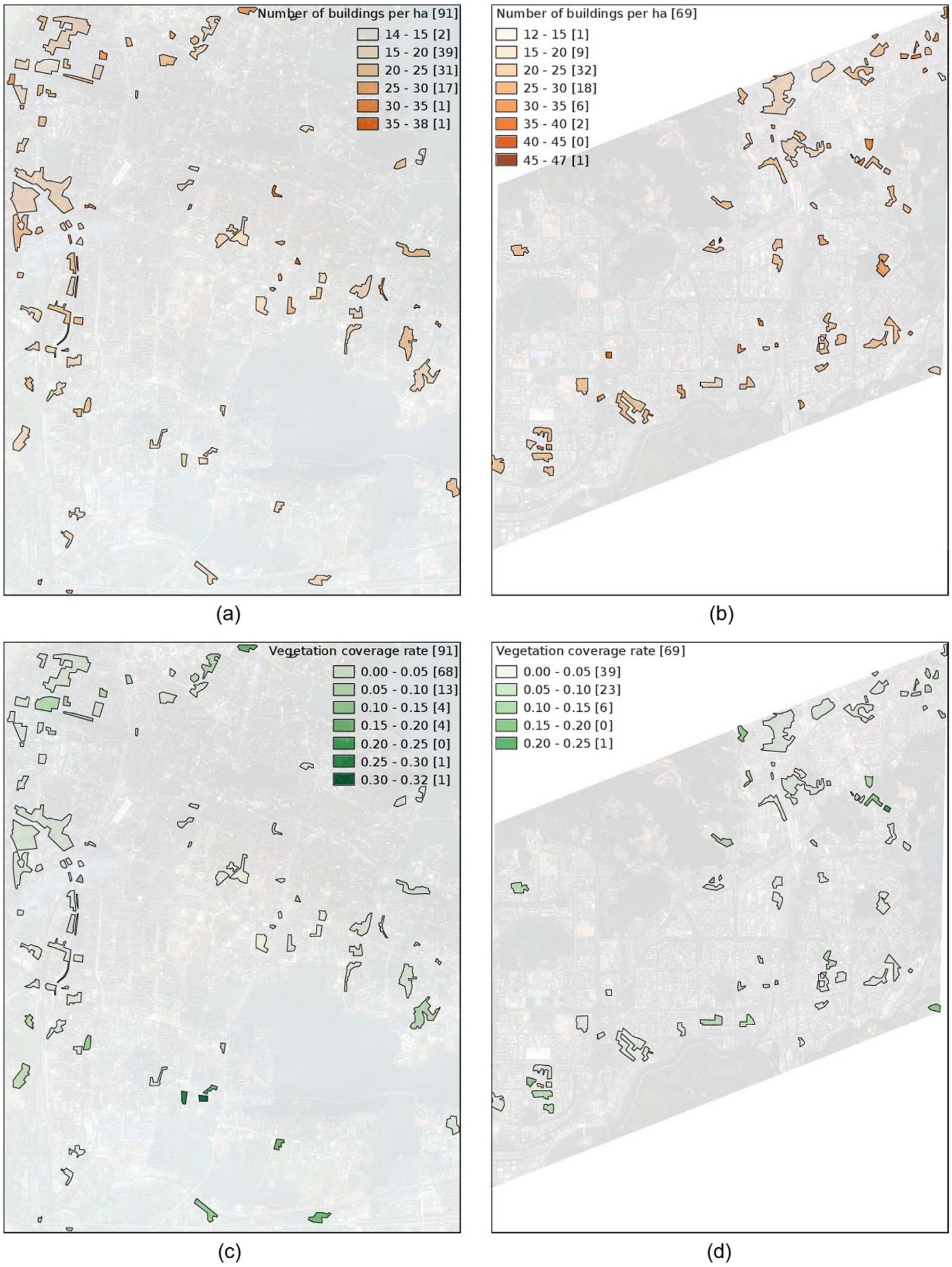


Fig. 15. Building density maps of UVs in (a) Wuhan and (b) Shenzhen and vegetation coverage rate maps of UVs in (c) Wuhan and (d) Shenzhen.

