A Novel Unsupervised Sample Collection Method for Urban Land-Cover Mapping Using Landsat Imagery

Jiayi Li, *Member, IEEE*, Xin Huang, *Senior Member, IEEE*, Ting Hu, *Student Member, IEEE*, Xiuping Jia, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*— Land-cover mapping over urban areas using Landsat imagery has attracted considerable attention in recent years as it can promptly and accurately reflect the biophysical composition status of the urban landscape and allow further applications such as urban planning and risk management. However, due to the large diversity across different urban landscapes, adequate training sample collection for urban area mapping is both challenging and time-consuming. In this paper, we propose a novel unsupervised sample collection method for mapping urban areas using Landsat imagery. Specifically, the idea is to select reliable, representative, and diverse training samples from the images in a two-stage and iterative manner, based on a set of spectral indices (vegetation, impervious surface, soil, water). To validate the effectiveness and robustness of the proposed method, a synthetic data set was designed and a series of Landsat images over 39 representative cities from different biomes across the world was employed. The effectiveness of the proposed algorithm was quantitatively validated by assessing the quality of the automatically collected samples and the accuracy of the mapping results. In terms of the mapping performance, the proposed automatic approach can achieve a comparable mapping accuracy to supervised classification with manually collected samples. On the basis of the freely accessed Landsat data, the proposed approach demonstrates a promising potential for automatic large-scale (i.e., global) mapping over urban areas.**

***Index Terms*— Classification, land-cover mapping, Landsat, spectral index, unsupervised learning, urban.**

Manuscript received May 27, 2018; revised September 28, 2018 and November 28, 2018; accepted December 16, 2018. Date of publication February 27, 2019; date of current version May 28, 2019. The research was supported by the National Natural Science Foundation of China under Grants 41701382 and 41771360, the National Program for Support of Top-Notch Young Professionals, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA029 and 2017CFB188, and the National Key R&D Program of China under Grant 2016YFB0501403. *(Corresponding author: Xin Huang.)*

J. Li and T. Hu are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China.

X. Huang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China, and also 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).

X. Jia is with the School of Engineering and Information Technology, The University of New South Wales, Canberra, ACT 2600, Australia.

J. A. Benediktsson is with the Department of Electrical and Computer Engineering, University of Iceland, 107 Reykjavik, Iceland.

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.2018.2889109

I. Introduction

As a reflection of human development, urbanization is taking place around the world at an unprecedented rate. With the increasing pressure of economic and population growth, the conversion from natural to anthropogenic landscapes has sharply increased, and the biophysical composition of urban areas has dramatically changed [1]. Although the rapid pace of recent urbanization has brought significant social and economic benefits (e.g., improved quality of life and economic prosperity [2]), it has also led to a number of negative effects (e.g., natural habitat loss [3] and urban heat islands [4]). Therefore, in order to provide reliable data for urban planning and management, it is imperative to automatically and accurately map the distribution and characteristics of urban areas.

Remote sensing imagery with various resolutions and flexible acquisition modes can provide us with synoptic views of the earth's surface, which in turn provide us with the potential to measure, analyze, and hence understand urban areas worldwide. The medium spatial resolution data (i.e., 10–100 m) with multispectral bands (i.e., 4–40 bands ranging from near infrared to visible), such as the free-access Landsat imagery, are appropriate data sources, in terms of the spatial detail, area coverage, revisit frequency, availability, and historical archives [5]. A number of methods have been proposed for the land-cover mapping of urban areas using Landsat imagery, which can basically be divided into three categories. The first category is based on spectral indices, e.g., the normalized difference built-up index (NDBI) [6], the normalized difference water index (NDWI) [7], the modified NDWI (MNDWI) [8], the normalized difference vegetation index (NDVI) [9], and the bare soil index (BI) [10]. These indices have the potential to highlight a set of land-cover classes in terms of their physical characteristics. In this kind of method, a threshold is usually used to determine the land-cover class, i.e., a pixel whose index value is larger than a specific threshold is labeled as the associated class, and vice versa. However, these information indices can be subject to a large number of commission or omission errors [5].

The second category is the supervised machine learning methods, which aim at establishing mapping criteria with the spectral/spatial features from the remote sensing

images and the collected training samples [11]. However, the manual collection of adequate training samples is a time-consuming process [5], [12]–[14], and the subjectivity of human intervention further limits the sample diversity [15]. Hence, training sample extraction from the existing products (e.g., the National Land Cover Database (2001–2011) [16]–[18] or the GlobeLand30 data sets (GLC30 for 2000 and 2010) [19]) is a practical alternative approach, especially when dealing with large-scale land-cover mapping. In this case, however, the registration error and date gap between the satellite Landsat images and the existing reference products still requires attention. For instance, it is still possible to extract erroneous samples from the existing products, due to land-cover change or spatial misregistration. In such a case, data cleaning, i.e., deleting the samples that appear in several classes [15], is needed to purify the training samples collected from the existing products. In summary, both the high cost of the manual sampling and the sample uncertainty of the existing products restrict the performance of the machine learning-based mapping methods. In this regard, direct and automatic sample collection from the target Landsat imagery has the potential to eliminate the aforementioned obstacles.

This paper presents a novel automatic sample collection method for land-cover mapping over urban areas using Landsat imagery. In this paper, we focus on the four basic categories in urban areas: (bright and dark) built-up (BU), vegetation (V), water surface (WS), and bare soil (BS). In particular, the category of BU is related to man-made structures, i.e., buildings, streets, impervious surfaces, and urban shadows [5]. The aim of our study was to propose an unsupervised training sample selection method based on a set of spectral indices. In the proposed method, the characteristics of the spectral indices, as well as the property of the label uniqueness for each pixel, are used to collect samples from the Landsat images in an iterative manner. The proposed method, without requiring training samples in advance, initializes the sample set with the pixels with the largest spectral index values, and sequentially collects samples in the order of the index values. In the main body of the proposed method, the sample collection process in each iteration includes an informative sample selection procedure and a label checker. The former procedure is similar to active learning [37], [44], [45], which is a supervised machine-learning paradigm that aims to discover the most suitable instances for labeling. However, as an unsupervised learning method with neither training samples nor human supervisor, the label checker is designed to ensure the label reliability of the augmented samples, based on the property of label uniqueness, i.e., each pixel belongs to only one category.

The rest of this paper is organized as follows. Section II presents the methodology of the automatic sample collection and classification method. Section III describes the experimental setup, including the real Landsat images and the synthetic data set. The results and analysis are reported in Section IV. Finally, our conclusions are given in Section V.

TABLE I
SPECTRAL INDICES CONSIDERED IN THIS PAPER

Index	Principle	Formula
NDVI	Constructed by exploiting the difference in the spectral reflectance of vegetation in the near-infrared and red bands.	$NDVI = \frac{NIR - R}{NIR + R}$
MNDWI	Used to enhance the water feature, based on the principle that water has strong reflectance in the green band, but strong absorption in the shortwave infrared band.	$MNDWI = \frac{G - SWIR}{G + SWIR}$
BI	Designed under the principle that bare soil has a high reflectance in both the shortwave infrared channel and red channel, but low reflectance in the near-infrared and blue channels.	$BI = \frac{(R + SWIR) - (B + NIR)}{(R + SWIR) + (B + NIR)}$
NDBI	Makes use of the characteristic that built-up areas with high albedo have high reflection in the shortwave infrared channel, but low reflection in the near-infrared channel.	$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$
NDWI	Can highlight the land-cover classes with strong reflectance in the green band, but strong absorption in the near-infrared band, including dark built-up and water.	$NDWI = \frac{G - NIR}{G + NIR}$

II. METHODOLOGY

A. Physical Characteristics of Spectral Indices and the Basic Principle

The first basic idea of the proposed sample collection algorithm originates from the physical characteristics of spectral indices, i.e., pixels with a large index value usually have a high confidence of being associated with the related land-cover class. In this paper, NDVI, MNDWI, and BI were used to produce the index images of V, WS, and BS, respectively. Detailed descriptions of each spectral index are provided in Table I. Owing to the high spectral variation caused by different structures, colors, and materials, the BU category is divided into bright BU and dark BU [20]. It is noted that NDBI can indicate bright BU; however, there is no customized spectral index to describe dark BU. To tackle this issue, a **s**ynthetic **d**ark BU **i**mage (SDBI for short) is designed during the sample collection to describe dark BU, based on the principle that NDWI can highlight the dark BU after the water surfaces have been masked out [21]. For instance, the NDBI image in Fig. 1 is transformed into a histogram, where the pixels with higher index values are more likely to be bright BU pixels than those with lower values [from right to left in Fig. 1(b)].

Label uniqueness is another basic principle, i.e., only one label can be assigned to each pixel. It implies that not only the spectral index associated with the potential class, but also those indices related to other classes, can be utilized for labeling a pixel in a synergistic manner. For instance, given a pixel whose NDBI and NDVI values are 0.8 and 0.05, respectively, this conveys the information that this pixel has a high probability of being bright BU and a low probability of being vegetation. In this way, the sample collection for multiple classes can be performed in an interactive manner.

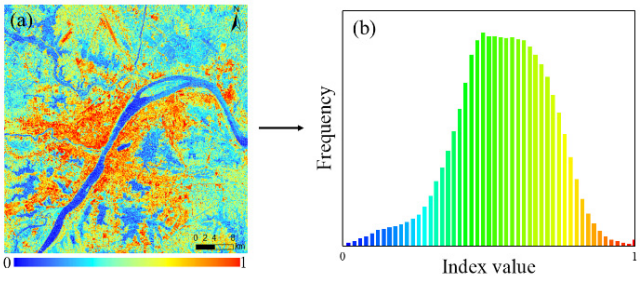


Fig. 1. (a) Spatial distribution and (b) histogram of the NDBI image (Wuhan). The proposed approach is sequentially carried out by selecting the candidate samples from the ones on the right with high values (in red) to the ones on the left with low values (in blue).

Based on the physical characteristics of the spectral indices, the proposed method initializes the sample set with the pixels with the largest spectral index values, and sequentially collects samples in the order of the index values. The collection process includes an informative sample selection procedure to maximize the sample set diversity, and a label consistency constraint to ensure the label reliability of the sample set. In view of this, the proposed sample collection procedure can be formulated as a quadruple (Q, S, T, U). U is the sample pool based on the physical characteristics of the spectral indices, and Q is the query function used to select the most informative unlabeled samples from U. S is the label checker that can classify the informative unlabeled samples of U and add them to the updated training sample set T. The proposed approach is an iterative process, where S interacts with U by labeling the most informative samples selected by the query function Q at each iteration. At the beginning, for each class, the initial training set T is composed of the pixels with the highest index values. After initialization, the query function Q selects a set of samples from the pool U, and the label checker S adds the reliable samples into T and excludes the outliers of T. Please kindly note that both Q and S are constructed on the basis of the current T. The iteration continues until the stopping criteria are met.

B. Candidate Pool, Query Function, and Label Checker Construction

1) Candidate Pool U: For the multiple-class sample collection, each iteration consists of several operations (each for one class). In this paper, the class set is defined as $C = \{c$: bright BU, dark BU, V, WS, and BS$\}$. For the sake of simplifying the description, we hereinafter take class c as an example. For the proposed approach, every index image is linearly normalized to [0, 1]. It is important to keep in mind that: 1) pixels with large index values should have a high probability of being assigned to the associated class and 2) pixels with close index values should have similar likelihoods with respect to the associated class. In view of this, for the ith iteration of class c, pixels whose associated index values in $[1-(i+1)/K, 1-i/K]$ (K is set as 1000 in this paper) are assigned to the candidate pool U of class c, e.g., all points in Fig. 2(b). Each pixel p_{s_j} in U of class c sized N_{uc} can be denoted by a triple $\{\boldsymbol{p}_{s_j}, w_{s_j} = 1, c\}$, $sj \in \{1, 2, \ldots, N_{\text{uc}}\}$, which represents the feature, the reliability (i.e., weight), and the label of pixel p_{s_j}, respectively. For Landsat imagery, the spectral bands (i.e., bands 1–5 and 7 for Landsat 5 and 7 images; bands 1–7 for Landsat 8 images) and five indices (MNDWI, NDVI, NDBI, NDWI, and BI) are considered as the input features, following the suggestions in [38] and [39]. The weight of each candidate pixel is initialized as 1.

