ISPRS Journal of Photogrammetry and Remote Sensing 90 (2014) 36–48

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas

Xin Huang *, Qikai Lu, Liangpei Zhang

The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

ARTICLE INFO

Article history:
Received 4 June 2013
Received in revised form 10 January 2014
Accepted 19 January 2014

Keywords:
High spatial resolution
Classification
SVM
Morphological
Texture
Feature extraction

ABSTRACT

In recent years, it has been widely agreed that spatial features derived from textural, structural, and object-based methods are important information sources to complement spectral properties for accurate urban classification of high-resolution imagery. However, the spatial features always refer to a series of parameters, such as scales, directions, and statistical measures, leading to high-dimensional feature space. The high-dimensional space is almost impractical to deal with considering the huge storage and computational cost while processing high-resolution images. To this aim, we propose a novel multi-index learning (MIL) method, where a set of low-dimensional information indices is used to represent the complex geospatial scenes in high-resolution images. Specifically, two categories of indices are proposed in the study: (1) *Primitive indices* (*PI*): High-resolution urban scenes are represented using a group of primitives (e.g., building/shadow/vegetation) that are calculated automatically and rapidly; (2) *Variation indices* (*VI*): A couple of spectral and spatial variation indices are proposed based on the 3D wavelet transformation in order to describe the local variation in the joint spectral-spatial domains. In this way, urban landscapes can be decomposed into a set of low-dimensional and semantic indices replacing the high-dimensional but low-level features (e.g., textures). The information indices are then learned via the multi-kernel support vector machines. The proposed MIL method is evaluated using various high-resolution images including GeoEye-1, QuickBird, WorldView-2, and ZY-3, as well as an elaborate comparison to the state-of-the-art image classification algorithms such as object-based analysis, and spectral-spatial approaches based on textural and morphological features. It is revealed that the MIL method is able to achieve promising results with a low-dimensional feature space, and, provide a practical strategy for processing large-scale high-resolution images.

1. Introduction

High-resolution remotely sensed data can provide a large amount of detailed ground information, and open new avenues for remote sensing applications such as precise land use/land cover mapping, landscape analysis, urban facility retrieval, regional environment monitoring. However, increase of the spatial resolution does not signify increase of the image processing accuracy. Specifically, increase of the intra-class variation and decrease of the inter-class variation lead to reduction of the separability of the spatial patterns in the spectral domain (Bruzzone and Carlin, 2006). As a result, the traditional pixel-based and spectral-based image classification techniques are inadequate for high-resolution data (Huang and Zhang, 2013). In this context, researchers proposed to exploit spatial information for complementing the spectral feature space and enhancing separability of the spectrally similar classes. The so-called spectral-spatial classification methods can be divided into the following two categories.

1.1. Exploration of spatial features

In this case, the structural and textural features are used as additional bands to enhance the spectral information and raise accuracies of high-resolution image classification. The commonly used spatial bands include wavelet textures (Ouma and Tateishi, 2008), gray-level co-occurrence matrix (GLCM) (Aguera et al., 2008), pixel shape index (Huang et al., 2007), geometric image features (Inglada 2007), Gabor textural features (Reis and Tasdemir, 2011), morphological filters (Pingel et al., 2013), and morphological profiles (Fauvel et al., 2012).

* Corresponding author. Tel.: +86 27 68778525.
E-mail address: huang_whu@163.com (X. Huang).

http://dx.doi.org/10.1016/j.isprsjprs.2014.01.008
0924-2716/

On the other hand, in recent years, the integrative algorithms combining multiple features have received increasing interest since a single feature descriptor is not adequate for representing complex high-resolution urban scenes. Huang and Zhang (2013) proposed to integrate the spectral, structural, and semantic features for classification of multi/hyperspectral imagery with high spatial resolution. Dell'Acqua et al. (2009) exploited boundary, textural, and morphological features for rapid mapping of high resolution SAR scenes. Fusion of LiDAR and optical data also received much attention for urban scene classification (Guo et al., 2011) and residential building detection (Awrangjeb et al., 2010), since the height information extracted from LiDAR is a useful feature source for urban images.

1.2. Object-based image analysis (OBIA)

Its basic idea is to segment the spatially adjacent pixels into spectrally similar homogeneous objects and then conduct image analysis on objects as the minimum processing unit (Blaschke, 2010). Compared to the pixel-based approach, advantages of the OBIA mainly lie in that it is able to reduce the local spectral variation and intra-class variance, and, hence, avoid the salt-pepper effect. OBIA is an active research area for remote sensing image interpretation. Some studies focus on segmentation, which is the core step of OBIA, such as adaptive mean-shift procedure (Huang and Zhang, 2008), anisotropic morphological leveling (Tzotsos et al., 2011), boundary-constrained multi-scale segmentation (Zhang et al., 2013). Other relevant literature refers to scale parameter selection (Liu et al., 2012) and optimization (Johnson and Xie, 2011) in order to adapt the segmentation. In recent years, OBIA has been successfully applied to mapping private gardens (Mathieu et al., 2007), urban objects extraction (Sebari and He, 2013), change detection (Hussain et al., 2013), tree crown delineation (Jing et al., 2012), spatial pattern analysis of vegetation cover (Yang et al., 2013), etc.

In spite of promising progress achieved for high-resolution image processing, few studies have paid attention to the practical techniques for information extraction from large-scale high-resolution remote sensing data. The spectral-spatial approach, which is the existing mainstream strategy for high resolution image classification, is always subject to the problem of high-dimensional feature space, constructed by a series of parameters such as scales, directions, statistical measures, and basis images. The high-dimensional space poses a big challenge to storage and computation cost for applications of high-resolution data.

In order to address this problem, we propose an innovative multi-index learning (MIL) method for urban scene classification based on high-resolution imagery. The notable characteristic of the MIL is to describe the urban landscape using a set of low-dimensional semantic indices that replace the high-dimensional and low-level features. The first step of the MIL framework is to decompose urban scenes into a series of information indices. In this study, two categories of indices are proposed.

1) Primitive indices (PI): Urban landscapes can be represented by a group of basic elements such as buildings, shadow, and vegetation. These basic elements are calculated automatically without training samples via the morphological building index (MBI) (Huang and Zhang, 2011), morphological shadow index (MSI) (Huang and Zhang, 2012), and normalized difference vegetation index (NDVI). It should be underlined that these indices actually provide more information than the three classes. They also contain information for non-buildings, non-shadow, and non-vegetation, respectively. Therefore, these primitive indices are potential for description of the semantic feature space in urban image scenes.
2) Variation indices (VI): Spectral and spatial variation indices are extracted by 3D wavelet transformation (3D-WT). The notable property of 3D-WT is that it processes an image as a cube, and, therefore, simultaneously describes variation information in the joint spectral-spatial feature space (Yoo et al., 2009). Accordingly, based on the 3D-WT, we propose a couple of variation indices as a representation of spectral and spatial information in a local image scene.

PI and VI are used to describe stationary (basic elements) and dynamic (variation information) features, respectively. Subsequently, we propose to use the multi-kernel learning (Tuia et al., 2010) for interpretation of the multi-index features. The proposed multi-index learning (MIL) method is tested through a series of sophisticated experiments, conducted on the GeoEye-1, QuickBird, WorldView-2, and ZY-3 images. The first three datasets are used for evaluation of MIL method by a detailed comparison to the existing state-of-the-art high-resolution image classification methods, e.g., object-based image analysis (Bruzzone and Carlin, 2006; Blaschke, 2010), spectral-spatial classification using multiscale and multidirectional gray level co-occurrence matrix (GLCM) (Pesaresi et al., 2008), and differential morphological profiles (DMP) (Pesaresi and Benediktsson, 2001). Furthermore, performance of the MIL method is then assessed using a large-scale high-resolution data, ZY-3 satellite image over the urban area of Wuhan city (260 km^2) in central China. In this case, the high-dimensional textural and structural features (e.g., GLCM, DMP) are impractical due to the limitation of memory and processor of a personal computer. In addition, ZY-3 satellite, launched on 9th January 2012, is the China's first civilian high-resolution satellite. To our best knowledge, this study is the first assessment of ZY-3 satellite imagery for information extraction and urban mapping.