Wuhan government toward UVs. Another important observation is that there are more UVs in the 2–4-km zone than other intervals in both cities, to some degree suggesting the tradeoff between the urban management and the convenient living of UV inhabitants.

D. Discussions

The analysis exhibits the variations of UVs within and across cities quantitatively. The variations root in the different socioeconomic factors. At the UV level, for example, the higher building density of UVs in Shenzhen than Wuhan could be attributed to the opening-up policy and the advantageous location, with the help of which Shenzhen attracts more people in the last 30 years than any other city including Wuhan. Within cities, the variations of UVs are usually irregular because the development of UVs is influenced by many factors [3], [38]. Taubenböck and Kraff [50] analyzed the physical parameters of slums in Mumbai, India, and the results also reveal the heterogeneity of slums within a city. For instance, there are 11, 21, and 26 buildings per hectare in three sample slums of Mumbai, which are comparable to the density of UVs. At the city level, viewing UVs as points regardless of their characteristics, we find many UVs in Shenzhen are very close to the prosperous areas (e.g., the CBD and commercial centers), and such UVs are rare in Wuhan. It is mainly because Shenzhen

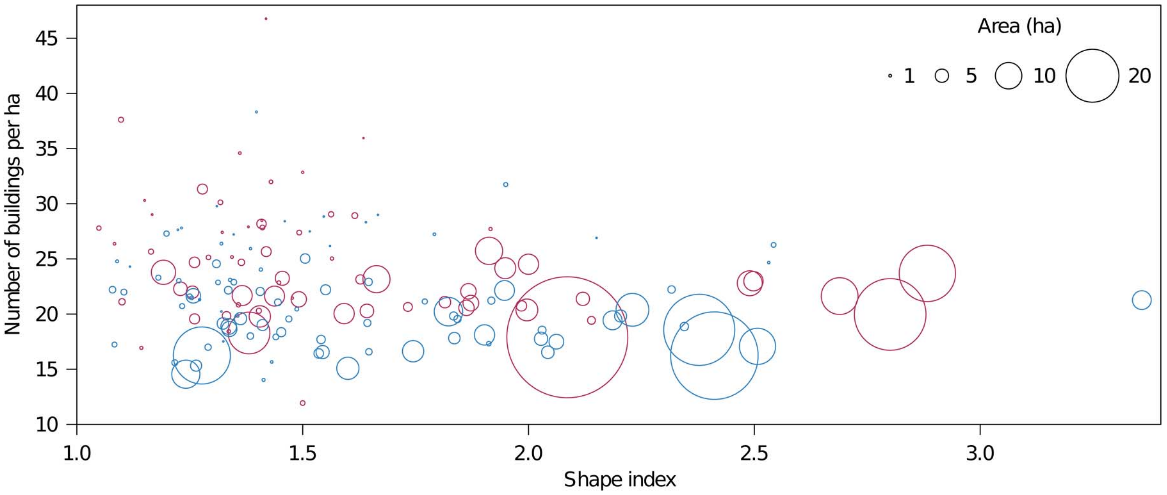


Fig. 16. Relationships between the area, the shape index, and the building density of UVs in Shenzhen (red) and Wuhan (blue). Each bubble represents a UV, whose size and coordinates indicate its area, shape index, and building density.