2) Query Function Q: The query function Q, which aims to not only extract the most informative samples from U, but also update the reliability of T, is built on the current training sample set T. Each pixel p_{m_i} in T of class c can also be denoted as a triple $\{\boldsymbol{p}_{m_i}, w_{m_i}, c\}$, where $mi \in \{1, 2, \ldots, N_{\text{tc}}\}$, e.g., all points in Fig. 2(a). The second component of the triple, i.e., the weight, which is a measure to estimate the reliability of the samples in T, is first initialized as 1, and is then updated during the query process. Given the training sample set sized N_{tc}, the average of the weighted spectral angle distance (WSAD) of all the pair-wise samples of class c, called $\text{WD}_c(\text{T})$, can be denoted by

$$\text{WD}_c(\text{T}) = \frac{2\sum_{m_i=1}^{N_{\text{tc}}}\sum_{n_k=1,n_k\neq m_i}(w_{m_i}\cdot w_{n_k})\left(\frac{\boldsymbol{p}_{m_i}^1\cdot\boldsymbol{p}_{n_k}}{\|\boldsymbol{p}_{m_i}\|_2\cdot\|\boldsymbol{p}_{n_k}\|_2}\right)}{n_{\text{tc}}(N_{\text{tc}}-1)} \tag{1}$$

where $\boldsymbol{p}_{m_i}^T$ means the transpose of feature $\boldsymbol{p}_{m_i}$. $\text{WD}_c(\text{T})$, i.e., the intraclass average WSAD, is inversely proportional to the average similarity of the pair-wise samples in T of class c. Considering that the Q function aims to pick out the unlabeled samples which can bring new information to T, the samples whose distance to the current training sample set T is larger than the intraclass average similarity are desirable. Thus, the distance between the unlabeled sample p_{s_j} in U and the current training sample set T of class c, shortened to $\text{WD}_c(p_{s_j})$, is modeled in a similar fashion as

$$\text{WD}_c(p_{s_j}) = \frac{\sum_{m_i=1}^{N_{\text{tc}}}(w_{m_i}\times 1)\left(\frac{\boldsymbol{p}_{m_i}^T\times\boldsymbol{p}_{s_j}}{\|\boldsymbol{p}_{m_i}\|_2\times\|\boldsymbol{p}_{s_j}\|_2}\right)^{-1}}{N_{\text{tc}}} \tag{2}$$

where $\text{WD}_c(p_{s_j})$ is inversely proportional to the average similarity between p_{s_j} and all the samples in T of class c. In this way, all the samples in which the distance to T is larger than $\text{WD}_c(\text{T})$, e.g., the solid points p_{s1}, p_{s2}, and p_{s3} in Fig. 2(b), comprise a set U_{temp}, which should be further checked in the next step.

Meanwhile, there are still some pixels [e.g., the hollow points p_{s5}, p_{s6}, and p_{s7} in Fig. 2(b)] that cannot increase the sample diversity [i.e., $\text{WD}_c(p_{s_j}) \leqslant \text{WD}_c(\text{T})$]. This implies that, for a pixel p_{s_j} in U, there is already a sample p_{m_i} in T whose spectral property is very similar to p_{s_j}, and pixel p_{m_i} has the smallest distance to p_{s_j}. In such a case, owing to the redundancy, pixel p_{s_j} should not be selected. However, on the other hand, the similarity between p_{s_j} and p_{m_i} actually verifies the reliability of sample p_{m_i} in T. Therefore, in order to strengthen its importance, the weight of sample p_{m_i} should

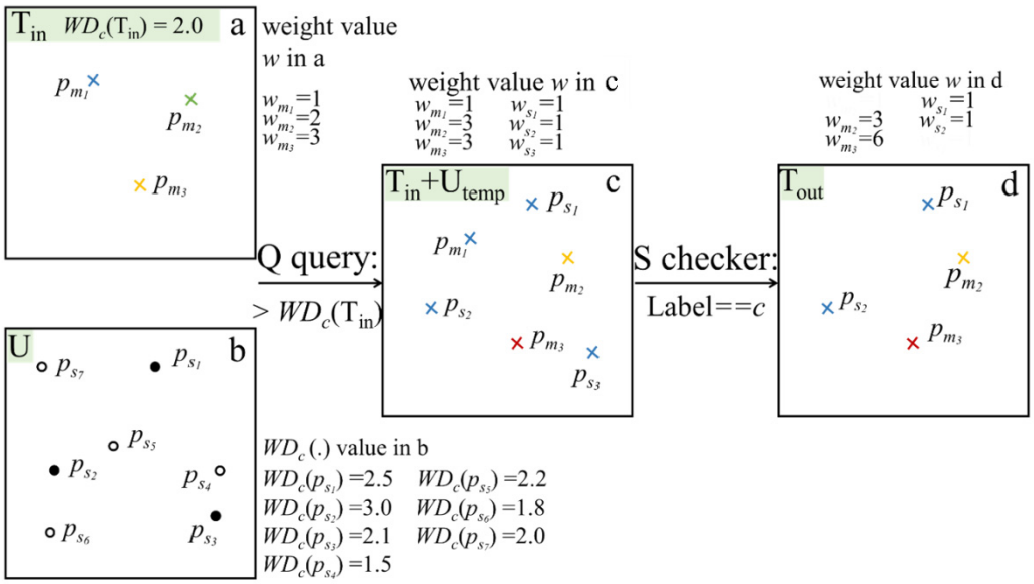


Fig. 2. Demonstration of the proposed sample selection process in one iteration for class c. (a) T_{in} means samples that have been selected in previous iterations, e.g., samples p_{m1}, p_{m2}, and p_{m3}. The weight of each sample is listed at the right side of (a), and the intraclass diversity measured by $\text{WD}_c(T_{\text{in}})$ is also presented. (b) Pixels in the candidate pool U, where the solid points denote the samples whose distance to T_{in} is larger than $\text{WD}_c(T_{\text{in}})$, and the hollow points represent the others. (c) Union of T_{in} and U_{temp}, i.e., samples after the Q query step, with the weight of each sample in T_{in} updated. To be more specific, the three hollow points (p_{s4}, p_{s5}, and p_{s6}) are similar to p_{m3}, and one point, called p_{s7} is similar to p_{m2}. In this way, the weight of p_{m3} has been increased by three, and that of p_{m2} has been increased by one. The newly added points are weighted as 1. (d) T_{out}, i.e., samples after the elimination of the S label checker step (e.g., samples p_{m1} and p_{s3} are removed, as they cannot be labeled as c by S).

be enlarged, as follows:

$$p_{m_i} = \arg\max_{p_{m_i}} \frac{\boldsymbol{p}_{m_i}^T \times \boldsymbol{p}_{s_j}}{\|\boldsymbol{p}_{m_i}\|_2 \times \|\boldsymbol{p}_{s_j}\|_2}, \quad \text{for } p_{m_i} \in \{p_{m_1}, \ldots, p_{N_{\text{tc}}}\}$$
$$w_{m_i} = w_{m_i} + 1. \tag{3}$$

As in (2), a sample with a larger weight will have a larger impact during the collection. Thus, the weight of the existing samples in T can be updated, e.g., the points p_{m2} and p_{m3} in Fig. 2(c).

3) Label Checker S: After the query process, ambiguous pixels from other classes might still be included. Therefore, the label checker S is designed to eliminate the unreliable samples, by considering the following criteria.

1) The uniqueness of the label: ambiguous samples in U_{temp} that simultaneously exist in at least two categories are removed.
2) The consistency of the sample: we import U_{temp} into T and check the label of each pixel in T in an iterative manner. For the kth iteration, E_k means the set of currently excluded samples, which is initialized as empty; and T'_k denotes the updated training samples after exclusion, where T'_0 is equal to T. In the kth iteration, for each sample p_{m_i}, we build a spectral angle mapper classifier, by using all the other samples in T'_{k-1}, except for p_{m_i} itself, as the reference spectra. If the temporary classification result is consistent with its label c, p_{m_i} is assigned to T'_k; otherwise, it is imported into E_k. The stopping condition of the iteration refers to the number of samples in E_k of class c being much less than those in T'_k of class c (i.e., less than 1% in this paper). In this way, the possible outliers, whether selected by the current query function, or wrongly collected by the previous iterations, can be further checked.

In this way, through the query process and the label check of the current iteration, the updated training sample set is input into the next iteration [e.g., p_{m2}, p_{m3}, p_{s1}, and p_{s2} in Fig. 2(d)], until the stopping condition is reached.

C. General Framework and Stopping Condition

Due to the lack of a ready-made spectral index to describe dark BU, four-category (bright BU, V, WS, and BS) sample collection is implemented in the first few iterations. The SDBI is then generated on the basis of the current samples, and five-category (bright BU, V, WS, BS, and dark BU) sample collection is finally put into practice with all five index images. In summary, the framework consists of the four-category stage (Q_1, S_1, T_1, U_1), the SDBI generation, and the five-category stage (Q_2, S_2, T_2, U_2) (see Algorithm 1):

As can be seen in Algorithm 1, there are two iterative stages with specific stopping conditions and an SDBI generation approach. As suggested in [8], compared with NDWI, MNDWI is more capable of highlighting the representative water surfaces with large MNDWI values. Thus, candidate pixels with large MNDWI values, i.e., the ones to be collected in the first few iterations, should be representative samples. As the goal of this stage is to produce the sample set for the following water mask, a smaller number of iterations for the first stage is appropriate. In this paper, we refer to the 50th iteration as the threshold for each class in the first stage. Next, the guidelines for setting T_c, i.e., the stopping conditions for each class of the five-category stage, are illustrated in Fig. 3 and summarized as follows.

Algorithm 1 The Proposed Unsupervised Training Sample Collection Algorithm

1. Input: Landsat image, NDBI, NDVI, MNDWI, BI, and NDWI

2. The four-category stage

Initialization: Linearly normalize NDBI, NDVI, MNDWI, and BI

Build the sample set (T_1) with the largest spectral index value for each class

Repeat for each class in alphabetical order:

Query a set of samples (with the query function Q_1) from the pool U_1

Label the selected samples and exclude the outliers of T_1 by a label checker S_1

Update the sample set T_1

Until a stopping criterion is satisfied

Output: temporary sample set T_1 for bright BU, V, WS, and BS

3. SDBI generation

Build a binary urban water mask by a classifier trained with T_1

Refine the urban water mask by a morphological dilation operation

Retain the value of each non-water pixel in the NDWI image and assign the values of all the water pixels to zero

4. The five-category stage

Initialization: Linearly normalize the SDBI

Build the sample set (T_2) by combining T_1 and an initial sample set with the largest SDBI index values for the dark BU class

Repeat for each class in alphabetical order:

Query a set of samples (with the query function Q_2) from the pool U_2

Label the selected samples and exclude the outliers of T_2 by a label checker S_2

Update the sample set T_2

Until a stopping criterion is satisfied

Output: T_2 for bright BU, dark BU, V, WS, and BS

5. Output: Sample set T_2, made up of BU (by merging the bright and the dark subclasses), V, WS, and BS

1) Considering the spectral variation of the BU category, we suggest that the thresholds for NDBI and SDBI are set in the range of 50 to 500.
2) For the V and WS categories, the suggested threshold value is 50, based on their homogenous spectral characteristics.
3) As the BS category has diverse spectral characteristics in urban areas, a reasonable value is between 50 and 200.

Please note that the thresholds for both stages are further analyzed in Section IV-D.

As shown in Fig. 4, the SDBI generation approach can be summarized as follows. A binary urban water mask can be produced by building a classifier (see Section II-D) with samples collected in the four-category stage (T_1). The derived water mask is then refined by a morphological dilation operation to reduce the noise and false alarms. The SDBI is then generated by retaining the value of each nonwater pixel in the NDWI image and assigning the values of all the water pixels to zero. In this way, the SDBI can be used to indicate the dark BU class.

D. Classification

In the proposed approach, with the automatically selected sample set as the training data, **c**ollaborative **r**epresentation-based **c**lassification (CRC) [22] is adopted to produce the final land-cover map. CRC is an effective supervised machine learning method, which has achieved outstanding performances in various fields, e.g., face recognition [22], [23], human action recognition [24], visual tracking [25], and hyperspectral image classification [26]–[28]. In CRC, for a labeling task, each test signal $\boldsymbol{p}$ can be linearly and densely represented by the collaborative codebook $\mathbf{A}$constructed by stacking all the training samples [27], [28]. Thereafter, a regression algorithm (e.g., least squares) is employed to assign the label with the smallest sub-regression residual. The mathematical expression is represented as follows:

$$\text{Collaborative representation: } \boldsymbol{\alpha} = \underset{\alpha}{\operatorname{argmin}}(||\boldsymbol{p} - \mathbf{A}\boldsymbol{\alpha}||_2 + \lambda||\boldsymbol{\alpha}||_2) \quad (4)$$

$$\text{Classification: class } (\boldsymbol{p}) = \underset{(i=1,\ldots M)}{\operatorname{argmin}} ||\boldsymbol{p} - \mathbf{A}_i\boldsymbol{\alpha}_i||_2 \quad (5)$$

where λ is a tradeoff between the data-fidelity term and the coefficient prior, which is often set as a small nonzero value by default, to avoid a trivial solution. $\mathbf{A}_i$, $\boldsymbol{\alpha}_i$ are subsets of the dictionary and coefficient vector associated with class i, respectively. CRC was naturally designed for classification by simultaneously using training samples from all the classes to stabilize the class pattern, and hence is especially suitable for multiclass applications with limited supervised training information [27]. In this manner, when dealing with a land-cover task involving similar classes (such as BS and BU), the CRC classifier can make use of the information from multiclass and complementary training samples. In addition, the dense representation means that the labeling process should not be heavily dependent on a few (sparse) samples [27], [28], so that it is capable of mitigating the effect of a small amount of wrongly labeled samples, thus strengthening the classification robustness.

In the classification procedure, the training samples are randomly generated from the automatically collected sample set. Ideally, more training samples can result in a more desirable classification performance. However, considering the computational cost, in this paper, 500 pixels are randomly selected for each class (see Section IV-D for a further discussion). The spectral bands, as well as the index images, are stacked as the input features. After classification, the bright and dark BU classes are merged into one class (i.e., the BU class), and the mapping result is finally generated.