The remainder of this paper is organized as follows. The multi-index feature extraction is described in Section 2. Then, the multi-index learning is introduced in Section 3. Experimental results and the comparative study are presented in Section 4, followed by the conclusions and remarks in Section 5.

2. Multi-index feature extraction

This section describes the primitive and variation indices, as well as a demonstration of the multi-index urban scene description.

2.1. Primitive indices (PI)

PI includes a group of basic urban classes: buildings, shadow, and vegetation, which are automatically calculated using MBI, MSI, and NDVI, respectively. Information for other primitive classes (e.g., roads, water) is implicitly contained in the three indices. For instance, MBI is able to discriminate between buildings and roads, while water and shadow can be distinguished based on MSI and NDVI.

2.1.1. Morphological building index (MBI)

The basic idea of MBI is to build the relationship between the spectral-spatial characteristics of buildings (e.g., local contrast, size, isotropy, and brightness) and the morphological operators (Huang and Zhang, 2011). It is constructed based on the fact that the relatively high reflectance of roofs and the spatially adjacent shadows lead to high local contrast of buildings (Pesaresi et al., 2008; Huang and Zhang, 2011). The calculation of MBI is briefly described as follows.

Step1. Brightness image. The maximum value of the visible bands for each pixel is recorded as its brightness. The visible bands are focused on since they have the most significant contribution to the spectral property of buildings (Pesaresi et al., 2008).

Step2. Top-hat morphological profiles. The spectral-structural characteristics of buildings are represented using the differential morphological profiles (Pesaresi and Benediktsson, 2001) of the top-hat transformation (DMP_{TH}),

$$DMP_{TH} = \{\Delta TH(d,s), d \in D, s \in S\}$$

with

$$\Delta TH(d,s) = |TH(d,s) - TH(d,s-\Delta s)| \quad (1)$$

where TH represents the top-hat by reconstruction (Huang and Zhang, 2011) of the brightness image; s and d indicate the scale and direction of a linear structural element, respectively; Δs is the interval of the profiles; S and D, respectively, represent the sets of scales and directions. The top-hat transform is used to measure the contrast as it is able to highlight the locally bright structures with a size up to a predefined value.

Step3. Calculation of MBI. It is defined by

$$MBI = \frac{\sum_d \sum_s DMP_{TH}(d,s)}{N_d \times N_s} \quad (2)$$

where N_d and N_s are the numbers of directions and scales, respectively. The average of the multiscale and multidirectional DMP_{TH} is defined as the building index since building structures have larger feature values in most of the scales and directions in the morphological profiles due to their local contrast and isotropy. As a result, the structures with large MBI values are more likely to be buildings. A challenging task for building detection is to automatically filter out roads that have similar spectral reflectance with buildings. Roads are always elongated in one or two directions while buildings are more isotropic. To this aim, in Eq. (2), MBI is implemented using a series of linear structural elements that are able to measure the directionality of local structures.

2.1.2. Morphological shadow index (MSI)

MSI is a twinborn index of MBI since shadows show spatially similar but spectrally contrary properties with buildings (Huang and Zhang, 2012). The calculation of MSI is based on the fact that shadows have high local contrast but low radiometric reflectance. Consequently, the black top-hat (BTH) is used to construct the shadow index in contrast to the white top-hat for the building index,

$$MSI = \frac{\sum_d \sum_s DMP_{BTH}(d,s)}{N_d \times N_s} \quad (3)$$

where the BTH is able to highlight the dark structures within the defined directions and scales.

2.1.3. Normalized difference vegetation index (NDVI)

NDVI is used to represent the urban vegetation components such as grass and trees. Its calculation is based on the principal that vegetation has a strong reflectance in the near infrared channel but a strong absorption in the red channel.

2.2. Variation indices (VI)

In order to complement the stationary indices (i.e., PI), in this study, we also propose a couple of variation indices, used to describe spectral and spatial variation information in a local image area. VI is defined based on the 3D wavelet transformation, which processes a multispectral remote sensing image as a cube (Yoo et al., 2009). 3D-WT can be expressed by a tensor product,

$$I^{(x,y,z)} = (L^x \oplus H^x) \otimes (L^y \oplus H^y) \otimes (L^z \oplus H^z) = \begin{cases} L^x L^y L^z \oplus L^x L^y H^z \oplus L^x H^y L^z \oplus L^x H^y H^z \\ \oplus H^x L^y L^z \oplus H^x L^y H^z \oplus H^x H^y L^z \oplus H^x H^y H^z \end{cases} \quad (4)$$

where $\oplus$ is the space direct sum, L and H represent the low- and high-pass filters along the x, y, and z axis, respectively. x and y denote the spatial coordinates of an image, and z stands for the spectral axis. Eq. (4) shows that one-level 3D-WT decomposition for an image cube yields a series of subbands, consisting of the following three blocks:

1) Approximation: LLL.
2) Spectral variation (**H): LLH, LHH, HLH.
3) Spatial variation (**L): LHL, HLL, HHL.

where the **H components denote spectral variation since the high-pass filter is used in the spectral domain, and the **L components represent the spatial variation since they describe high-frequency details (horizontal, vertical, and diagonal information) in the spatial domain and low-frequency information in the spectral domain. Accordingly, for a pixel i in a local image cube w, its spectral and spatial variation indices (VI_{spe} and VI_{spa}, respectively) are defined by

$$VI_{spe}(i) = \frac{1}{N_w} \sum_w \left\{ \frac{E^{LLH}_{i \in w}(i) + E^{LHH}_{i \in w}(i) + E^{HLH}_{i \in w}(i)}{E^{LLL}_{i \in w}(i)} \right\}_w \quad (5)$$

$$VI_{spa}(i) = \frac{1}{N_w} \sum_w \left\{ \frac{E^{HLL}_{i \in w}(i) + E^{LHL}_{i \in w}(i) + E^{HHL}_{i \in w}(i)}{E^{LLL}_{i \in w}(i)} \right\}_w \quad (6)$$

where E is the energy function, which is defined as the quadratic sum of the 3D wavelet coefficients for each subband. It can be seen that Eqs. (5) and (6) describe the normalized '**H' and '**L' components in a local image cube, representing the spectral and spatial variation indices, respectively. The energy of the LLL subband is used for normalization considering that the approximation component in the 3D-WT decomposition contains majority of the energy of wavelet coefficients. As shown in Eqs. (5) and (6), a multiscale approach is utilized for calculation of VI by considering multiple local image cubes centered by the pixel i. The multiscale approach is able to alleviate the boundary blur caused by the moving window processing, and meanwhile take advantage of multiscale characteristics in the high-resolution urban scenes. It should be noted that in order to reduce the computational cost, the variation indices are calculated with non-overlapped windows. Unlike other textural or spatial features that only focus on the spatial domain, VI is able to simultaneously describe the spectral-spatial joint feature space.

2.3. Multi-index urban scene description

In this subsection, effectiveness of the proposed multi-index features is demonstrated by comparison with two well-known unsupervised feature extraction methods, PCA (principal component analysis), and ICA (independent component analysis). Similarly to PCA and ICA, PI can be viewed as an unsupervised image feature transformation approach, resulting in a subspace of the original image. However, PI is a semantic-oriented transformation, related to a set of specific objects of interest, while PCA and ICA correspond to the constraints of variance and independence, respectively, which is a data-oriented transformation (Zhang et al., 2006).