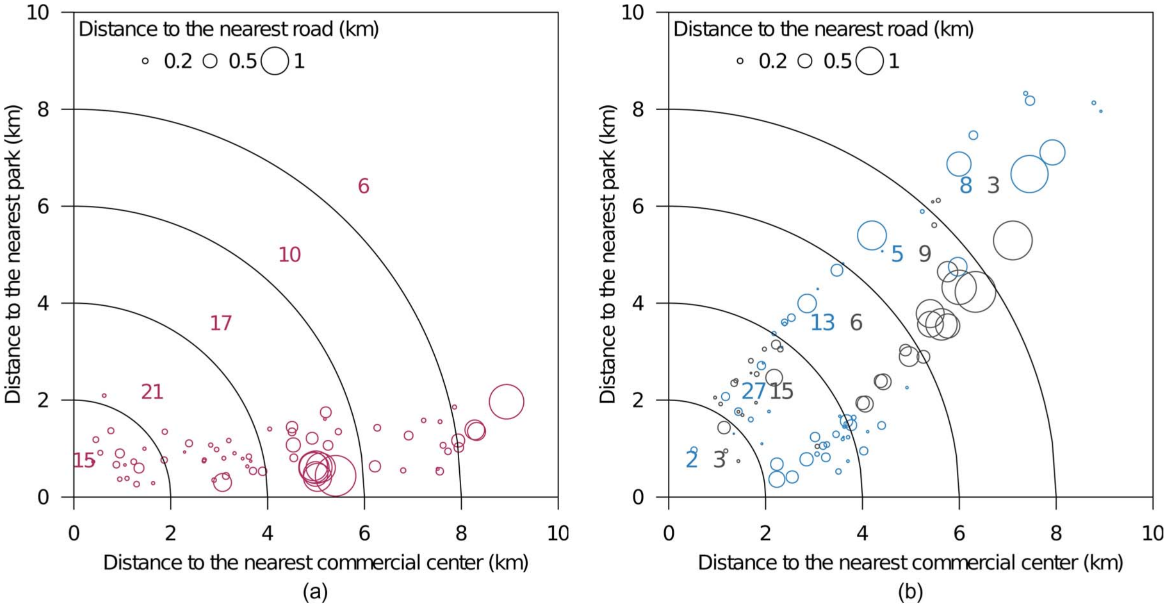


Fig. 17. Relationships between the distances to the nearest public facilities of UVs in (a) Shenzhen and (b) Wuhan. Blue and gray bubbles in (b) denote UVs in 2012 and UVs redeveloped in 2009–2012, respectively.

is a young city, and its urban areas and UVs are developed at the same time. Thus, urban areas and UVs are intermingled with each other. In Wuhan, the downtown areas actually formed before the fast urban expansion of recent decades, and many UVs close to these areas have been also demolished recently (see Fig. 13).

Despite the large diversity of UVs, the similarities shared by UVs (e.g., overcrowded buildings) overwhelm the variations in the sense of distinguishing between UVs and other urban areas. Indeed, it is according to these similarities that we proposed the index-based model that yielded satisfactory results. Moreover, due to these common characteristics and the resulting poor living condition, the change of UVs to urbanized areas would be inevitable and irreversible. As shown in Figs. 12 and 13, UVs in the core urban areas have been decreasing recently. Hao *et al.* [37] also observed the slight decline of UVs during 2004–2009 in Shenzhen. Such changes should be credited to the redevelopment programs made by the government. In 2005, Shenzhen proposed The Master Plan of Urban Village Redevelopment, aiming to redevelop 20% of UVs inside the main city areas by 2010. In 2009, Wuhan planned to redevelop all UVs within the second ring road by the end of 2011. In contrast to the incomplete statistics of UVs in the study areas (see Section IV-A), however, the aim set by the government was just partly fulfilled. From the socio-economic perspective, many studies have discussed the factors influencing the redevelopment, e.g., the high cost, the access to employment, and the housing demand of huge inhabitants [3], [38]. In fact, the redevelopment of informal settlements elsewhere also faces considerable hindrances [50].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Overall, due to the huge demand of migrant workers and the poor for low-cost housing, UVs will exist in a long term, particularly in the urban fringe. Moreover, UVs have been evolving to adapt to the urban development and the government administration, and some studies [36] also argue that the value of UVs in urbanization should be recognized, and the redevelopment should be carefully considered. For the future management of UVs, we could see the great potential of remotely sensed data in urban monitoring (e.g., an up-to-date and large-scale detection at the UV level) due to high time and spatial resolution. Beyond the detection, we can also use remote sensing data to characterize UVs by computing VCR or estimating the building density as shown in Section IV-B, where semantic features (e.g., MBI and NDVI) would be more helpful than low-level features.