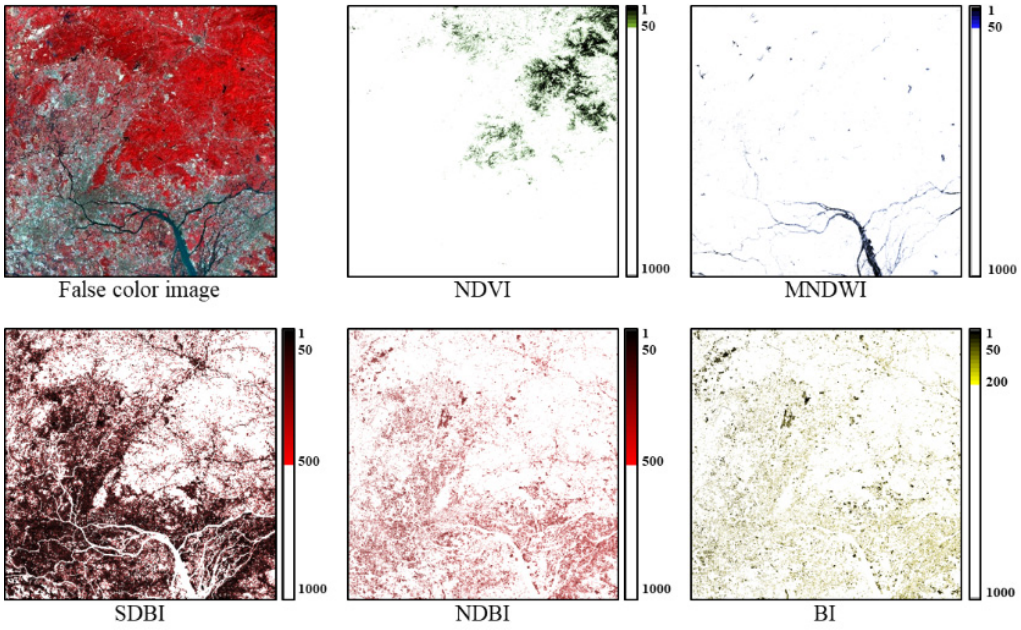


Fig. 3. Distribution and the specific index values of the pixels that meet the guidelines to set the T_c parameters for each class, taking the Guangzhou Landsat image as an instance. Each legend bar is consistent with the guidelines for the associated land-cover type.

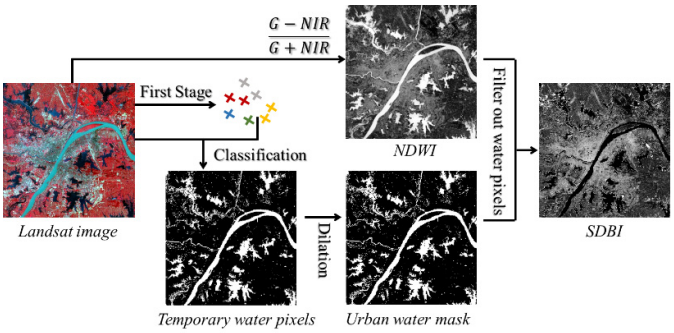


Fig. 4. Production of the synthetic dark BU image.

E. Computational Complexity

In theory, the computational cost mainly consists of the four-and five-category stages and the two classifications, one of which is implemented during the SDBI generation, and the other is the final land-cover labeling. In view of the sample collection stage, for the ith iteration, the training sample set of class c is sized $N_{\mathrm{tc},i}$, and the candidate pool of class c is sized $N_{\mathrm{uc},i}$. The query processing step thus takes $O(N_{\mathrm{uc},i}N_{\mathrm{tc},i})$, and the label checker step takes $O(g(C-1)N_{\mathrm{uc},i}N_{\mathrm{tc},i})$, where C is the number of classes and g (a small value in practice) is the number of iterations in the ith label checker step. In view of this, given L iterations as the maximum threshold of the second stage, the computational cost of the sample collection stage is $O(C^2\sum_i^L N_{\mathrm{uc},i}N_{\mathrm{tc},i}) = O(C^2LN_{\mathrm{uc}}N_{\mathrm{tc}})$, where N_{tc} and N_{uc} are the maximum sizes of the training sample set and candidate pool for a single class, respectively. Meanwhile, for a Landsat image with F pixels, the time complexity of the two classifications is $O(Fd\sum n_{\mathrm{tc}})$, where d is the dimension of the features, and n_{tc} is the size of the randomly selected subset of N_{tc} (500 in this paper), which is much smaller than N_{tc}. In summary, it can be said that the computational complexity, i.e., $O(Fd\sum n_{\mathrm{tc}} + C^2LN_{\mathrm{uc}}N_{\mathrm{tc}})$, is linear with regard to the number of pixels F in the Landsat image.

III. Experimental Data Sets and Setup

A. Study Area and Data Sources

In this paper, according to the stratification of urban eco-regions developed by [29], 39 representative cities (Fig. 5) were selected throughout these eco-regions (except for the Boreal forest region and the tundra/permanent ice/snow areas) to validate the effectiveness of the proposed method. These cities are located in various biomes, with different urban topologies and economic development situations. The related parameters [i.e., population (POP) and area (available at: http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/GHSL/)] are also listed for each city.

This research relied on terrain-corrected Level 1T Landsat 30-m resolution data, which can be freely acquired from the United States Geological Survey (USGS) (http://earthexplorer.usgs.gov/). For each city, cloud-free Landsat scenes were selected to cover the study area. We chose adjacent scenes acquired on the same date if more than one scene was needed to cover a city, to avoid any differences in imaging conditions. To obtain the study area of each city, three specific masks, i.e., Fmask [30], an urban area mask, and an open water surface mask, were utilized. Fmask records every pixel in the scene as valid (surface reflectance) or not (e.g., cloud, cloud shadow, saturation, out of satellite swath, and striping). The urban area mask (http://e4ftl01.cr.usgs.gov/MOTA/MCD12Q1.051/) was used to focus our test images

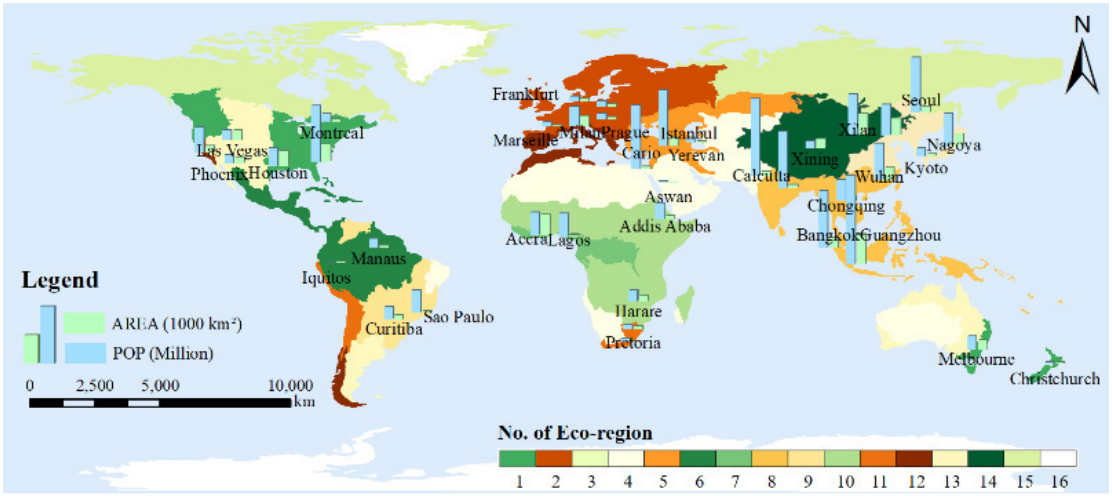


Fig. 5. Locations of the 39 cities considered in this paper (Fig. 6). Spectral profiles of the different land-cover classes in the real and synthetic data sets. (a) Real average profiles for each land-cover class. (b)–(e) Synthetic BU, V, WS, and BS profiles, respectively.

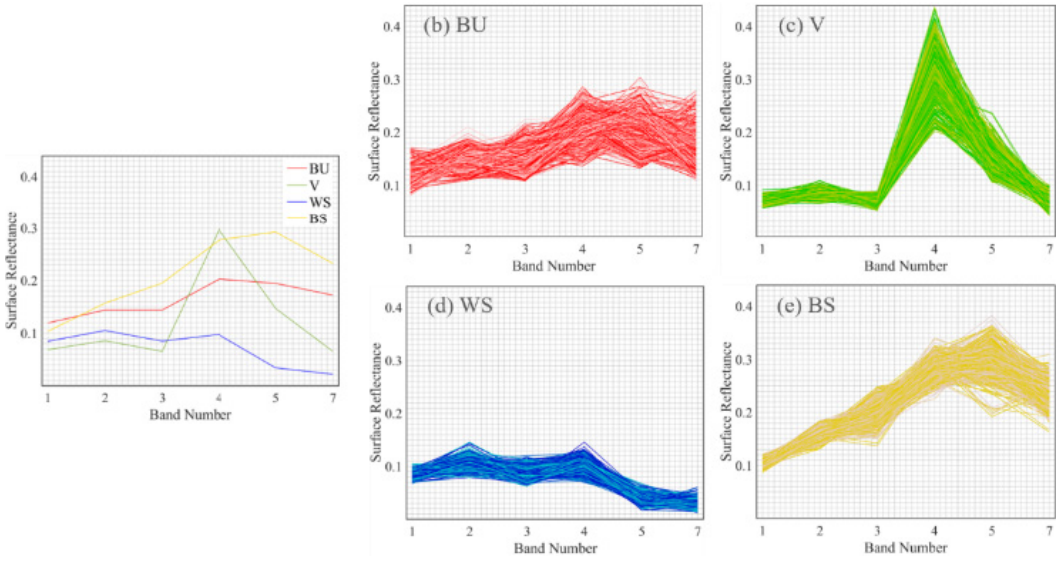


Fig. 6. Spectral profiles of different land-cover classes in the real and synthetic data sets. (a) Real average profiles for each land-cover class. (b)–(e) Synthetic BU, V, WS, and BS profiles, respectively.

on the urban areas. Open water surfaces (i.e., oceans, rivers, and lakes with large areas, as defined in [31]) outside the urban areas were masked out using the global open water body product from the European Space Agency [spatial resolution: 150 m; overall accuracy (OA): 96%; Kappa: 0.91; website: http://maps.elie.ucl.ac.be/CCI/viewer/download.php]. Thus, only urban water bodies (such as paddy fields, ponds) were considered in our research. In total, 41 Landsat image scenes for 39 cities were selected, comprising 32 Thematic Mapper (TM), four Enhanced Thematic Mapper Plus (ETM+), and five Operational Land Imager OLI scenes (Table II).

B. Synthetic Data Set

A synthetic data set was used to quantitatively evaluate the quality of the samples collected by the proposed method and the effect of label noise on the classification. In this paper, the reference samples collected by field survey in the area of the Wuhan Landsat image were used to produce a synthetic data set (referred to as *SimuGT*), involving 80 000 pixels (20 000 pixels for each class). Based on the assumption of a normal distribution for each land-cover type and high spectral similarity within an object, we used an object-level simulation procedure [32] to bridge the gap between real and synthetic data. The normal distribution parameter for each class (i.e., mean μ_c and standard deviation σ_c for class c) was estimated by the average spectral profile of each region of interest (ROI). In total, 20 000 pixels for each class were evenly grouped as 1000 objects, with each object containing 200 pixels. For each object of class c, the range of the spectral profile was randomly selected from (μ_c, σ_c), and the spectral profile of each pixel in this object was then randomly set within this range. In this way, for the objects in the same class, samples within an object should be more similar than those across different objects. After all the objects were generated, uniform noise was added to the whole synthetic data set. Fig. 6 presents some examples from the synthetic data set.