In Fig. 1, a high-resolution urban scene as well as its PCA, ICA, PI, and VI features are shown for a visual comparison. It is clearly seen that PI presents better discrimination ability for urban structures than PCA and ICA. In the pseudo-color image of PI, the classes of buildings, roads, soil, vegetation, and shadow are well represented

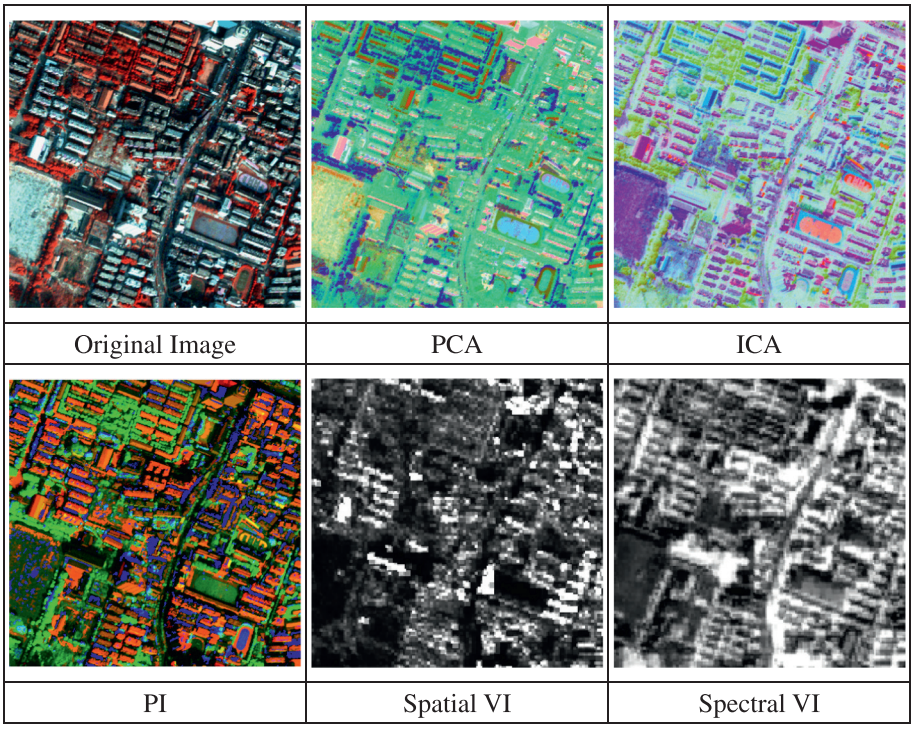


Fig. 1. Comparison and demonstration of unsupervised image feature transformation for PCA, ICA, PI (Red = MBI, Green = NDVI, Blue = MSI), and VI, respectively.

as red, black, gray, green, and blue, respectively. Moreover, the Jeffries–Matusita (JM) distance (Richards and Jia, 1999) is used to measure the separability of several spectrally similar class pairs with the transformed features (Table 1). The JM distance indicates how well the selected class pairs are statistically separate. A high value indicates that the feature space can be well separated, while a low one indicates that the feature space is not well separated. In Table 1, the JM values also support the observation that the three-channel primitive indices (PI) are more potential for discrimination between urban information classes than PCA and ICA, since the JM values of PI are significantly higher than others. Moreover, it can be also seen that the discrimination ability is further enhanced when PI and VI are combined, showing that the variation indices are complementary to the primitive indices for the purpose of urban classification.

3. Multi-index learning

In this section, a novel multi-index learning (MIL) framework is proposed for interpretation of the multiple information indices as well as the multispectral signals in an urban scene. In recent years, the support vector machine (SVM) has received much interest for remote sensing image classification and its efficiency has been proved in many applications, such as spectral-spatial classification (Huang et al., 2007), multiscale interpretation (Bruzzone and Carlin, 2006), and multi-sensor fusion (Waske and Benediktsson, 2007). SVM is chosen as the base classifier of the multi-index learning due to the following factors:

Table 1
JM distance for spectrally similar class pairs with different image feature transformation methods.

Subspace	Buildings–roads	Buildings–soil	Roads–soil	Trees–grass
PCA	1.10	1.34	0.98	1.61
ICA	1.13	1.29	0.91	1.59
PI	1.33	1.37	1.31	1.64
PI + VI	**1.53**	**1.89**	**1.71**	**1.90**

The largest values are highlighted as bold.

1) Rapid and adaptive learning.
2) Enhancing the separability of low-dimensional feature space via the nonlinear mapping and kernel trick.
3) Not constrained to prior assumptions on the distribution of input features (e.g., normal distribution).

It should be noted that, in addition to the simple single-kernel SVM (S-SVM), we propose to exploit multi-kernel SVM (M-SVM) for multi-index learning. M-SVM overcomes the drawbacks of the S-SVM that works as a black box and gives no insight about the importance of the distinct features (Tuia et al., 2010). As a result, the M-SVM is able to weight the multiple indices adaptively according to their relative importance for a specific task. The basic principles of the single and multi-kernel SVMs are described below.

Given a set of n labeled training samples $\{(x_i, y_i)\}_{i=1}^{n}$ with $y_i \in \{-1, +1\}$ and a nonlinear mapping $\phi(\cdot)$, the SVM classifier needs to solve

$$\min_{w,\xi_i,b}\left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i\right\} \tag{7}$$

where $\mathbf{w}$ represents the vector of parameters for the optimal decision hyperplane $f(x) = \langle \mathbf{w}, \phi(\mathbf{x})\rangle + b$ with b the bias. C is a regularization parameter and ξ_i is a slack variable. The problem can be converted into maximization of the margin in the high-dimensional space:

$$\max_{\alpha}\left\{\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j \langle \varphi(\mathbf{x_i}), \varphi(\mathbf{y_i})\rangle\right\} \tag{8}$$

$$\text{s.t.}\ \begin{cases}\sum_{i=1}^{n}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C, \forall i = 1, 2, 3, \ldots n\end{cases}$$

where α_i is the Lagrange multipliers that determine the support vectors.

The standard SVM uses one kernel for projecting the original data to the high-dimensional feature space. Moreover, the SVM classification highly depends on the kernel function utilized. The use of a single kernel does not consider the heterogeneous distribution of multi-source information, and, hence decrease the generalization ability of the SVM. In this study, therefore, we also propose to use the multiple-kernel framework for multi-index learning. The M-SVM can be constructed with a linear combination of several basis kernels,

$$K(\mathbf{x_i}, \mathbf{x_j}) = \sum_{m=1}^{M} d_m K_m(\mathbf{x_i}, \mathbf{x_j}) \tag{9}$$

$$\text{s.t.} \begin{cases} d_m \geqslant 0 \\ \sum_{m=1}^{M} d_m = 1 \end{cases}$$

where M is the number of candidate kernels; K_m and d_m represent the m-th kernel and its weight, respectively.

The basic idea of the M-SVM is to determine the preserved kernels and the corresponding weights. By optimizing the weight of each basis kernel, the useful information hidden in the multi-index features can be effectively mined and better exploited. However, solving the M-SVM might be complicated and computationally unaffordable when a high number of training samples or kernels is used (Tuia et al., 2010). In order to overcome this problem, an efficient algorithm, SimpleMKL (Rakotomamonjy et al., 2008), which makes multi-kernel learning tractable for large-scale problems, is used in this study. SimpleMKL aims at wrapping a multi-kernel SVM solver with a single kernel (e.g., the linear combination in (9)). Specifically, by introducing (9) into (8), the multiple-kernel dual problem can be expressed as

$$\max_{\alpha} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \sum_{m=1}^{M} d_m K_m(\mathbf{x_i}, \mathbf{x_j}) \right\}. \tag{10}$$

It is shown that maximizing the dual problem in (10) is equivalent to solving the following problem (Rakotomamonjy et al., 2008)

$$\min_{d} J(d) \text{ such that } \sum_{m=1}^{M} d_m = 1, d_m \geqslant 0, \tag{11}$$

$$\text{with } J(d) = \min_{w,b,\xi} \frac{1}{2} \sum_{m=1}^{M} \frac{1}{d_m} \|w_m\|^2 + C \sum_{i=1}^{n} \xi_i,$$

$$\text{subject to } \begin{cases} y_i \left(\sum_{m=1}^{M} \langle w_m, \phi_m(x_i) \rangle + b \right) \geqslant 1 - \xi_i \\ \xi_i \geqslant 0 \end{cases}$$

Actually, the objective function $J(d)$ is an optimal SVM objective value, which is solved by a simple gradient method. Readers can refer to Rakotomamonjy et al. (2008) for more details about multi-kernel learning and the SimpleMKL strategy.