V. Conclusion

The main purpose of this paper was to analyze the spatiotemporal patterns of UVs. First, we proposed three scene-based methods, including a novel index-based approach, for detecting UVs. The index-based approach models UVs as the proportions and the spatial configuration of building and vegetation objects with MBI and NDVI and can integrate the prior knowledge of UVs easily. Moreover, the multitemporal classification was conducted to evaluate the transferability of these methods. Second, the spatiotemporal changes of UVs were systematically analyzed based on the detection results. We demonstrated the variations of UVs within and across cities and revealed their dependence on the socio-economic factors. We also summarized the spatiotemporal changes of UVs over the recent years and found the decline of UVs in two cities.

To our knowledge, this is the first study of UVs using remotely sensed data, and it would be valuable for the future redevelopment and management of UVs in Shenzhen, Wuhan, and other cities of China. Moreover, this paper exhibits the ability of the scene-based approach for the detection of complex urban structures. In particular, the results indicate the great potential of high-level features, e.g., MBI and NDVI, for modeling and characterizing complex scenes. High-level features will also play an important role in transferable researches.

Urbanization in China is still rapidly increasing beyond the mega cities, and new UVs may emerge in the urban fringe along with the urban expansion. Given the diversity of UVs across cities shown in this paper, UVs in cities with different sizes should be considered in further research. In addition, due to the limited ability of the optical imagery in urban mapping (e.g., extraction of building heights and sizes), future research could take into account multisource data such as LiDAR and GIS data.

Acknowledgment

The authors would like to thank the anonymous reviewers for their insightful suggestions, which significantly improved the quality of this paper.

References

[1] UN-Habitat, Slums of the world: The face of urban poverty in the new millennium? 2003.

[2] H. Chung, "Building an image of villages-in-the-city: A clarification of China's distinct urban spaces," *Int. J. Urban Reg. Res.*, vol. 34, no. 2, pp. 421–437, Jun. 2010.

[3] P. Hao, P. Hooimeijer, R. Sliuzas, and S. Geertman, "What drives the spatial development of urban villages in China?" *Urban Studies*, vol. 50, no. 16, pp. 3394–3411, Dec. 2013.

[4] J. Shen, "Rural development and rural to urban migration in China 1978–1990," *Geoforum*, vol. 26, no. 4, pp. 395–409, Nov. 1995.

[5] X. Yang, "Determinants of migration intentions in Hubei province, China: Individual versus family migration," *Environ. Plan. A*, vol. 32, no. 5, pp. 769–788, Nov. 2000.

[6] H. Chung, "The planning of 'villages-in-the-city' in Shenzhen, China: The significance of the new state-led approach," *Int. Plan. Stud.*, vol. 14, no. 3, pp. 253–273, 2009.

[7] H. Chung and S.-H. Zhou, "Planning for plural groups? villages-in-the-city redevelopment in Guangzhou city, China," *Int. Plan. Stud.*, vol. 16, no. 4, pp. 333–353, 2011.

[8] H. Taubenböck *et al.*, "Delineation of central business districts in mega city regions using remotely sensed data," *Remote Sens. Environ.*, vol. 136, pp. 386–401, Sep. 2013.

[9] R. Mathieu, C. Freeman, and J. Aryal, "Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery," *Landsc. Urban Plan.*, vol. 81, no. 3, pp. 179–192, 2007.

[10] M. Molinier, J. Laaksonen, and T. Hame, "Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 861–874, Apr. 2007.

[11] P. Hofmann, J. Strobl, T. Blaschke, and H. Kux, "Detecting informal settlements from QuickBird data in Rio de Janeiro using an object based approach," in *Object-Based Image Analysis*, T. Blaschke, S. Lang, and G. Hay, Eds. New York, NY, USA: Springer-Verlag, 2008, pp. 531–553.