TABLE II
LIST OF CITIES AND THE LANDSAT DATA USED IN THIS STUDY

No.	Eco-region	Urban topology	City (Country)	Sensor	Date (Path/row)
1	Temperate forest in North America	Rectilinear grid structure, segregated land use, high-density core, low-density suburbs	New York (United States)	Landsat TM	2010-09-22 (13/32)
			Houston (United States)	Landsat TM	2010-08-25 (25/39)
			Montreal (Canada)	Landsat TM	2009-09-10 (14/28)
			Melbourne (Australia)	Landsat TM	2010-10-24 (93/86)
			Christchurch (New Zealand)	Landsat TM	2006-11-02 (73/90)
2	Temperate forest in Europe	Radial street structure, varying building density, blurred urban/rural boundaries	Marseille (France)	Landsat ETM+	2000-06-16 (169/30)
			Prague (Republic of Czech)	Landsat TM	2010-07-03 (191/25)
			Frankfurt (Germany)	Landsat ETM+	2000-08-12 (195/25)
			Stuttgart (Germany)	Landsat TM	2010-07-08 (194/26)
3	Temperate forest in East Asia	Rectilinear grid structure, high-density rect. buildings, newly developed outer nuclei/towns	Wuhan (China)	Landsat TM	2009-09-06 (123/39)
			Shenyang (China)	Landsat TM	2010-08-12 (119/31)
			Kyoto (Japan)	Landsat ETM+	2000-08-25 (110/36)
			Seoul (South Korea)	Landsat TM	2011-09-27 (116/34)
			Nagoya (Japan)	Landsat TM	2009-11-07 (109/36)
4	Temperate grassland in North/South America	Low-density, rectilinear grid structure, segregated land use, wide streets	Cairo (Egypt)	Landsat ETM+	2000-08-23 (176/39)
			Aswan (Egypt)	Landsat TM	2002-06-20 (174/43)
5	Temperate grassland in Middle East, Asia	Rectilinear grid structure, high-density block structure	Istanbul (Turkey)	Landsat TM	2010-09-08 (180/32)
					2010-09-08 (180/31)
			Yerevan (Armenia)	Landsat TM	2010-08-01 (170/32)
6	Tropical broadleaf forest in South America	Rectilinear grid structure, central axis layout, typical new frontier cities	Manaus (Brazil)	Landsat OLI	2016-07-27 (231/62)
			Iquitos (Peru)	Landsat OLI	2016-08-20 (6/63)
7	Tropical broadleaf forest in Africa	Compact cities/towns, high density, mixed land use, varied structure	Lagos (Nigeria)	Landsat OLI	2015-12-08 (191/55)
			Accra (Ghana)	Landsat OLI	2014-03-22 (193/56)
8	Tropical, sub-tropical forest in Asia	High density, tightly spaced buildings, narrow streets, mixed land use	Chongqing (China)	Landsat TM	2010-08-04 (127/40)
			Guangzhou (China)	Landsat TM	2011-09-21 (122/44)
			Bangkok (Thailand)	Landsat TM	2005-03-13 (129/50)
					2005-03-13 (129/51)
			New Delhi (India)	Landsat TM	2010-10-12 (146/40)
			Calcutta (India)	Landsat TM	2010-12-23 (138/44)
9	Tropical, sub-tropical savannah in South America	Rectilinear grid structure, central axis, dense block structure, segregated land use	Curitiba (Brazil)	Landsat OLI	2015-08-29 (220/78)
			Sao Paulo (Brazil)	Landsat TM	2010-04-18 (219/76)
10	Tropical, sub-tropical savannah in Africa	Irregular street structure, medium density, vegetated core, high-density outskirts	Addis Ababa (Ethiopia)	Landsat TM	2010-12-09 (168/54)
			Harare (Zimbabwe)	Landsat TM	2008-08-11 (170/72)
11	Tropical, sub-tropical grassland	Radial structure, central axes, high–medium density core/outskirts, segregated land use	Johannesburg (South Africa)	Landsat TM	2009-04-08 (170/78)
			Pretoria (South Africa)	Landsat TM	2009-04-08 (170/78)
12	Temperate Mediterranean	Radial, semi-irregular street structure, high–medium building density	Milan (Italy)	Landsat TM	2009-09-07 (194/28)
			Los Angeles (United States)	Landsat TM	2011-08-28 (41/36)
13	Arid, semi-arid desert, shrubland	Varied	Las Vegas (United States)	Landsat TM	2010-08-11 (39/35)
			Phoenix (United States)	Landsat TM	2011-07-15 (37/37)
14	Arid, semi-arid steppe in Central Asia	Irregular street layout, conjoined buildings, segregated land use	Xi'an (China)	Landsat TM	2010-06-17 (127/36)
			Xining (China)	Landsat TM	2011-08-07 (143/30)
15	Boreal forest, tundra	Small towns, villages	-	-	-
16	Permanent ice/snow	-	-	-	-

C. Reference Data Set

To assess the accuracy of the classification, the ground references were obtained through in-lab inspection of the Landsat data, photointerpretation based on Google Earth high-resolution images, and field campaigns to the relevant Chinese cities. Evenly distributed object patches (i.e., ROIs) for different land-cover types were manually delineated for every Landsat scene. The statistics of the ROIs in each study site are detailed in Table IX. On average, for each study site, the number of ROIs was 164, 113, 57, and 76 for BU, V, WS, and BS, respectively. Since the samples within an ROI have similar spectral and spatial characteristics, only one pixel was randomly selected from each ROI, to avoid spatial autocorrelation [35]. To allow us to achieve an unbiased accuracy assessment, the test sample selection and accuracy assessment were independently repeated 10 times during the validation procedure, and the statistical accuracy was recorded.

In order to assess the accuracy of the selected samples in the real data sets, we chose six cities in different classification accuracy levels (Montreal and Shenyang with an OA of less than 90%, Stuttgart and Addis Ababa with an OA of 90%–95%, and Wuhan and New Delhi with an OA of higher than 95%). In total, 600 samples (100 samples for each city) were randomly selected and used to calculate the correctness, by visual comparison with the high-resolution images from Google Earth.

D. Experimental Setup

1) Parameter Setting for the Proposed Method:

1) Considering the computational cost of the proposed approach, 500 training pixels per class were randomly selected to implement the classification.
2) Considering that a large number of pixels within an index range will be spectrally similar and redundant,

TABLE III
CONFUSION MATRIX FOR THE SAMPLES COLLECTED IN THE TWO STAGES FROM THE SYNTHETIC DATA SET

T_1 (T_2)*	Ground reference					
	BU	V	WS	BS	Total	UA (%)
BU	915(3434)	0(0)	0(0)	0(81)	915(3515)	100.0(97.7)
V	0(0)	4058(4058)	0(0)	0(0)	4058(4058)	100.0(100.0)
WS	0(0)	0(0)	2103(2103)	0(0)	2103(2103)	100.0(100.0)
BS	206(68)	0(0)	0(0)	992(1223)	1198(1291)	82.80(94.73)
Total	1121(3502)	4058(4058)	2103(2103)	992(1304)	8274(10967)	-
PA (%)	81.62(98.06)	100.0 (100.0)	100.0(100.0)	100.0(93.79)	-	-
	OA (%)	97.51(98.64)		Kappa	0.962(0.981)	

*T_1 (T_2) indicates the number of samples collected by the four-category (five-category) stage.

during each iteration; 2000 pixels were randomly selected to implement the proposed sample collection.

3) For the second stage, the thresholds for both vegetation and water surfaces were fixed as 50, and those for the other two classes were empirically tuned (see Table VIII), based on the guidelines mentioned before.
4) The threshold for the first stage was set as 50 iterations, in consideration of the high variance of MNDWI [8] and the homogenous spectral characteristics of water surfaces.
5) According to the previous studies on CRC [27], [28], a nonzero but small λ is desirable. In this paper, λ was set as $1e^{-3}$.

Please refer to Section V-B for a detailed analysis of the sensitivity of the parameters mentioned above. In addition, considering that the spatial resolution of Landsat imagery is 30 m, we fixed a 5×5 disk for the morphological dilation operation in the SDBI generation.

2) Compared Mapping Methods: In this paper, we utilized two other methods to test the performance of the proposed algorithm.

1) The "simple sampling" method is based on random forest (RF) and four spectral indices, i.e., NDBI, NDVI, MNDWI, and BI. Specifically, candidate instances for all four classes were first extracted according to manually tuned thresholds (i.e., NDBI for BU, NDVI for vegetation, MNDWI for water, and BI for bare soil). The ambiguous samples that simultaneously existed in at least two categories were then removed. Finally, a RF classifier was constructed by the use of the remaining instances.
2) To compare the proposed method with supervised classification based on manually selected training samples, additional ROIs (detailed in Table X) were manually selected in a similar way to train the classifier. In the following, we refer to this method as the "manual sampling" method. For the purpose of a fair comparison, CRC was also taken as the classifier.

3) Compared Classifiers and the Experimental Environment: To analyze the suitability of utilizing the CRC classifier in the proposed framework, support vector machine (SVM) and RF classifiers were used as a comparison, implemented in LibSVM (version: 3.23) software [40] and the MATLAB 2017b built-in TreeBagger function, respectively. The parameters of SVM were set as kernel = radial basis function (RBF), and the penalty coefficient and RBF bandwidth were tuned by tenfold cross-validation. For the RF classifier, 500 trees were constructed, considering the tradeoff between computational burden and classification accuracy [41]. A random subset of $\sqrt{n}$ features was used for RF at each node, where n is the number of features [42]. All the experiments were carried out using MATLAB on a PC with one 4-GHz processer and 64 Gb of RAM.

IV. RESULTS AND ANALYSIS

A. Sample Quality Assessment

For the synthetic data set, the indicators of OA, Kappa coefficient (Kappa), producer's accuracy (PA), and user's accuracy (UA), derived from the confusion matrix, were used to assess the accuracy of the selected samples. In Table III, T_1 (containing 8274 samples) and T_2 (containing 10 967 samples) represent the sample set after the first stage (i.e., the four-category stage) and the second stage (i.e., the five-category stage), respectively. In general, both stages present pleasing accuracies (Kappa $>$ 0.96), and the second stage shows a more desirable performance. For the classes such as V and WS, there was no sample elimination during the second stage, which was mainly due to their relatively homogeneous spectral characteristics. For the other two land-cover types, changes of the diagonal elements in Table III indicate the addition of newly selected samples during the second stage. In particular, please note that, with the addition of the SDBI in T_2, some dark BU samples that were wrongly identified as BS in T_1 were eliminated (from 206 to 68).

From the 39 cities considered in this paper, we chose six cities in different classification accuracy levels (Montreal and Shenyang with an OA of less than 90%, Stuttgart and Addis Ababa with an OA of 90%–95%, and Wuhan and New Delhi with an OA of higher than 95%). In total, 600 samples (100 samples for each city) were randomly selected and used to calculate the correctness by visual comparison with the high-resolution images from Google Earth. The sample accuracies for both WS and V are 100%. Although a small number of mistakes exist in the BU (accuracy: 91%) and BS samples (accuracy: 84%), the effectiveness of the proposed approach is confirmed, with the total sample accuracy reaching 94%. Furthermore, to visually investigate the diversity of the collected samples in the real Landsat images, the selected samples in the Wuhan (accuracy: 92%) and Montreal (accuracy: 92%)

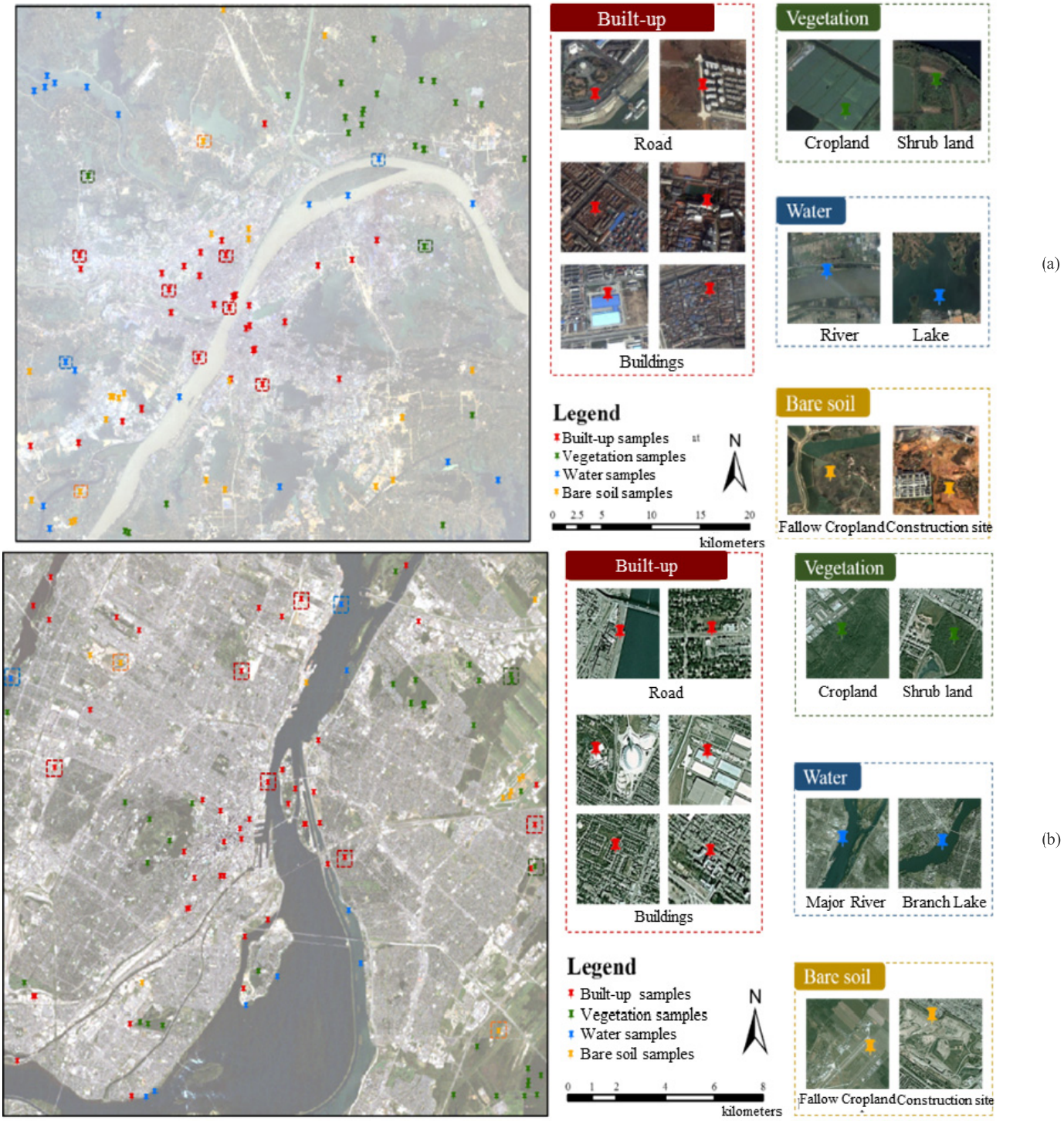


Fig. 7. Visual representation of the selected samples for (a) Wuhan image and (b) Montreal image.

images are displayed in Fig. 7. By referring to the Google Earth images, it can be seen that the proposed method can select diverse samples. For instance, as shown in the zoomed-in images of the samples of the BU class in the Wuhan image [Fig. 7(a)], the BU samples not only cover buildings with bright and dark colors, but also involve different types of roads, such as arterial roads and pathways. The samples of the V class contain various forms of vegetated land, including crops and scrubland. The WS samples include both river and lake areas, and the BS samples are distributed in both fallow cropland and construction sites. Similar examples can also be observed in the other cities [e.g., Montreal in Fig. 7(b)].