4. Experiments and discussion

The processing chain of the proposed multi-index learning method is shown in Fig. 2. This section refers to a series of elaborate experiments, consisting of the following parts:

1) Introduction to the datasets and the experimental setup.
2) Evaluation of the proposed MIL method on three high-resolution datasets: GeoEye-1 (GE-1) and QuickBird (QB) images over urban areas, and a WorldView-2 (WV-2) image over a rural area.
3) A comparative study on the three high-resolution datasets between the proposed MIL method and a set of state-of-the-art high-resolution image classification methods.
4) A large-scale high-resolution urban mapping (260 km^2 covering the Wuhan City center) based on the ZY-3 satellite (the first civilian high-resolution satellite of China).

4.1. Datasets and experimental setup

The three test datasets as well as their PI feature images are shown in Fig. 3. It can be observed that the PI feature space exhibits more discrimination information between urban structures compared to the traditional spectral space. The characteristics of the three datasets are listed below:

- GE-1 Wuhan: It shows a typical urban landscape of Wuhan City, with dense residential areas, sparse vegetation, and bare land for construction.
- QB Wuhan: It covers a campus scene, including regular buildings with heterogeneous roofs, forests, meadows, water body, etc.
- WV-2 Hainan: It is a natural landscape in the Hainan Island, with a lot of grasslands, soil, small buildings, water areas, and a golf course.

The aforementioned images with different characteristics are used to test the robustness of the proposed classification algorithms. In order to assess the classification accuracy, a reference map for each test image is manually delineated based on visual interpretation (WV-2 Hainan) and a field campaign (GE-1 and QB Wuhan). The number of the reference samples for each information class in the three test datasets is provided in Table 2. The training and test samples are randomly selected from the reference samples for training and validation of the classification results, respectively. The number of the training samples for each class is 100, 100, and 50, respectively, for GE-1, QB, and WV-2 images.

The parameters used in the three test images are provided below.

- MBI and MSI: $D = \{0°, 45°, 90°, 135°\}$ for the directions, $S = \{3, 11, 19, 27\}$ for the scale parameters of the linear SE.

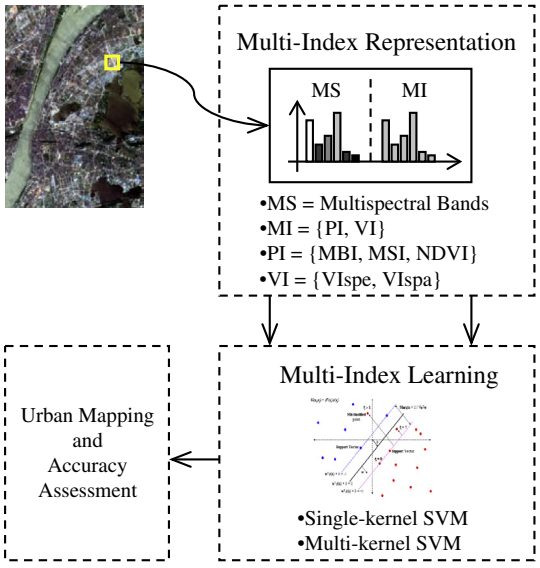
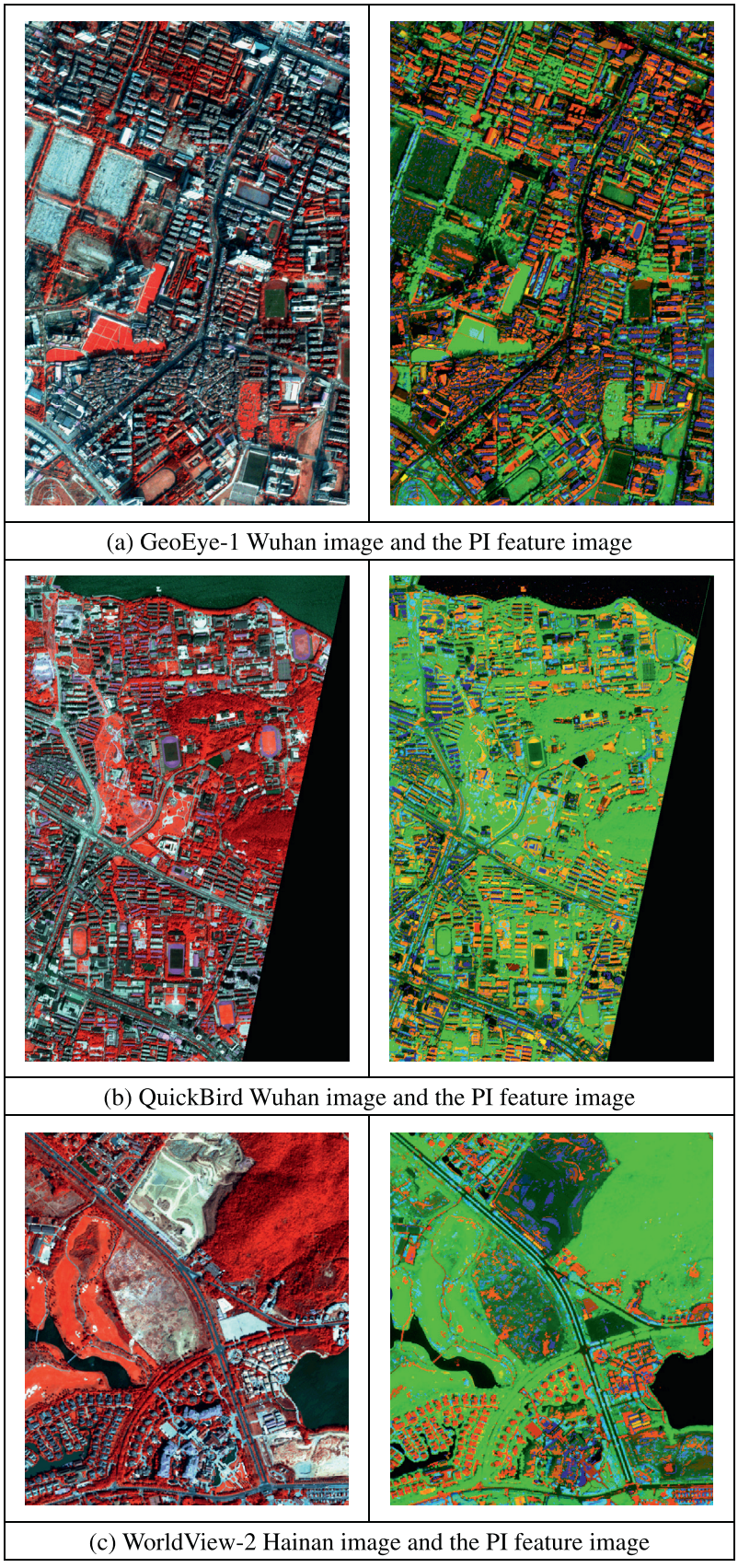


Fig. 2. The processing flow of the proposed multi-index learning (MIL) framework for urban mapping.



(a) GeoEye-1 Wuhan image and the PI feature image

(b) QuickBird Wuhan image and the PI feature image

(c) WorldView-2 Hainan image and the PI feature image

Fig. 3. High-resolution test images and their PI (primitive index) feature images for (a) GeoEye-1 Wuhan, (b) QuickBird Wuhan, and (c) WorldView-2 Hainan, respectively. Note that the color-infrared multispectral images are shown, i.e., the near infrared channel is used as the Red component in the RGB space.

Table 2
Number of reference and test samples (in pixels) for the three high-resolution datasets.