[12] P. Hofmann, "Detecting informal settlements from IKONOS image data using methods of object oriented image analysis—An example from Cape Town (South Africa)," in *Remote Sensing of Urban Areas*, C. Jügens, Ed., Regensburger, Germany: Regensburger Geographische Schriften, 2001, pp. 41–42.

[13] P. Hurskainen and P. Pellikka, "Change detection of informal settlements using multi-temporal aerial photographs—The case of Voi, SE-Kenya," presented at the 5th African Assoc. Remote Sens. Environ. Conf., Nairobi, Kenya, 2004.

[14] S. Niebergall, A. Loew, and W. Mauser, "Object-oriented analysis of very high-resolution QuickBird data for mega city research in Delhi/India," presented at the *Urban Remote Sens. Joint Event.*, Apr. 2007.

[15] H. Rhinane, A. Hilali, A. Berrada, and M. Hakdaoui, "Detecting slums from SPOT data in Casablanca Morocco using an object based approach," *J. Geogr. Inf. Syst.*, vol. 3, no. 3, pp. 217–224, 2011.

[16] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification via pLSA," in *Computer Vision—ECCV*, A. Leonardis, H. Bischof, and A. Pinz, Eds., New York, NY, USA: Springer-Verlag, 2006, vol. 3954, pp. 517–530.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2006, pp. 2169–2178.

[18] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inform. Syst.*, 2010, pp. 270–279.

[19] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.

[20] A. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 439–451, Jan. 2014.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[22] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.

[23] D. Bratasanu, I. Nedelcu, and M. Datcu, "Bridging the semantic gap for satellite image annotation and automatic mapping applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 193–204, Mar. 2011.

[24] R. R. Vatsavai, A. Cheriyadat, and S. Gleason, "Unsupervised semantic labeling framework for identification of complex facilities in high-resolution remote sensing images," in *Proc. IEEE ICDMW*, Dec. 2010, pp. 273–280.

[25] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Red Hook, NY, USA, Curran Associates, Inc., 2008, pp. 121–128.

[26] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

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

[28] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation svm classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, Apr. 2009.

[29] P. Hofmann, T. Blaschke, and J. Strobl, "Quantifying the robustness of fuzzy rule sets in object-based image analysis," *Int. J. Remote Sens.*, vol. 32, no. 22, pp. 7359–7381, 2011.

[30] T. Novack, H. Kux, R. Q. Feitosa, and G. A. O. P. Costa, "A knowledge-based, transferable approach for block-based urban land-use classification," *Int. J. Remote Sens.*, vol. 35, no. 13, pp. 4739–4757, 2014.

[31] G. O'Hare and M. Barke, "The favelas of Rio de Janeiro: A temporal and spatial analysis," *GeoJournal*, vol. 56, no. 3, pp. 225–240, 2002.

[32] H. Taubenböck, M. Wegmann, A. Roth, H. Mehl, and S. Dech, "Urbanization in India—Spatiotemporal analysis using remote sensing data," *Comput., Environ. Urban Syst.*, vol. 33, no. 3, pp. 179–188, May 2009.

[33] P. Griffiths, P. Hostert, O. Gruebner, and S. van der Linden, "Mapping megacity growth with multi-sensor data," *Remote Sens. Environ.*, vol. 114, no. 2, pp. 426–439, Feb. 2010.

[34] O. Kit and M. Lüdeke, "Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 130–137, Sep. 2013.

[35] K. K. Owen and D. W. Wong, "An approach to differentiate informal settlements using spectral, texture, geomorphology and road accessibility metrics," *Appl. Geogr.*, vol. 38, pp. 107–118, Mar. 2013.

[36] Y. P. Wang, Y. Wang, and J. Wu, "Urbanization and informal development in China: Urban villages in Shenzhen," *Int. J. Urban Reg. Res.*, vol. 33, no. 4, pp. 957–973, Dec. 2009.

[37] P. Hao, S. Geertman, P. Hooimeijer, and R. Sliuzas, "Spatial analyses of the urban village development process in Shenzhen, China," *Int. J. Urban Reg. Res.*, vol. 37, no. 6, pp. 2177–2197, Nov. 2013.