B. Classification Results

Representative mapping results for the various eco-regions are shown in Fig. 8. In general, the identified BU areas

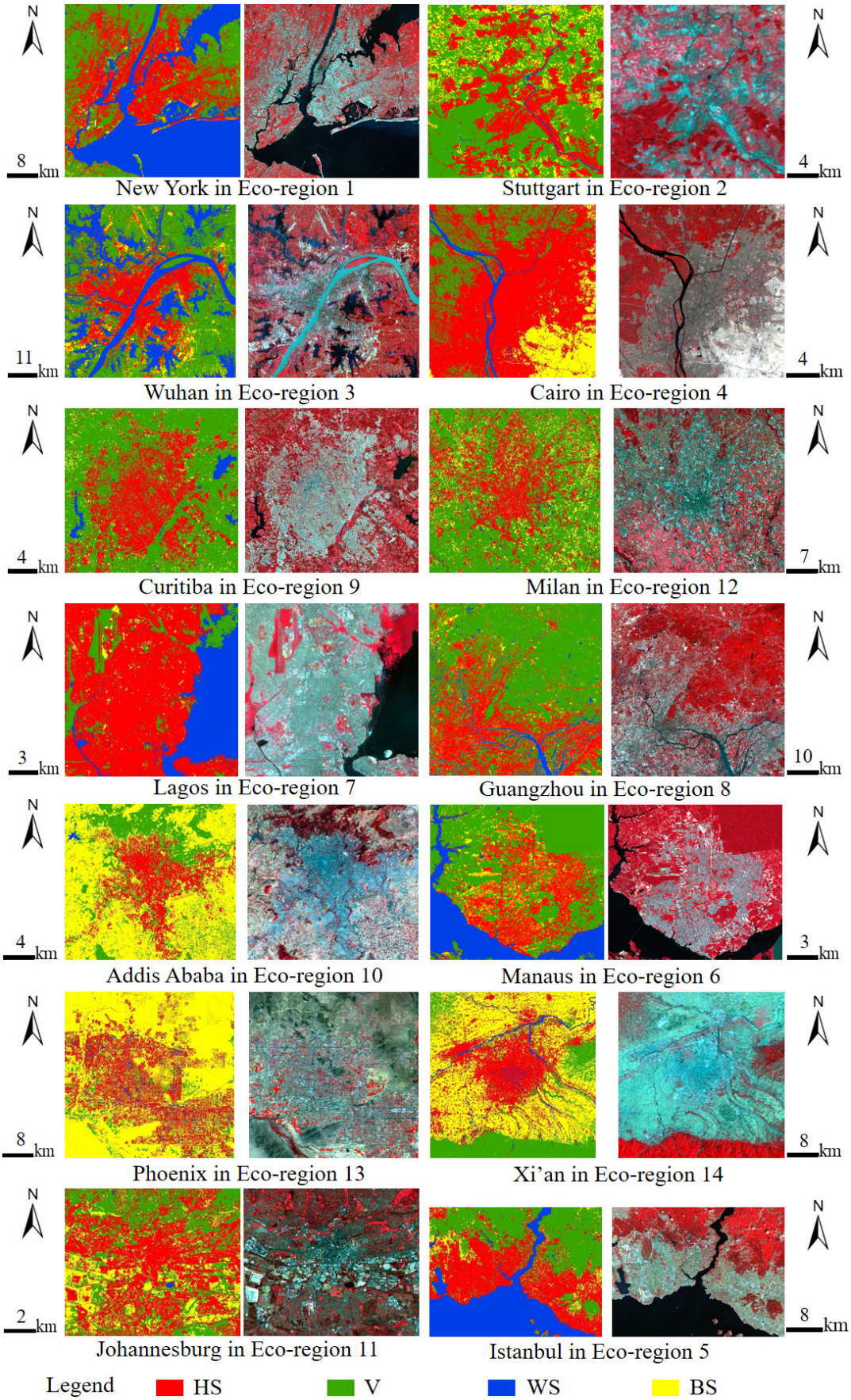


Fig. 8. Results for the representative cities in each eco-region. (Left) Land-cover mapping results of the proposed method. (Right) False-color images.

TABLE IV
ACCURACY ASSESSMENT OF THE CLASSIFICATION RESULTS FOR THE 39 CITIES

Study site	Proposed automatic sampling										Simple sampling		Manual sampling	
	PA (%)				UA (%)				OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
	BU	V	WS	B	BU	V	WS	B						
New York	96.7	98.0	74.1	95.8	76.8	99.5	93.3	99.6	91.2	0.88	88.8	0.78	94.5	0.90
Melbourne	99.5	100.0	54.6	78.0	87.9	99.7	99.0	98.6	93.3	0.89	84.2	0.74	95.7	0.93
Prague	88.6	100.0	93.1	98.5	92.0	99.3	100.0	89.6	95.0	0.93	81.6	0.72	95.8	0.94
Frankfurt	98.0	96.7	90.7	99.1	93.4	100.0	100.0	97.2	96.9	0.96	75.3	0.65	97.9	0.97
Stuttgart	99.0	100.0	83.0	81.6	90.6	93.7	98.5	98.9	93.8	0.91	75.9	0.64	98.7	0.98
Wuhan	97.0	96.6	95.8	91.7	93.7	99.6	92.3	95.8	95.7	0.94	90.9	0.88	94.9	0.93
Shenyang	89.9	100.0	86.1	73.1	88.8	89.9	99.6	82.0	89.8	0.86	83.4	0.76	96.8	0.96
Kyoto	95.0	100.0	80.9	84.3	96.1	99.9	99.6	71.2	94.3	0.91	87.8	0.82	96.9	0.95
Seoul	95.0	100.0	80.9	84.5	96.1	99.9	100.0	71.5	94.3	0.91	86.6	0.78	99.5	0.99
Nagoya	94.8	98.4	91.3	98.9	96.2	99.5	98.6	89.2	96.0	0.94	78.3	0.68	91.5	0.88
Cairo	96.0	98.7	99.0	94.7	96.0	100.0	100.0	90.1	97.0	0.96	96.2	0.95	98.1	0.97
Aswan	91.6	100.0	79.3	82.0	82.8	99.4	100.0	87.3	90.3	0.86	73.6	0.62	90.4	0.87
Istanbul	96.5	100.0	85.2	92.2	95.3	100.0	100.0	88.7	96.5	0.95	87.0	0.79	98.8	0.98
Yerevan	85.5	99.8	69.5	81.5	84.6	89.4	89.7	85.9	87.0	0.82	77.4	0.70	91.6	0.88
Manaus	94.0	98.5	81.0	92.6	96.1	99.3	83.1	82.4	94.6	0.92	51.3	0.11	95.8	0.93
Iquitos	98.1	100.0	80.2	91.8	94.0	98.3	99.5	82.8	94.1	0.92	49.0	0.28	95.6	0.94
Lagos	98.9	94.3	99.1	98.8	99.2	100.0	89.1	96.3	97.7	0.97	48.4	0.08	96.7	0.95
Accra	87.9	98.9	84.3	71.6	77.1	96.2	100.0	78.9	87.1	0.82	81.6	0.74	86.9	0.85
Chongqing	94.0	99.0	97.1	72.9	86.3	96.3	95.0	92.0	91.6	0.88	78.4	0.69	97.1	0.96
Guangzhou	97.7	97.2	99.7	97.3	94.8	99.8	99.9	97.6	98.0	0.97	81.0	0.72	98.1	0.97
Bangkok	99.5	99.3	99.9	93.7	97.2	100.0	100.0	98.6	98.8	0.98	86.5	0.80	99.2	0.99
New Delhi	99.4	99.8	96.8	98.5	99.2	98.8	99.5	98.4	99.0	0.99	83.9	0.76	97.6	0.96
Calcutta	95.2	89.6	91.9	81.9	93.8	89.6	99.9	69.2	91.5	0.88	86.5	0.80	92.5	0.89
Curitiba	98.8	100.0	90.4	94.8	97.0	97.9	100.0	98.0	97.7	0.97	51.2	0.19	96.3	0.95
Sao Paulo	90.7	100.0	77.4	75.6	91.7	99.6	99.3	56.7	90.9	0.85	88.8	0.81	94.1	0.90
Addis Ababa	82.0	100.0	80.0	100	99.3	98.2	100.0	86.6	93.6	0.91	93.3	0.90	96.4	0.95
Harare	92.8	84.9	90.2	79.9	75.9	98.3	100.0	86.0	86.4	0.81	67.7	0.56	90.8	0.87
Johannesburg	96.8	100.0	44.7	85.3	85.0	97.4	100.0	95.4	91.6	0.88	66.4	0.50	95.4	0.93
Pretoria	91.7	100.0	95.2	70.0	88.5	94.2	93.1	85.4	90.2	0.86	76.1	0.63	85.4	0.78
Milan	88.1	100.0	89.0	95.6	98.0	94.8	100.0	78.2	92.7	0.90	77.9	0.66	96.2	0.94
Los Angeles	96.3	98.2	70.8	96.7	96.4	95.4	84.0	93.5	94.7	0.92	76.4	0.59	94.6	0.91
Las Vegas	88.3	100.0	24.2	89.7	80.3	98.1	43.2	87.8	84.3	0.77	52.9	0.41	89.5	0.85
Phoenix	81.2	97.8	89.7	91.4	88.2	99.3	90.8	80.2	89.0	0.85	62.7	0.52	91.6	0.88
Xi'an	99.2	99.6	98.1	91.3	92.5	99.8	100.0	98.7	96.9	0.96	74.3	0.65	96.3	0.95
Xining	83.9	98.7	94.8	69.3	76.7	100.0	93.9	78.2	86.0	0.81	67.7	0.53	75.5	0.65
Mean	93.9	98.5	81.4	87.1	90.6	97.8	93.9	88.1	92.9	0.90	77.1	0.65	94.5	0.92
SD	5.1	3.0	20.3	10.5	6.7	3.0	13.0	9.8	3.8	0.05	12.9	0.21	4.6	0.06

are located in the downtown and scattered around the city suburbs; BS pixels are effectively recognized in both the city centers and agricultural land; and V and WS regions are also well discriminated in the mapping results. The detailed quantitative assessment for each city is presented in Table IV. In summary, the proposed automatic mapping method achieves a promising accuracy, as the OA ranges from 84.3% to 99% (average: 92.9% $\pm$ 3.8%) and the Kappa coefficient ranges from 0.77 to 0.99 (average: 0.90 ± 0.05) for all the cities. No significant differences are captured across eco-regions (Table IV), while cities in the arid, semiarid, and shrubland regions (i.e., the 13th eco-region, OA: 89.5%$\pm$7.35%, Kappa: 0.85 ± 0.10) show a slightly inferior performance.

Compared with the two benchmark methods, the results show that, in general, the proposed automatic sampling method can obtain a comparable classification accuracy to the manual sampling method, and it is significantly superior to the simple sampling method. On the one hand, due to the lack of a ready-made spectral index to describe the dark BU class, the simple sampling method cannot extract this kind of land cover. On the other hand, despite the fine-tuned threshold for each index used in the simple sampling method, informative samples for each class still cannot be extracted, which leads to the inferior classification result.

The results show that the manual sampling method only outperforms the proposed method by 1.6%$\pm$3.3% in OA, with no significant difference captured by the ANOVA significance test (OA: p-value $= 0.1025$). Taking Wuhan (proposed method: OA $= 95.7\%$, Kappa $= 0.94$; manual method: OA $= 94.9\%$, Kappa $= 0.93$) as an example (Fig. 9), in Site 2, both the BS and BU classes are correctly identified by the automatic [Fig. 9(e)] and manual [Fig. 9(f)] methods. However, in Site 1, the proposed method [Fig. 9(b)] outperforms the manual [Fig. 9(c)] method, and the BU class is overestimated when using the manually selected samples. However, as can be seen in Fig. 9(b), by the use of the proposed method, both the BU and BS classes can be clearly discriminated from the vegetation background.