Classes	GeoEye-1 Wuhan		QuickBird Wuhan		WorldView-2 Hainan	
	Reference	Test	Reference	Test	Reference	Test
Roads	3187	2000	5103	5000	5357	5000
Grass	4098	2000	9179	5000	7417	5000
Buildings	20,074	2000	18,296	5000	11,578	5000
Soil	18,249	2000	3709	3000	22,189	5000
Shadow	1330	1000	4378	4000	1427	1400
Trees	5377	2000	17,415	5000	14,086	5000
Water body	NA	NA	16,614	5000	11,209	5000

- VI: $w = \{4 \times 4 \times 4,\ 8 \times 8 \times 4\}$ for GE-1 and QB, and $w = \{4 \times 4 \times 8, 8 \times 8 \times 8\}$ for WV-2.
- S-SVM: Regularization parameter $C = 100$, kernel = RBF (radial basis function), RBF bandwidth $\sigma = 1/n$ (n is the dimension of the input features).
- M-SVM: Four RBF kernels for the multispectral (MS) features and four RBF kernels for the multi-index (MI) features are used in the multi-kernel learning. The bandwidths of the four kernels are set to $\sigma = \{0.1, 0.25, 0.35, 0.5\}$ for both MS and MI features. Note that all the MS and MI features have been linearly normalized into [0,1] before they are processed.

The parameters for the multi-index calculation are determined according to the spatial resolution and sizes of the objects of interest, while the parameters for the SVM-based classifiers are tuned manually. It should be underlined that further optimization of the parameter selection might improve the classification accuracy, however, the improvement is not significant and this discussion is beyond scope of this paper.

4.2. Results of the GE-1, QB, and WV-2 experiments

The classification accuracies (overall accuracy and Kappa coefficient) of the proposed multi-index learning method for the three test datasets are shown in Table 3. In order to analyze the effect of each feature source for urban classification, different combinations of information indices as well as the multispectral features are tested. The accuracies of the MS features are used as benchmarks representing the traditional spectral-based classification method, and all the results better than the benchmarks are highlighted as gray in the table. The comments on the experiments are summarized as follows.

1) Individual use of VI or PI does not produce satisfactory results since their OA and kappa values are much lower than the MS classification. However, it is seen that PI significantly outperforms VI in terms of accuracies, which shows that primitive indices are more informative than the local variation features for urban classification.
2) Combination of PI and VI can strikingly enhance the performance of their individual use. Comparing VI and 'VI + PI', the improvements of OA achieved by the latter are 20%, 30%, and 14% for GE-1, QB, and WV-2, respectively. As for PI and 'VI + PI', the improvements of OA are also significant: 12%, 9%, and 8%, respectively, for the three datasets. This observation reveals that the proposed PI and VI can complement each other and provide more effective discrimination between urban land cover classes. In addition, it is found that 'PI + VI' gives better results than the MS classification in the urban areas (GE-1 and QB datasets).
3) Combination of MS (multispectral) and MI (multi-index): Stacking of MS and MI gives the highest accuracies in all the three experiments and their results are very promising considering that the overall accuracies are higher than 90% with a large number of test samples. In addition, it is found that the M-SVM slightly outperforms S-SVM, and the improvements are marginal (smaller than 1%).

The classification maps of the traditional multispectral (MS) classification and the proposed multi-index learning (MIL) method are compared in Fig. 4 for a visual inspection. Moreover, several subsets are extracted from the three test images in order to show the advantages of the MIL method compared to the MS classification (Fig. 5).

Subset (a): It can be seen that a large number of misclassifications occur between soil, buildings, and roads for the MS classification. However, these spectrally similar classes are correctly identified by the MIL method.

Subset (b): Similarly to the case (a), the buildings that are wrongly classified as roads by the MS are correctly identified by the MIL. In addition, it can be seen that the road in the center of the scene is correctly detected by the MIL but is wrongly labeled as buildings by the MS.

Subset (c): The confusion between buildings and bare soil in the MS classification can be overcome by the MIL.

4.3. Comparison

A comparative study is conducted in order to further verify the effectiveness of the proposed multi-index learning method for urban classification. To this aim, a series of state-of-the-art classification methods for high-resolution images are carried out in this study for the purpose of comparison, including: object-based image analysis (OBIA) (Bruzzone and Carlin, 2006), spectral-spatial approach based on the gray level co-occurrence matrix (GLCM) (Huang and Zhang, 2013), and spectral-spatial classification using the well-known differential morphological profiles (DMP) (Pesaresi and Benediktsson, 2001). These algorithms are briefly described as follows:

Algorithm (1): OBIA.

Step1. Adaptive segmentation. The first step of the object-based classification is to segment the image into a set of objects. In this study, the adaptive mean-shift procedure (Huang and Zhang, 2008) is employed considering that it is a robust feature space analysis approach and the scale parameter can be adaptively determined according to the spatial homogeneity around the object considered.

Step2. Spectral-spatial feature extraction. After segmentation, a set of spectral and spatial features are extracted for each object. The spectral properties refer to the mean and standard deviation of the pixels within each object. The spatial features include length, width, compactness, solidity, length–width ratio, and extent:

- Length/Width is defined as the length of the major/minor axis of the ellipse that has the same second moments as the object considered.

Table 3
Classification accuracies (OA = overall accuracy, K = Kappa coefficient) of the proposed multi-index learning method (MS = MultiSpectral, VI = Variation Index, PI = Primitive Index) for GE-1, QB, and WV-2 experiments.

Datasets	MS	VI	PI	VI + PI	MS + PI	MS + PI + VI (S-SVM)	MS + PI + VI (M-SVM)
GE-1							
OA	80.9%	63.4%	71.9%	83.9%	88.6%	90.4%	90.0%
K	0.769	0.560	0.663	0.805	0.862	0.883	0.879
QB							
OA	86.4%	57.6%	78.5%	87.4%	89.8%	92.1%	92.2%
K	0.841	0.504	0.748	0.852	0.881	0.907	0.908
WV-2							
OA	86.4%	59.1%	65.2%	73.7%	92.2%	91.9%	92.9%
K	0.839	0.528	0.589	0.694	0.908	0.905	0.917

- Compactness is computed as $\frac{2}{\text{Diameter}} \cdot \sqrt{\left(\frac{\text{Area}}{\pi}\right)}$, where Diameter indicates the diameter of a circle with the same area as the object.
- Solidity is defined as the proportion of the pixels in the convex hull that are also in the object.
- Extent is defined as the proportion of the pixels in the bounding box that are also in the object. It is computed as the area of the object divided by the area of the bounding box.

A notable advantage of the OBIA is that it is convenient to extract geometrical properties of objects, which are potential for discrimination between spectrally similar urban classes.
Step3. The spectral and spatial properties for each object can be expressed as a vector:

$$\left\{(\text{mean}(n),\ \text{std}(n))_{n=1}^{N},\ (\text{L, W, Ratio, Com, Solidity, Ext})\right\}$$

where N is the number of spectral bands. (mean, std) represent the spectral features of an object (mean and standard deviation), and (L, W, Ratio, Com, Solidity, Ext) are the geometrical attributes of an object (Length, Width, Length–Width Ration, Compactness, Solidity, and Extent, respectively). The spectral-spatial feature vector is then input into the S-SVM for object-based classification.

Algorithm (2): GLCM-based spectral-spatial classification.

Step1. The first two spectral principal components (PC1 and PC2) extracted by the principal component analysis (PCA), containing more than 90% information of the original spectral signals, are used as the basis images for the subsequent texture calculation.
Step2. The GLCM texture feature can be expressed as a function of window size (w), statistical measure (m), direction (d), and basis images (*base*): GLCM(w, m, d, *base*). In this study, the parameters are set as: $w = \{5 \times 5, 9 \times 9\}$, m = {Contrast, Homogeneity}, $d = \{0°, 45°, 90°, 135°\}$, *base* = {PC1, PC2}. The window sizes are chosen according to the spatial resolution and the sizes of the objects in the study area. The Contrast and Homogeneity are used as textural measures since they are commonly used in urban areas, and they represent the spatial property of heterogeneity and homogeneity in a local region, respectively.
Step3. The spectral and textural features are written as a vector:

$$\left\{(\text{SPE}(n))_{n=1}^{N},\ \text{GLCM}\right\},$$

where SPE(n) represents the spectral feature value for band n. The hybrid feature vector is then input into the SVM for spectral-spatial classification.