[38] J. Zacharias and Y. Tang, "Restructuring and repositioning Shenzhen, China's new mega city," *Prog. Plan.*, vol. 73, no. 4, pp. 209–249, May 2010.

[39] X. Huang, Q. Lu, and L. Zhang, "A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 36–48, Apr. 2014.

[40] W. Luo, H. Li, G. Liu, and L. Zeng, "Semantic annotation of satellite images using author-genre-topic model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1356–1368, Feb. 2014.

[41] L. Zhang, S. X. B. Zhao, and J. P. Tian, "Self-help in housing and chengzhongcun in China's urbanization," *Int. J. Urban Reg. Res.*, vol. 27, no. 4, pp. 912–937, 2003.

[42] L. Tian, "The chengzhongcun land market in China: Boon or bane?—A perspective on property rights," *Int. J. Urban Reg. Res.*, vol. 32, no. 2, pp. 282–304, Jun. 2008.

[43] Y. Liu, S. He, F. Wu, and C. Webster, "Urban villages under China's rapid urbanization: Unregulated assets and transitional neighbourhoods," *Habitat Int.*, vol. 34, no. 2, pp. 135–144, Apr. 2010.

[44] Y. Lin, B. de Meulder, and S. Wang, "Understanding the 'village in the city' in Guangzhou: Economic integration and development issue and their implications for the urban migrant," *Urban Studies*, vol. 48, no. 16, pp. 3583–3598, Dec. 2011.

[45] B. Zhang, F. Zhang, Y. Gao, C. Li, and F. Zhu, "Identification and spatial differentiation of rural settlements' multifunction," *Trans. Chin. Soc. Agric. Eng.*, vol. 30, no. 12, pp. 216–224, June 2014.

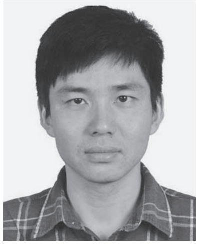
[46] B. Sirmacek and C. Unsalan, "Urban-area and building detection using sift keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1156–1167, Apr. 2009.

[47] C. Ioannidis, C. Psaltis, and C. Potsiou, "Towards a strategy for control of suburban informal buildings through automatic change detection," *Comput., Environ. Urban Syst.*, vol. 33, no. 1, pp. 64–74, Jan. 2009.

[48] K. Zhang, J. Yan, and S.-C. Chen, "Automatic construction of building footprints from airborne LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2523–2533, Sep. 2006.

[49] K. McGarigal and B. J. Marks, "Spatial pattern analysis program for quantifying landscape structure," USDA Forest Service Gen., Washington, DC, USA, Tech. Rep. PNW-GTR-351, 1995.

[50] H. Taubenböck and N. Kraff, "The physical face of slums: A structural comparison of slums in Mumbai, India, based on remotely sensed data," *J. Hous. Built Environ.*, vol. 29, no. 1, pp. 15–38, Mar. 2014.



Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009, working with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS).

He is currently a Full Professor with LIESMARS. He has published more than 50 peer-reviewed articles in international journals. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Dr. Huang was a recipient of the Top-Ten Academic Star of Wuhan University in 2009, the Boeing Award for Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing in 2010, the New Century Excellent Talents in University from the Ministry of Education of China in 2011, and the National Excellent Doctoral Dissertation Award of China in 2012. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as a Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the winner of the IEEE GRSS 2014 Data Fusion Contest. Since 2014, he has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Hui Liu received the B.S. degree in mathematics from Wuhan University, Wuhan, China, in 2012, where he is currently working toward the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

His research interests include image processing, object recognition, and remote sensing applications.

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 Chinese Academy of Sciences, Xian, China, in 1988; and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a 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. He has more than 300 research papers. He is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governor) of the China National Committee of the International Geosphere–Biosphere Programme, and an Executive Member of the China Society of Image and Graphics. He regularly serves as a Cochair for the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and Geoinformatics symposiums. He also serves as an Associate Editor for the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, the *Journal of Remote Sensing*, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.