Overall, the results of the automatic method can be regarded as satisfactory, considering that the proposed strategy is able

TABLE V

MAPPING ACCURACY (OA: %) USING DIFFERENT COMBINATIONS OF SVM, RF, AND CRC IN THE THREE REPRESENTATIVE CITIES

	SVM1-SVM2	*SVM1-CRC2*	*RF1-RF2*	*RF1-CRC2*	*CRC1-RF2*	*CRC1-SVM2*	*CRC1-CRC2*
Prague	93±1.39	92.1±1.01	87.8±1.74	92.1±0.85	88.3±1.44	92.2±1.26	95.3±0.91
New York	83.8±1.09	83.3±0.79	84.6±1.61	92.3±0.98	84.9±1.53	83.6±1.74	91.8±1.05
Guangzhou	95.3±1.11	96.1±0.74	96.4±1.86	97.3±0.66	94.4±1.32	96.3±1.18	97.9±0.88

TABLE VI

MAPPING ACCURACY USING COMBINATIONS OF SVM, RF, AND CRC IN THE SYNTHETIC DATA SET

		C_2-RF	C_2-SVM	C_2-CRC	T_2-RF	T_2-SVM	T_2-CRC
PA (%)	BU	82.01±0.97	96.89±0.79	99.23±0.47	94.88±0.86	99.31±1.04	98.01±0.51
	V	99.28±0.02	91.15±1.01	100.0±0.00	99.27±0.01	74.09±3.47	100.0±0.00
	WS	99.66±0.05	88.88±1.23	99.54±0.04	99.66±0.05	43.15±8.03	100.0±0.00
	BS	99.82±0.21	96.93±0.84	99.86±0.09	97.00±0.27	89.87±1.72	98.79±0.17
UA (%)	BU	99.16±0.73	98.10±0.67	99.39±0.07	96.15±0.77	51.67±1.87	98.78±0.62
	V	99.94±0.02	100.0±0.00	100.0±0.00	99.97±0.01	100.0±0.00	99.94±0.01
	WS	100.0±0.00	100.0±0.00	100.0±0.00	100.0±0.00	100.0±0.00	100.0±0.00
	BS	84.38±0.75	96.93±0.91	99.23±0.06	94.76±0.53	99.24±0.18	98.09±0.08
OA (%)		95.19±0.42	93.75±0.84	99.66±0.24	97.70±0.39	76.60±3.39	99.20±0.35
Kappa		0.932±0.006	0.917±0.011	0.995±0.004	0.969±0.005	0.688±0.057	0.989±0.005

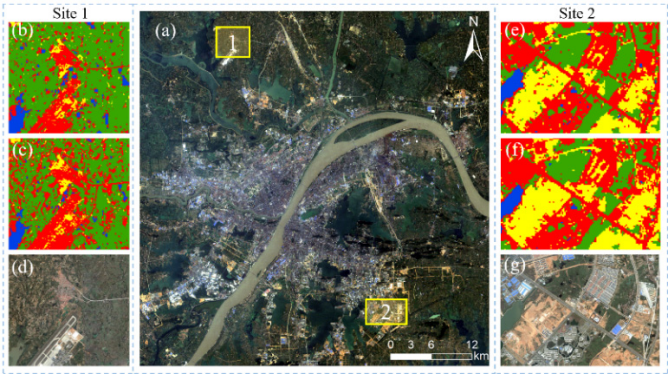


Fig. 9. Visual comparison of the mapping result for Wuhan. (a) Landsat image (R: red; G: green; B: blue). (b) and (c) Classification maps obtained by the automatic and the manual sampling methods, respectively, for Site 1. (d) Corresponding Google Earth images. (e) and (f) Classification maps of the automatic and the manual methods, respectively, for Site 2. (g) Google Earth image for Site 2.

to automatically classify the urban Landsat imagery, and the time-consuming manual sample collection can be avoided. Furthermore, the automatic sampling approach, without requiring human intervention in selecting the training samples, can significantly alleviate the labor cost and increase the efficacy of the mapping process, compared to the manual sampling method.

C. Analysis of the Classifiers and Sensitivity to Label Noise

To analyze the relevance of the choice of classifier, we compared the CRC classifier with SVM and RF classifiers. Prague (~520 km^2), New York (~3865 km^2), and Guangzhou (~6438 km^2) were chosen in this test, considering their different urban sizes, different countries, and different eco-regions. Please note that in the proposed method, the classifier was used twice: first to produce the SDBI after the four-category stage during the sample collection, and second to map the study area after the five-category stage. Accordingly, Table V presents the mapping accuracy with different classifier combinations. For instance, SVM1-CRC2 stands for using SVM in the sample collection (to produce the SDBI after the four-category stage) and CRC in the subsequent land-cover mapping (after the five-category stage). From Table V, the following observations can be made.

1) Compared to SVM1-SVM2 and RF1-RF2, the combination *of CRC1-CRC2* obtains a higher mapping accuracy.
2) When fixing CRC2 and investigating the contribution of constructing the SDBI, CRC is comparable to RF, and both are superior to SVM (see the shaded columns in Table V). One possible reason for this could be the lack of intra-class diversity of T_1, as 50 iterations might result in a premature stopping condition for the first stage. In this case, the advantage of CRC in dealing with multiclass applications with limited training information may partly explain the superior performance of CRC1-CRC2.
3) When fixing CRC1 and analyzing the contribution of the final urban mapping, CRC is superior to RF and SVM (see the last three columns in Table V). In this case, the superiority of RF and CRC can partly be attributed to their robustness in dealing with a small amount of wrongly labeled samples [27], [42].

To further unfold the effect of training label noise on the mapping result, the synthetic data set was again used. In Table VI, for example, T_2-RF stands for using T_2 as the training sample set and RF as the classifier for the land-cover mapping. In the following, we recall the definitions of the sample sets collected in the synthetic data set. T_2 refers to the set selected by the proposed sample collection algorithm after the five-category stage. C_2 indicates the subset of T_2, with its mislabeled samples removed. From Table VI, the following observations can be made.

1) In the noise-free cases (i.e., C_2), all the classifiers show reasonable performances, but CRC presents the best performance.
2) When using the training sample set with label noise (i.e., T_2), SVM is sensitive to noise (PA of WS: 43.15%;

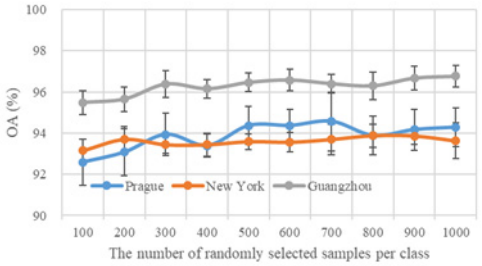


Fig. 10. Classification accuracy versus the number of randomly selected training sample per class for the Prague, New York, and Guangzhou Landsat images.

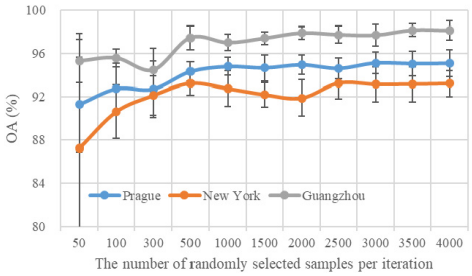


Fig. 11. Classification accuracy versus the number of randomly selected samples per iteration for the Prague, New York, and Guangzhou Landsat images.

PA of V: 74.09%; UA of BU: 51.67%; and inferior OA and Kappa values), while the other classifiers (i.e., RF and CRC) show more robust results. As in [43], the random and bootstrapped characteristics of RF enable this classifier to be robust and error tolerant. As has been claimed before, the dense representation of CRC ensures the capability of mitigating the effect of a small amount of label noise, and strengthens the classification robustness.

D. Parameter Analysis

1) Number of Automatically Collected Samples Per Class: In this section, we analyze the effect of the number of randomly selected samples per class on the classification accuracy, based on the three representative cities. For the purpose of a fair comparison, the numbers of selected samples for each class were kept equal, and 100 to 1000 samples per class (with an interval of 100) were randomly chosen to train the classifier. In these sets, training samples contained in one set were also contained in the affiliated set with a larger number of samples. The horizontal axis in Fig. 10 is the number of randomly selected samples per class, and the vertical axis shows the OA of the mapping result. The accuracy was averaged over 10 runs to reduce the possible bias induced by the random sampling. In general, the accuracies increase slightly with the increment of the training sample size. As can be seen in Fig. 10, it appears reasonable to set the sample size as 500, considering both the computational cost and mapping performance.

2) Number of Randomly Selected Samples Per Interval: Fig. 11 represents the effect of the number of randomly selected samples per iteration on the classification accuracy. Likewise, 50 to 4000 samples were chosen per iteration for each class to run the automatic sample collection procedure. The horizontal axis in Fig. 11 denotes the number of randomly selected samples per iteration, and the vertical axis shows the OA of the mapping result. The accuracy was again averaged over 10 runs. It can be clearly seen from Fig. 11 that at the beginning, the OA increases as the number of samples increases. Subsequently, when the number is larger than 1000, the accuracy curves for all three study areas reach the maximum and become stable from then on. Therefore, in this

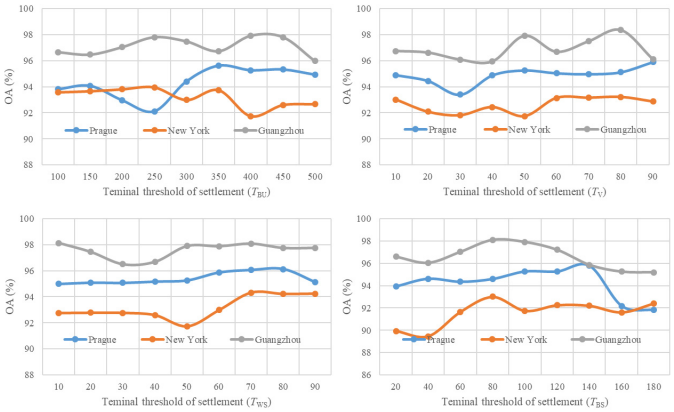


Fig. 12. Threshold sensitivity for the sample selection of (a) BU, (b) V, (c) WS, and (d) BS in the Prague, New York, and Guangzhou Landsat images.

TABLE VII
RUN TIME ANALYSIS FOR THE PROPOSED METHOD AND THE SIMPLE SAMPLING METHOD FOR THE PRAGUE, NEW YORK, AND GUANGZHOU LANDSAT IMAGES

Test site	Size (pixels)	Times (s)	
		Simple sampling	Automatic sampling
Guangzhou	2676×2673	12255.79	7596.19
New York	2061×2084	2059.19	4307.86
Prague	846× 657	215.86	493.19

TABLE VIII
SAMPLE SELECTION THRESHOLD T_c FOR BU AND BARE SOIL

Study site	T_c		Study site	T_c	
	Built-up (BU)	Bare soil (BS)		Built-up (BU)	Bare soil (BS)
New York	400	100	Yerevan	100	50
Houston	400	100	Manaus	400	100
Detroit	400	200	Iquitos	500	100
Toronto	400	100	Lagos	100	100
Montreal	100	100	Accra	100	200
Melbourne	500	100	Chongqing	400	100
Christchurch	400	100	Guangzhou	400	100
London	300	100	Singapore	100	100
Marseille	100	100	Bangkok	200	100
Lyon	300	100	New Delhi	400	100
Yekaterinburg	500	100	Calcutta	50	100
Prague	400	100	Curitiba	100	50
Frankfurt	400	100	Sao Paulo	100	50
Stuttgart	200	50	Addis Ababa	50	100
Shanghai	400	100	Harare	300	150
Wuhan	100	100	Johannesburg	500	100
Shenyang	100	100	Pretoria	500	200
Nagoya	200	100	Milan	400	100
Kyoto	200	100	Los Angeles	400	100
Seoul	200	100	Las Vegas	500	100
Busan	100	200	Phoenix	100	50
Cairo	100	100	Xi'an	100	100
Aswan	100	100	Xining	500	200
Istanbul	300	100			