Algorithm (3): DMP-based spectral-spatial classification.

Step1. The PCA transformation is used to generate the basis images for the subsequent DMP computation.
Step2. DMP is a well-known structural feature extraction approach for high-resolution images and it is effective in describing multiscale shape profiles of objects. It can be expressed as DMP(SE, m, *base*), where m stands for the morphological operators, and SE indicates the shape and sizes of the structural element (SE). In this study, the parameters of the DMP are set as: *base* = {PC1, PC2}, SE = {shape = 'disk', diameter = (3, 5, 7, 9)}, and m = {opening/closing by reconstruction}.
Step3. The spectral and morphological features form a vector $\{(\text{SPE}(n))_{n=1}^{N}, \text{DMP}\}$, which is then input into the SVM for spectral-spatial classification.

The classification accuracies of the state-of-the-art high-resolution image classification algorithms as well as the proposed MIL method are compared in Table 4. Generally speaking, the proposed MIL method achieves the best results, except for the case of the QB experiment, where OBIA slightly outperforms MIL. In particular, it should be underlined that the dimension of the MIL input features is much lower than other methods (9-dimensional feature for the GE-1 and QB (4-d MS, 3-d PI, and 2-d VI), and 13-dimensional feature for the WV-2 (8-d MS, 3-d PI, and 2-d VI)). The results of the OBIA are also very promising in all the experiments, especially for the QB dataset (OA = 92.8%, the first ranking among all the methods). The performance of the DMP-based spectral-spatial classification is satisfactory since it gives overall accuracy larger than 91% in QB and WV-2 experiments. As for the GLCM, it obtains high accuracy in QB dataset (OA = 91.9%), but weak results in GE-1 and WV-2.

4.4. Large-scale experiment based on ZY-3 satellite

The proposed MIL method aims at achieving accurate image classification over urban areas with a low-dimensional feature space, and, hence, it is a practical technique for large-scale (regional scale) high-resolution urban mapping. To this aim, we design an experiment of large-scale high-resolution urban classification based on ZY-3, which is the China's first civilian high-resolution mapping satellite. Note that this is the first result of ZY-3 satellite for urban mapping to our knowledge.

The study area is located at Wuhan, capital of Hubei Province and one of the largest cities in China. The dataset was acquired on 22 April 2012. It contains four spectral channels (Blue, Green, Red, and Near Infrared) with a spatial resolution of 5.8 meter. As shown in Fig. 6, the test image covers the city center of Wuhan (260 km^2). It is a typical urban landscape of China, where the residential area is composed of high-density buildings but with

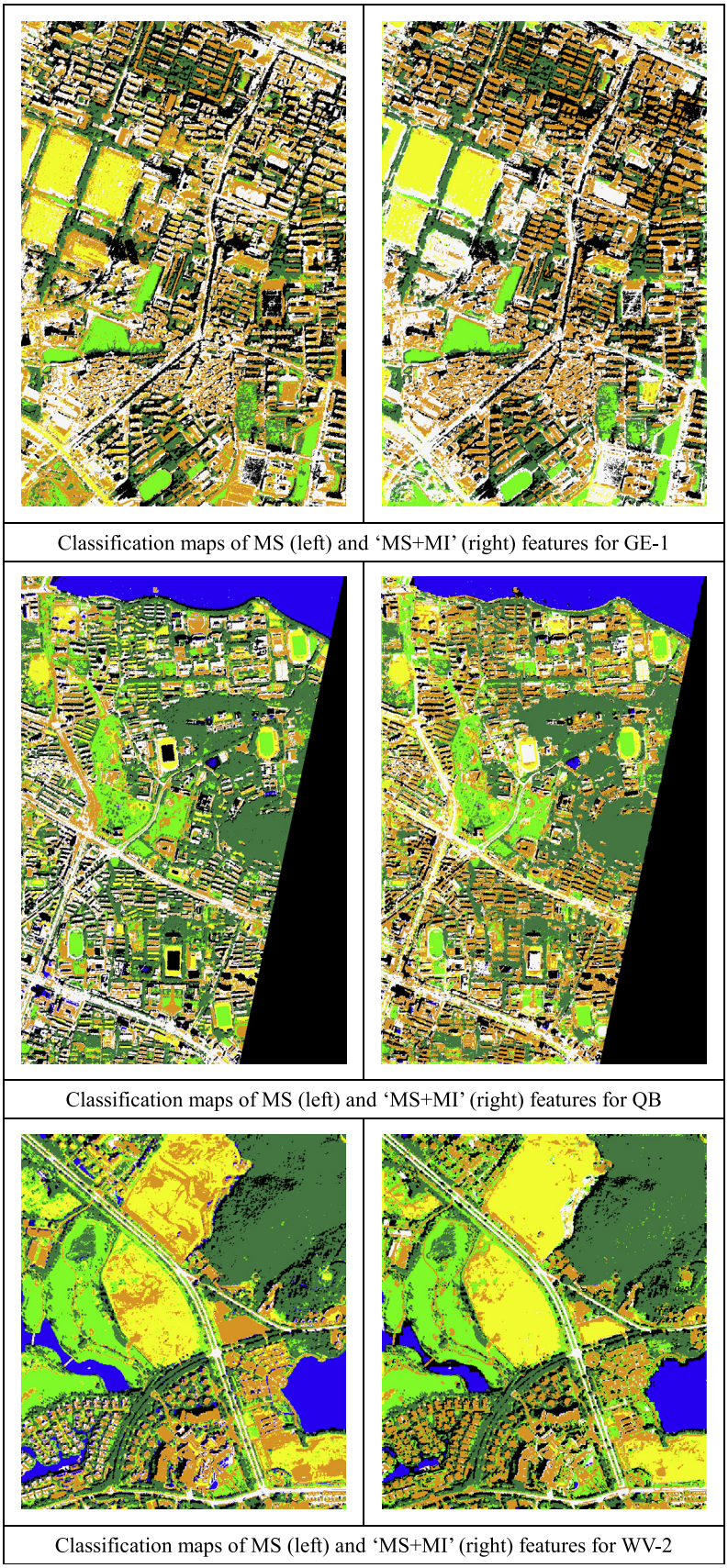


Fig. 4. Classification maps of MS (left column) and 'MS + MI' (right column) features for the three datasets (Orange = buildings, White = roads, Blue = water, Dark Green = trees, Light Green = grass, Black = shadow, Yellow = Soil).

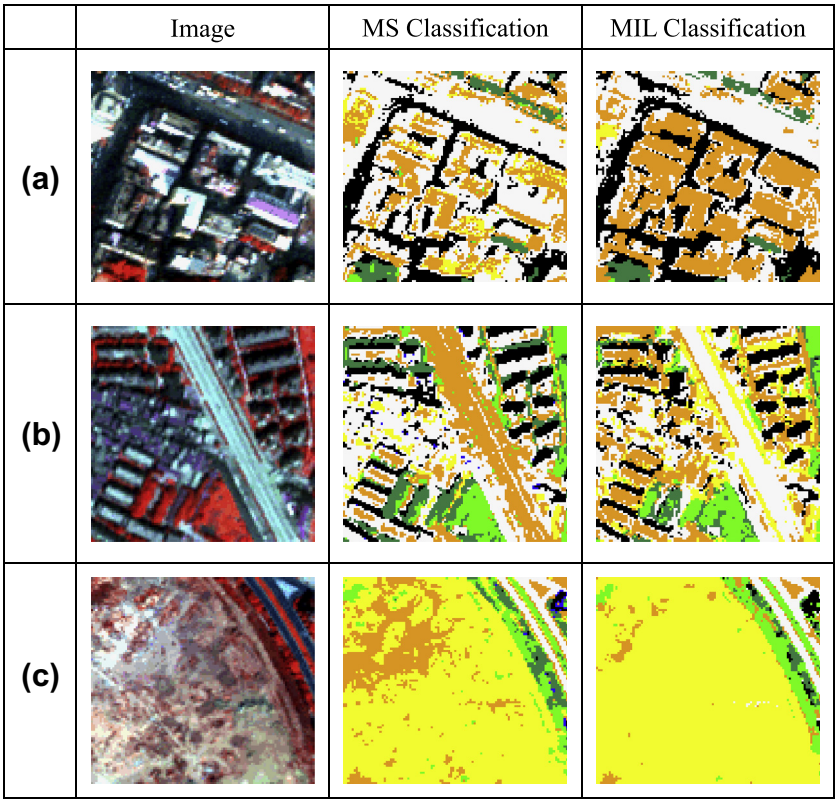


Fig. 5. Examples showing that the proposed multi-index learning method improves the traditional spectral classification, extracted from (a) GE-1, (b) QB, and (c) WV-2, respectively.