TABLE IX
STATISTICAL INFORMATION FOR THE ROIS IN THE REFERENCE DATA SET

Study site	Number of ROIs											
	Built-up (BU)			Vegetation (V)			Water (WS)			Bare soil (BS)		
	Count	Median	IQR	Count	Median	IQR	Count	Median	IQR	Count	Median	IQR
New York	409	13.0	21.0	144	80.5	57.0	58	55.0	36.0	63	14.5	17.0
Houston	246	26.0	42.0	190	95.0	86.0	177	70.0	44.0	76	39.0	30.5
Montreal	165	18.0	24.0	104	28.0	23.0	29	59.0	28.0	33	12.0	15.0
Melbourne	209	19.3	30.0	142	57.0	42.5	37	30.8	19.0	54	35.0	34.0
Christchurch	105	16.5	22.0	99	31.5	34.0	18	23.0	9.5	76	39.0	37.0
Marseille	98	16.0	17.5	63	31.0	36.0	20	31.5	25.5	56	13.5	12.0
Prague	146	16.0	26.0	98	72.0	59.5	44	71.5	39.5	60	35.0	25.0
Frankfurt	140	21.0	24.0	111	54.3	52.0	55	53.5	38.0	99	18.8	28.0
Stuttgart	173	13.3	22.0	119	80.8	54.0	46	31.0	40.5	97	14.8	19.0
Wuhan	111	79.0	119.0	116	116.0	90.0	60	51.0	45.0	65	38.8	45.0
Shenyang	128	74.0	114.0	113	181.3	242.0	59	482.5	179.0	58	34.0	34.0
Kyoto	108	13.0	29.0	99	36.3	41.0	32	18.5	21.5	35	9.5	10.0
Seoul	188	15.5	23.5	140	95.0	77.5	34	1254.0	841.0	47	15.0	16.0
Nagoya	94	12.0	4.0	63	15.0	5.0	38	12.0	3.0	45	12.0	4.0
Cairo	113	28.3	48.0	109	19.5	28.0	51	243.5	56.0	49	104.8	81.0
Aswan	127	8.0	19.0	108	7.0	20.0	41	9.0	16.0	90	11.0	23.0
Istanbul	152	24.0	27.5	133	77.0	49.0	25	58.8	58.0	45	30.0	31.0
Yerevan	65	15.5	29.0	54	3.0	11.0	20	5.0	10.0	48	15.5	29.0
Manaus	150	8.0	17.0	112	93.0	86.0	20	78.5	27.5	38	12.0	13.0
Iquitos	83	12.8	17.0	82	66.0	73.0	51	139.3	127.0	45	6.0	8.0
Lagos	99	20.5	33.0	51	80.3	45.0	23	33.0	16.0	24	12.0	18.0
Accra	112	4.0	17.0	91	7.0	20.0	81	7.0	20.0	61	4.3	18.0
Chongqing	111	33.8	57.0	98	53.0	56.0	41	1247.0	683.0	65	22.8	27.0
Guangzhou	301	52.3	77.0	135	731.0	232.0	191	311.8	178.0	145	29.0	35.0
Bangkok	251	11.0	20.0	186	24.0	33.0	159	114.5	55.0	93	18.5	20.0
New Delhi	233	19.0	29.0	144	22.0	25.5	80	114.5	44.0	87	17.8	17.0
Calcutta	120	21.5	11.0	73	14.0	9.0	105	22.0	15.3	36	16.0	11.5
Curitiba	227	10.0	18.0	161	41.0	52.0	48	546.5	107.0	98	12.0	16.0
Sao Paulo	170	14.0	27.5	98	15.0	17.0	34	23.0	22.0	27	8.5	9.0
Addis Ababa	115	13.0	21.0	101	12.0	17.0	12	109.0	116.5	132	11.0	20.0
Harare	99	6.0	15.0	74	6.0	15.0	45	3.0	14.0	108	6.0	15.5
Johannesburg	121	6.3	18.0	92	9.0	17.0	17	37.3	22.0	94	11.0	13.0
Pretoria	99	8.8	23.0	63	5.8	22.0	27	8.5	20.0	45	5.3	20.0
Milan	321	15.0	26.0	191	36.5	45.0	86	70.0	58.0	141	33.0	34.0
Los Angeles	255	19.0	29.0	105	22.0	20.0	40	15.0	8.0	99	18.0	18.0
Las Vegas	179	30.8	37.0	115	20.0	24.0	50	24.0	26.5	90	160.0	140.0
Phoenix	306	9.0	16.0	198	6.0	22.0	144	5.0	10.0	207	11.0	18.0
Xi'an	140	144.0	168.0	107	600.3	71.0	48	341.5	132.0	120	85.5	99.0
Xining	119	9.0	25.0	108	7.0	25.0	42	5.0	12.0	90	9.0	24.0

Count: number of ROIs in second reference dataset; Median, IQR: median value and interquartile range of the number of pixels within the ROI.

paper, we set this parameter as 2000 for all the study sites, considering both the computational cost and performance.

3) Sensitivity of Parameter T_c: In Fig. 12, the effect of the termination threshold T_c ($c \in$ {bright BU, V, WS, BS, dark BU}) of the automatic sample collection is analyzed, by taking the three cities as examples. With regard to the reference ROIs (see Table X), the bright BU and the dark BU classes were merged as one class, and the termination thresholds for these two subclasses were set as the same. We fixed the other parameters as the optimal values and focused on the influence of T_c, as shown in Fig. 12. The horizontal axis denotes the termination threshold for each land-cover type, and the vertical axis shows the corresponding OA. First, the flat curves of V and WS can be attributed to their relatively homogenous spectral characteristics. With respect to BU, the performance for this land-cover type is also relatively stable. BS shows the largest variation, and it is suggested that a suitable threshold would be around 100, as a small value of T_c can degrade the diversity of the collected samples.

4) Sensitivity of the Threshold for the First Stage: In the following, we examine the effect of the threshold for the first stage, i.e., the approach designed for SDBI generation, on the classification performance of the proposed algorithm. According to the large variance of the MNDWI image, it is able to highlight the strong contrast between water and nonwater surfaces [8]. In this context, it is reasonable to set the threshold of the first stage to a small number of iterations.

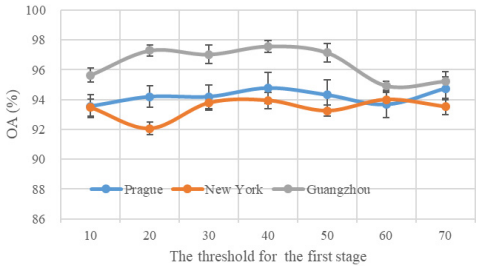


Fig. 13. Classification accuracy versus the threshold for the first stage for the Prague, New York, and Guangzhou Landsat images.

Taking Prague, New York, and Guangzhou as instances, as shown in Fig. 13, the horizontal axis denotes the threshold for the first stage, and the vertical axis shows the corresponding OA. From Fig. 13, it can be seen that the performances for all the test sites are relatively stable as the threshold increases. In short, it is reasonable to set the threshold as 50.

5) Regularization Parameter λ in CRC: The effect of the regularization parameter λ on the classification performance for the three test sites is shown in Fig. 14. The experiments were again repeated 10 times using different randomly chosen training sets to reduce the possible bias induced by the random sampling. The horizontal axis shown in Fig. 11 is the value range of λ, while the vertical axis shows the OA. In Fig. 14, all the curves are quite stable and show a pleasing accuracy

TABLE X

STATISTICS OF THE MANUALLY COLLECTED ROIS USED IN THE SUPERVISED CLASSIFICATION COMPARISON EXPERIMENTS

Study site	Number of ROIs											
	BU			V			WS			BS		
	Count	Median	IQR	Count	Median	IQR	Count	Median	IQR	Count	Median	IQR
New York	45	17.3	23.0	16	47.0	54.0	6	20.0	39.5	7	8.5	11.0
Houston	27	37.3	46.0	21	73.8	78.0	19	26.3	26.0	8	32.0	22.5
Montreal	18	23.0	26.5	11	26.0	27.0	3	92.3	14.0	3	9.8	18.0
Melbourne	23	31.8	28.0	15	95.5	81.0	4	10.0	9.5	6	40.0	29.5
Christchurch	10	28.0	27.5	11	29.3	35.0	2	101.0	55.5	8	29.0	25.5
Marseille	11	22.8	22.0	7	55.8	49.0	2	9.0	22.5	6	8.0	7.5
Prague	16	17.5	24.5	11	135.3	49.0	4	27.5	59.0	6	49.0	43.5
Frankfurt	15	17.5	22.0	12	48.0	48.5	6	66.0	43.0	11	36.0	28.0
Stuttgart	19	10.8	19.0	13	73.5	74.0	5	34.0	14.0	10	8.0	19.0
Wuhan	12	79.0	103.5	11	116.8	94.0	6	15.0	36.5	7	91.0	38.0
Shenyang	14	51.0	120.0	12	139.0	232.5	6	226.0	167.5	6	39.0	34.0
Kyoto	12	22.0	29.0	11	37.5	53.0	4	41.0	23.0	4	15.5	10.0
Seoul	20	14.5	29.0	15	47.8	63.0	3	1516.5	606.0	5	6.0	22.0
Nagoya	11	11.0	3.5	8	14.5	5.0	5	14.5	10.5	6	11.5	3.0
Cairo	12	29.0	51.0	12	21.5	29.0	5	634.5	447.0	5	197.3	143.0
Aswan	15	5.0	19.0	12	8.5	22.5	5	4.3	13.0	10	14.0	24.5
Istanbul	17	17.5	31.0	14	155.0	41.5	2	167.0	143.5	5	33.0	25.0
Yerevan	8	14.5	31.5	7	5.5	9.0	3	6.0	11.0	6	10.0	29.0
Manaus	16	8.0	18.0	12	116.0	84.5	2	13.0	16.5	4	12.0	19.0
Iquitos	9	6.3	12.0	9	52.8	44.0	4	48.0	181.5	5	5.0	11.0
Lagos	11	29.3	31.0	5	169.0	150.0	2	67.0	43.5	2	6.0	10.0
Accra	13	6.3	20.0	11	7.0	22.0	10	6.0	20.5	7	8.3	18.0
Chongqing	12	18.5	49.0	10	29.0	59.0	4	872.5	316.0	7	6.0	30.0
Guangzhou	33	54.5	72.0	15	486.8	311.0	21	463.8	97.0	16	44.0	39.0
Bangkok	28	8.5	18.5	20	18.5	27.5	17	137.3	34.0	10	10.0	27.0
New Delhi	26	25.0	28.5	16	27.5	36.0	7	75.0	55.0	9	12.8	18.0
Calcutta	14	13	21.5	9	13.75	22	12	10	21.5	4	6.5	18
Curitiba	25	12.5	23.0	18	45.0	64.0	4	143.5	15.5	11	10.0	15.0
Sao Paulo	18	13.0	25.0	10	14.0	25.0	3	14.3	28.0	3	8.3	12.0
Addis Ababa	12	8.5	18.0	11	16.5	16.0	1	0.0	175.0	14	15.0	20.0
Harare	12	6.0	21.0	9	6.5	18.0	5	6.5	16.0	12	3.0	15.5
Johannesburg	13	6.5	17.0	10	9.0	15.0	1	0.0	75.0	10	6.0	14.5
Pretoria	11	6.3	20.0	7	4.8	20.0	4	4.0	15.0	5	11.0	23.0
Milan	35	16.0	31.0	21	44.8	43.0	9	68.0	62.0	15	36.0	41.0
Los Angeles	29	13.3	29.0	12	11.0	16.5	5	50.0	25.0	11	18.3	20.0
Las Vegas	19	37.8	40.0	11	17.8	27.0	5	34.8	23.0	10	57.0	73.5
Phoenix	34	11.0	14.5	23	5.8	22.0	16	6.0	9.0	23	8.5	15.0
Xi'an	15	102.3	134.0	11	95.8	112.0	5	128.0	82.0	13	88.3	118.0
Xining	14	6.0	27.5	12	9.0	31.5	5	9.0	7.0	10	16.0	31.0

Count: number of ROIs in the manually collected ROIs used in the supervised classification comparison experiments; Median, IQR: median value and interquartile range of the number of pixels within the ROI.

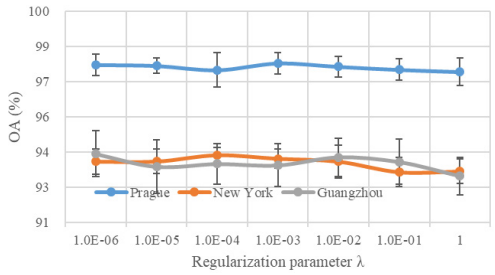


Fig. 14. Classification accuracy versus regularization parameter λ for the Prague, New York, and Guangzhou Landsat images.

when the regularization parameter λ ranges from $1e^{-6}$ to $1e^{-2}$, which suggests the validity of setting λ as $1e^{-3}$.

E. Run Time Analysis

Taking Prague (~520 km^2), New York (~3865 km^2), and Guangzhou (~6438 km^2) as instances, the detailed run times for both the proposed method and the simple sampling method are listed in Table VII. For the simple sampling method, the computational cost of setting the thresholds by trial and error, which is a time-consuming and subjective matter, is not included. For the proposed method, the stopping condition for all three test sites was 400 iterations, and the size of the training sample set was 500 for each class. In Table VII, it can be seen that the simple sampling method is faster than the proposed method for the two smaller test sites, but it takes more time for the Guangzhou image with the largest size. For the Guangzhou image, the proposed method is not only faster than the simple sampling method, but it also presents a superior classification result (proposed method: OA = 98%; simple sampling method: OA = 81%). In more detail, it is noted that the computational cost of the simple sampling method increases exponentially with the image size, while the computational burden of the proposed method increases relatively slowly as the image size grows. The burdensome computational cost of the simple sampling method mainly comes from the huge size of the collected sample sets. In general, although the proposed method does require some computing time, it is, however, reasonable to believe that with the rapid development in computer hardware, the time cost of the proposed method will soon no longer be an issue.