Table 4
Classification accuracies (OA = overall accuracy, *K* = Kappa coefficient) for comparison between several state-of-the-art high-resolution image classification algorithms (OBIA, GLCM and DMP-based spectral-spatial classification) and the proposed MIL method (Dim = the dimension of the input features).

Datasets	OBIA	GLCM	DMP	MIL
GE-1				
Dim	14	36	20	9
OA	88.4%	81.5%	86.9%	90.4%
K	0.860	0.775	0.842	0.883
QB				
Dim	14	36	20	9
OA	92.8%	91.9%	91.1%	92.1%
K	0.916	0.906	0.896	0.907
WV-2				
Dim	22	40	24	13
OA	91.2%	83.5%	91.3%	91.9%
K	0.896	0.805	0.898	0.905

inadequate green space (e.g., gardens, trees, meadows). Most of the buildings have seven to ten floors. It should be kept in mind that in this experiment a personal computer (e.g., equipped with Intel (R) Core2 CPU and 2.0G RAM in this study) cannot support the data processing for the high-dimensional feature extraction and classification (e.g., OBIA, GLCM, and DMP) due to the limitation of memory space and computational cost.

Table 5 shows the training and test samples used in this experiment. The four reference images presented in Fig. 6 are delineated manually according to a filed campaign and our prior knowledge on the study area. The test samples consists of four validation areas, (a) HanKou (commercial area), (b) QingShan (residential area), (c) WuChang (educational area), and (d) HongShan (industrial area), covering the main administrative districts of Wuhan City center.

The accuracies of urban mapping by the proposed MIL method for the four validation regions are provided in Table 6. The conclusions are very clear:

1) In all the cases, the proposed MIL method achieves the best results in terms of the accuracies. Compared to the MS classification, the improvements of OA are 9%, 4%, 7%, and 11%, respectively, for the four test regions.
2) For the case of 'MS + VI', it can be seen that addition of VI can enhance the traditional spectral classification in all the experiments. As for the 'MS + PI', the results are more promising since it gives much higher accuracies than the MS and 'MS + VI'.

The results of the Wuhan urban mapping for the multispectral classification and the proposed MIL method are compared in Fig. 7 for a visual inspection. It is shown that the multi-index learning approach can obtain satisfactory result for regional urban mapping since the overall accuracies are 82% for test region (a) and (d), and 90% for (b) and (c), with a large number of test samples (about 220,000 pixels) but a limited training samples (100 pixels for each class).

4.5. Discussions

A notable characteristic of the proposed method is to represent the complicated urban image scenes using a set of low-dimensional semantic indices. The effectiveness for each information index is analyzed as follows:

1) MBI: It aims at modeling the spectral-spatial characteristics of buildings based on a series of morphological operators (brightness, local contrast, size, and isotropy).

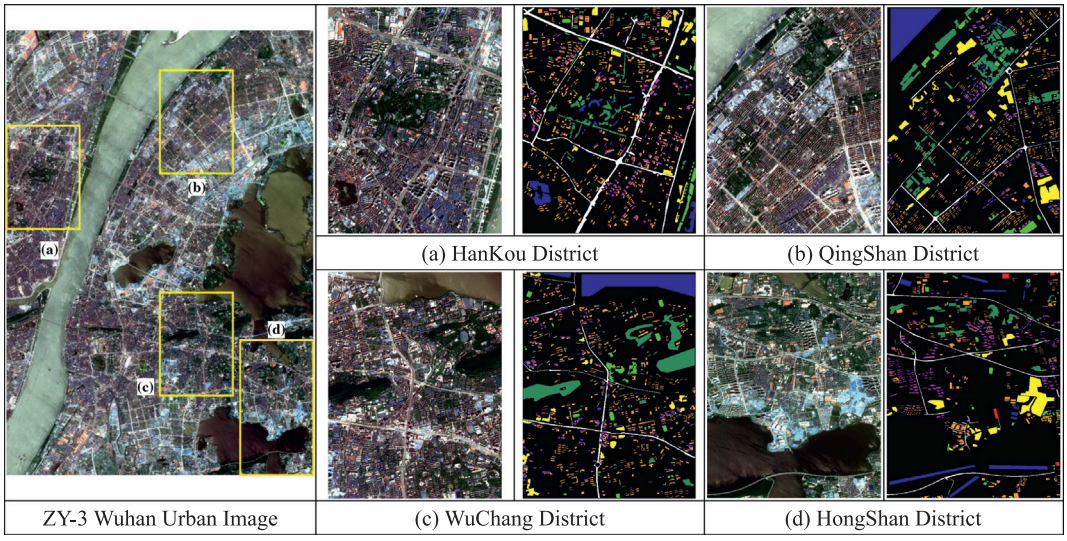


Fig. 6. The ZY-3 Wuhan city image with four validation areas, corresponding to the four main administrative districts of Wuhan (Orange = buildings, White = roads, Blue = water, Dark Green = trees, Light Green = grass, Magenta = shadow, Yellow = Soil). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Training and test samples used in the ZY-3 experiment.

Class	# Trainings samples	# Test samples
Roads	100	31,191
Buildings	100	48,106
Shadow	100	18,289
Bare Soil	100	28,685
Grass	100	12,382
Trees	100	43,032
Water	100	38,114
Total	700	219,799

Table 6
Classification accuracies (OA = overall accuracy, K = Kappa coefficient) of the proposed multi-index learning method (MS = MultiSpectral, VI = Variation Index, PI = Primitive Index) for the four validation regions of ZY-3 experiment.

Datasets	MS	MS + VI	MS + PI	MS + PI + VI (S-SVM)
(a) HanKou				
OA	73.4%	78.4%	81.7%	**82.7%**
K	0.668	0.730	0.770	**0.782**
(b) QingShan				
OA	85.7%	87.2%	89.1%	**89.4%**
K	0.828	0.846	0.869	**0.872**
(c) WuChang				
OA	83.4%	86.8%	90.1%	**90.8%**
K	0.794	0.837	0.875	**0.885**
(d) HongShan				
OA	71.5%	72.8%	82.2%	**82.3%**
K	0.660	0.676	0.787	**0.788**

The largest values are highlighted as bold.

- Brightness and local contrast: The relatively high reflectance of roofs and the spatially adjacent shadows lead to high local contrast of buildings compared to their surroundings. Accordingly, the top-hat transformation is used to measure the local contrast.
- Size: Sizes for most of buildings are within a scale range, and, hence, the lengths of the linear structural elements are used to define the spatial scale of buildings.
- Isotropy: Buildings and roads are spectrally similar urban structures, thus, it is difficult to discriminate between them without considering their difference in the spatial properties. Buildings show high local contrast in more directions than roads, as buildings are compact and isotropic but roads are elongated in one or two directions. In this regard, MBI defines the isotropy by measuring the multidirectional contrast based on the linear structural elements.