V. CONCLUSION

In this paper, we have proposed a novel unsupervised sample selection method for urban land-cover mapping. The proposed method was effectively implemented in 39 representative cities located in different eco-regions across the world, using Landsat TM, ETM+, and OLI images. In these study areas, the OA ranged from 84.3% to 99% and the Kappa coefficient ranged from 0.77 to 0.99, demonstrating the potential for automatic urban mapping. The sample accuracy was taken into account to validate the effectiveness of the proposed method. When compared with the results of supervised classification

with manually selected training samples, the proposed method obtained a comparable and promising accuracy. Furthermore, the experimental results showed that the proposed method significantly outperformed the use of the simple index-based threshold for different land-cover classes.

In practice, land-cover mapping, especially large-scale mapping, depends on data availability, time and cost constraints, and the efficiency of the processing scheme. With the proposed sample selection scheme, by the use of simply calculated and widely used spectral indices, it is possible to automatically collect reliable, diverse, and concise sample sets, which are essential for gaining an accurate classification result. Subsequently, using CRC, the urban areas can be efficiently mapped into four basic land-cover types using the automatically collected samples, and the final mapping results will be comparable to those of the manual sampling approach. As vegetation, impervious surfaces, soil, and water surfaces can be regarded as the four basic biophysical components of the urban environment, their automatic extraction will be of great interest for urban planning and studies of urbanization and the urban environment. It is noted that some subdivision of these basic land covers may be of further interest, but this is beyond the scope of this paper. Based on the proposed sample collection procedure, given an effective index, it has the potential to extract the subtypes of the land cover. In short, this approach shows a promising potential for automatic and global mapping of urban areas.

APPENDIX

See Tables VIII–X.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful and constructive suggestions.

REFERENCES

[1] J. Liu, H. Tian, M. Liu, D. Zhuang, J. M. Melillo, and Z. Zhang, "China's changing landscape during the 1990s: Large–scale land transformations estimated with satellite data," *Geophys. Res. Lett.*, vol. 32, pp. 1–5, 2005.

[2] X. Bai, J. Chen, and P. Shi, "Landscape urbanization and economic growth in China: Positive feedbacks and sustainability dilemmas," *Environ. Sci. Technol.*, vol. 46, no. 1, pp. 132–139, Nov. 2011.

[3] C. He, Z. Liu, J. Tian, and Q. Ma, "Urban expansion dynamics and natural habitat loss in China: A multiscale landscape perspective," *Global Change Biol.*, vol. 20, no. 9, pp. 2886–2902, Sep. 2014.

[4] L. Zhou *et al.*, "Evidence for a significant urbanization effect on climate in China," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 26, pp. 9540–9544, Jun. 2004.

[5] X. Li and P. Gong, "An 'exclusion-inclusion' framework for extracting human settlements in rapidly developing regions of China from Landsat images," *Remote Sens. Environ.*, vol. 186, pp. 286–296, Dec. 2016.

[6] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 583–594, 2003.

[7] X. Huang, C. Xie, X. Fang, and L. Zhang, "Combining pixel- and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2097–2110, May 2015.

[8] H. Xu, "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3025–3033, Jan. 2006.

[9] X. Huang, H. Liu, and L. Zhang, "Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3639–3657, Jul. 2015.

[10] W. Chen, L. Liu, C. Zhang, J. Wang, J. Wang, and Y. Pan, "Monitoring the seasonal bare soil areas in Beijing using multitemporal TM images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2004, pp. 3379–3382.

[11] L. Yang, C. Huang, C. G. Homer, B. K. Wylie, and M. J. Coan, "An approach for mapping large-area impervious surfaces: Synergistic use of Landsat-7 ETM+ and high spatial resolution imagery," *Can. J. Remote Sens.*, vol. 29, no. 2, pp. 230–240, 2003.

[12] M. Chellasamy, T. P. A. Ferré, and M. H. Greve, "An ensemble-based training data refinement for automatic crop discrimination using worldview-2 imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4882–4894, Oct. 2015.

[13] C. Deng and C. Wu, "BCI: A biophysical composition index for remote sensing of urban environments," *Remote Sens. Environ.*, vol. 127, pp. 247–259, Dec. 2012.

[14] X. Huang, C. Weng, Q. Lu, T. Feng, and L. Zhang, "Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas," *Remote Sens.*, vol. 7, no. 12, pp. 16024–16044, Dec. 2015.

[15] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sens.*, vol. 9, no. 1, p. 95, Jan. 2017.

[16] G. Xian, C. Homer, and J. Fry, "Updating the 2001 national land cover database land cover classification to 2006 by using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1133–1147, Jun. 2009.

[17] C. Homer *et al.*, "Completion of the 2001 national land cover database for the counterminous United States," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 4, p. 337, Apr. 2007.

[18] J. Fry *et al.*, "Completion of the 2006 national land cover database for the conterminous United States," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 9, pp. 858–864, 2011.

[19] J. Chen *et al.*, "Global land cover mapping at 30 m resolution: A POK-based operational approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 7–27, May 2015.

[20] D. Lu, G. Li, E. Moran, M. Batistella, and C. C. Freitas, "Mapping impervious surfaces with the integrated use of Landsat Thematic Mapper and radar data: A case study in an urban–rural landscape in the Brazilian Amazon," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 6, pp. 798–808, Nov. 2011.

[21] G. Sun, X. Chen, X. Jia, Y. Yao, and Z. Wang, "Combinational build-up index (CBI) for effective impervious surface mapping in urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 2081–2092, May 2016.

[22] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. (2012). "Collaborative representation based classification for face recognition." [Online]. Available: https://arxiv.org/abs/1204.2358

[23] P. Zhu, W. Zuo, L. Zhang, S. C. K. Shiu, and D. Zhang, "Image set-based collaborative representation for face recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 7, pp. 1120–1132, Jul. 2014.

[24] T. Guha and R. Ward, 'Learning sparse representations for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1576–1588, Aug. 2012.

[25] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[26] S. Jia, L. Shen, and Q. Li, "Gabor feature-based collaborative representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1118–1129, Feb. 2015.

[27] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.

[28] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.

[29] A. Schneider, M. A. Friedl, and D. Potere, "Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions,"' *Remote Sens. Environ.*, vol. 114, no. 8, pp. 1733–1746, Aug. 2010.

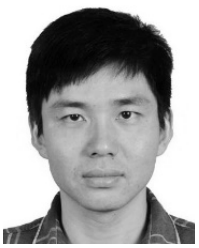
[30] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, Mar. 2012.

[31] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," *Nature*, vol. 540, pp. 418–422, Dec. 2016.
[32] C. Pelletier, S. Valero, J. Inglada, N. Champion, C. M. Sicre, and G. Dedieu, "Effect of training class label noise on classification performances for land cover mapping with satellite image time series," *Remote Sens.*, vol. 9, no. 2, p. 173, Feb. 2017.
[33] I. L. Thomas, N. P. Ching, V. M. Benning, and J. D'Aguanno, "Review article a review of multi-channel indices of class separability," *Int. J. Remote Sens.*, vol. 8, no. 3, pp. 331–350, 1987.
[34] S. M. Davis *et al.*, *Remote Sensing: The Quantitative Approach*. New York, NY, USA: McGraw-Hill, 1978. p. 405.
[35] D. Chen and D. Stow, "The effect of training strategies on supervised classification at different spatial resolutions," *Photogramm. Eng. Remote Sens.*, vol. 68, no. 11, pp. 1155–1162, Nov. 2002.
[36] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-9, no. 1, pp. 62–66, Jun. 1979.
[37] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.
[38] A. Schneider, "Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach," *Remote Sens. Environ.*, vol. 124, pp. 689–704, Sep. 2012.
[39] R. Goldblatta *et al.*, "Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover," *Remote Sens. Environ.*, vol. 205, pp. 253–275, Feb. 2018.
[40] C.-C. Chang and C.-J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[41] X. Huang, X., Han, L. Zhang, J. Gong, W. Liao, and J. A. Benediktsson, "Generalized differential morphological profiles for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1736–1751, Apr. 2016.
[42] B. S. C. Wade, S. H. Joshi, B. Gutman, and P. M. Thompson, "Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods," *Pattern Recogn.*, vol. 63, pp. 731–739, Mar. 2017.
[43] T. Hu, X. Huang, J. Li, and L. Zhang, "A novel co-training approach for urban land cover mapping with unclear Landsat time series imagery," *Remote Sens. Environ.*, 217, pp. 144–157, Nov. 2018.
[44] J. Munoz-Mari, D. Tuia, and G. Camps-Valls, "Semisupervised classification of remote sensing images with active queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, Oct. 2012.
[45] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and M. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics. Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

Jiayi Li (S'12–M'15) received the B.S. degree from Central South University, Changsha, China, in 2011, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2016.

She is currently an Assistant Professor with the School of Remote Sensing and Information Engineering, Wuhan University. She has authored over 30 peer-reviewed articles (SCI papers) in the international journals. Her research interests include hyperspectral imagery, sparse representation, computation vision and pattern recognition, and remote sensing images.

Dr. Li is a Reviewer for more than 10 international journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE SIGNAL PROCESSING LETTERS, and the *International Journal of Remote Sensing*. She is the Guest Editor of the special issue on *Change Detection Using MultiSource Remotely Sensed Imagery for the Remote Sensing* (An open access journal from MDPI).



Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, where he is also a Luojia Distinguished Professor and he teaches remote sensing, photogrammetry, and image interpretation. He is the Founder and the Director with the Institute of Remote Sensing Information Processing, School of Remote Sensing and Information Engineering, Wuhan University. He has authored over 130 peer-reviewed articles (SCI papers) in the international journals. His research interests include remote sensing image processing methods and applications.

Dr. Huang has been supported by The National Program for Support of Top-notch Young Professionals in 2017, the China National Science Fund for Excellent Young Scholars in 2015, and the New Century Excellent Talents in University from the Ministry of Education of China in 2011. He was a recipient of the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) in 2010, the second place recipient for the John I. Davidson President's Award from ASPRS in 2018, 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 the Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the Winner of the IEEE GRSS 2014 Data Fusion Contest. He was the Lead Guest Editor of the special issue on *Information Extraction From High-Spatial-Resolution Optical Remotely Sensed Imagery* for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (vol. 8, no. 5, May 2015), and the special issue on *Sparsity-Driven High Dimensional Remote Sensing Image Processing and Analysis* for the *Journal of Applied Remote Sensing* (vol. 10, no. 4, Oct 2016). He has been serving as an Associate Editor for *the Photogrammetric Engineering and Remote Sensing* since 2016, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2014, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING since 2018. He has been an Editorial Board Member of *Remote Sensing* (an open access journal from MDPI) since 2018.

Ting Hu (S'18) received the B.S. and M.S. degrees from Wuhan University, Wuhan, China, in 2012 and 2015, respectively, where she is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering.

Her research interests include machine learning, remote sensing mapping, and remote sensing applications.

Xiuping Jia (M'93–SM'03) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from The University of New South Wales, Sydney, NSW, Australia, in 1996.

Since 1988, she has been with the School of Engineering and Information Technology, The University of New South Wales at Canberra, Canberra, ACT, Australia, where she is currently an Associate Professor. She has authored or co-authored over 200 referred papers, including over 100 journal papers. She has co-authored the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd (1999) and 4th eds. (2006)]. Her research interests include remote sensing, image processing, and spatial data analysis.

Dr. Jia is a Subject Editor of the *Journal of Soils and Sediments* and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Jón Atli Benediktsson ((S'86–M'90–SM'99–F'04)) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

In 2015, he became the President and the Rector at the University of Iceland. From 2009 to 2015, he was the Pro Rector of Science and Academic Affairs. Since 1991, he has been with the University of Iceland and has been a Professor of electrical and computer engineering since 1996. He is on the 2018 list of the Highest Cited Researchers. He has authored extensively in his research fields. His research interests include remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing.

Dr. Benediktsson is a fellow of SPIE. He was a member of the 2014 IEEE Fellow Committee. He received the Stevan J. Kristof Award from Purdue University in 1991 as an Outstanding Graduate Student in remote sensing. He was a recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997, he was granted the IEEE Third Millennium Medal in 2000, and he was a co-recipient of the University of Iceland's Technology Innovation Award in 2004. He received the Yearly Research Award from the Engineering Research Institute of the University of Iceland in 2006, the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007, and the OECE Award from the School of ECE, Purdue University in 2016. He was a co-recipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award and the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Highest Impact Paper Award. He received the IEEE/VFI Electrical Engineer of the Year Award in 2013. He was a co-recipient of the International Journal of Image and Data Fusion Best Paper Award in 2014 and 2018. He received the David A. Landgrebe Award in 2018 from the IEEE GRSS for outstanding contributions in the field of remote sensing image analysis. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scinetiarum Islandica, and Tau Beta Pi. He was the 2011–2012 President of the IEEE GRSS and has been on the GRSS AdCom since 2000. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has served as an Associate Editor for TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and the IEEE ACCESS since 2013. He is on the Editorial Board of the PROCEEDINGS OF THE IEEE, the International Editorial Board of the *International Journal of Image and Data Fusion*, the Editorial Board of the *Remote Sensing*, and was the Chairman of the Steering Committee of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2007 to 2010. He is a Co-Founder of the biomedical startup company Oxymap (www.oxymap.com).