2) MSI: Construction of MSI is based on the observation that buildings and shadow exhibit spatially similar but spectrally contrary properties. Shadow, having low intensity value, also shows high local contrast in different directions. Consequently, MSI can be naturally defined within the framework of MBI, by replacing the white top-hat (highlight bright structures) by the black top-hat (highlight dark structures).
3) NDVI: Its effectiveness is guaranteed by the biophysical model of vegetation.
4) VI: The primitive indices (MBI, MSI, NDVI) represent stationary image semantics, while VI aims at describing the image spectral-spatial variations. The principle of VI is that urban classes have different degrees of variations in spectral and spatial domains, which provides discriminative information for them.

The effectiveness of the information indices for modeling urban landscapes is demonstrated in Fig. 8. The QuickBird and GeoEye-1 images are taken as examples since they exhibit typical and complicated urban scenes. In the figure, the vertical axis represents the average feature values which are calculated based on all the reference samples per each class and then scaled into the range between 0 and 255. By focusing on the primitive indices, it can be found that MBI, MSI, and NDVI are effective in highlighting buildings, shadow, and vegetation, respectively, and, at the same time suppressing other classes. In the case of spatial variation index (VI_{spa}), it can be seen that it is able to describe the complexity of different classes. For instance, buildings and shadow have high VI_{spa} values in both images since they present more structural variation. On the other hand, water, bare soil, and grass, however, are relatively homogeneous in the spatial domain, and, hence, they have small VI_{spa} values. With respect to the spectral variation index (VI_{spe}), it aims at depicting the image spectral variation across multispectral bands. From Fig. 8, it can be observed that there is not a regular pattern for the VI_{spe}, since it is related to specific image

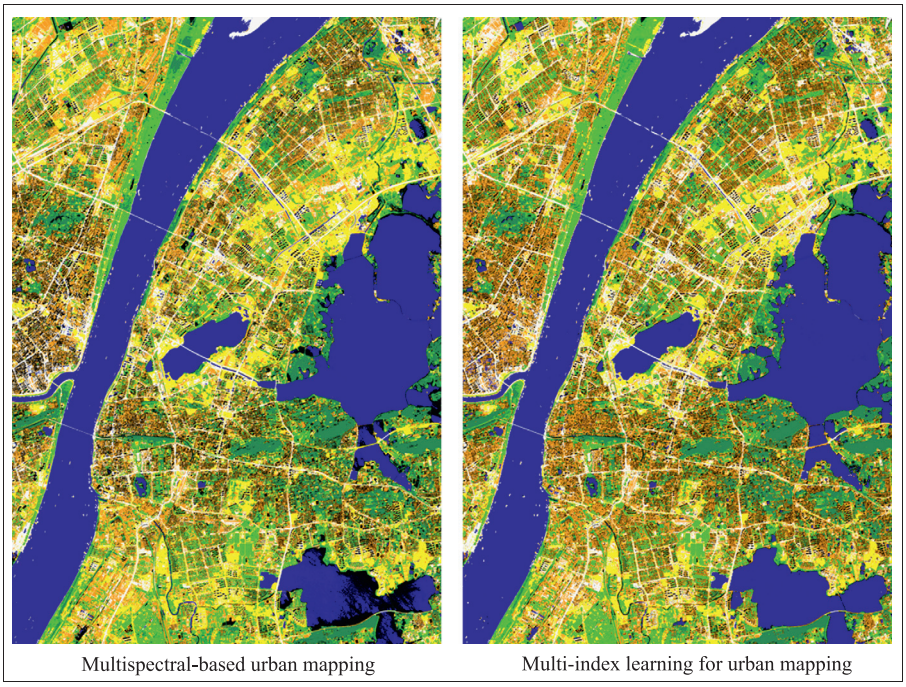


Fig. 7. Classification maps of the multispectral-based and MIL-based approaches for urban mapping over Wuhan City Center (260 km^2) (Class legend: Orange = buildings, White = roads, Blue = water, Dark Green = trees, Light Green = grass, Black = shadow, Yellow = Soil).

scenes, e.g., illumination, contrast, landscapes. However, it can provide discrimination information for different classes from the spectral interpretation viewpoint.

The proposed multi-index learning (MIL) method aims to interpret the urban image landscapes using a set of low-dimensional information indices, which provides a practical strategy for urban mapping. The high efficiency and satisfactory performance of the MIL method can be attributed to the factor that the low-dimensional feature space is composed of semantic information, which has high discrimination ability for urban land cover classes.

The computational burden of the MIL method consists of two parts: the multi-index calculation and learning. The primitive indices (MBI, MSI, and NDVI) can be calculated rapidly and automatically, while the variation indices based on the 3D wavelet transformation need more computation time since they are extracted for each local image cube. However, our experiments show that the primitive indices combined with the spectral bands can achieve satisfactory classification accuracies (see Tables 3 and 6). We can, therefore, only choose the PI as the input features for urban mapping when the computational efficiency is of primary importance. On the other hand, concerning the multi-index learning, the experimental results reveal that the single-kernel SVM (S-SVM) can obtain similar classification accuracy with the multi-kernel SVM (M-SVM) (see Table 3). Thus, the former can be used as a reasonable alternative for rapidly interpreting the multiple information indices. In addition, it should be noted that the parallel computing can be considered in future for further improving the efficiency of the proposed workflow.

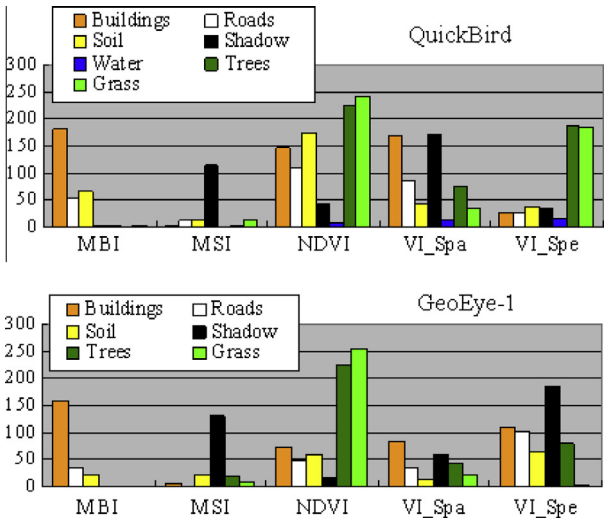


Fig. 8. Demonstration of the information indices for discrimination between different information classes in QuickBird and GeoEye-1 images (VI_Spe and VI_Spa denote spectral and spatial variation index, respectively).

The transferability of the proposed multi-index learning, which is not addressed in this study, will be included in our future research. The active learning (Tuia et al., 2011) and transfer learning (Pan and Yang, 2010) can be considered for transferability of the multi-index learning across image scenes and across different sensors, in order to deal with the so-called covariate shift problem, i.e., the discrepancy between training and test data.

5. Conclusions

The contributions of this paper lie in the following three aspects:

1) An innovative multi-index learning (MIL) method is proposed for high-resolution image classification over urban areas. The notable advantage of the MIL is that it is capable of achieving accurate classification results with a low-dimensional semantic feature space, including a set of stationary primitive indices (MBI, MSI, NDVI) and a couple of dynamic indices (spectral and spatial variation).

Remarks: The effectiveness of the MIL method was verified in the experiment Section 4.2, in which the feature combination

'MS + PI + VI' provided the highest accuracies in all the experiments. In addition, it was shown that the PI and VI were complementary since their combination could significantly improve the performance of their individual use.

2) A comparative study was conducted for comparison between the proposed MIL method and several state-of-the-art high-resolution image classification algorithms that had been proved to be effective in many applications, such as the object-based image analysis (OBIA), multiscale textural classification (GLCM), and the differential morphological profiles (DMP).

Remarks: The superiority of the proposed MIL method compared to the well-known OBIA, GLCM, and DMP algorithms was verified in the experiment Section 4.3. It was shown that the MIL gave very promising results with a much lower feature space compared to the existing methods (Table 4).

3) A large-scale high-resolution urban mapping (Wuhan City center, 260 km^2) based on ZY-3, which is the China's first civilian high-resolution satellite, was conducted in order to test the feasibility of the MIL method for regional mapping. Note that it is the first announcement of the results for urban mapping from ZY-3 images.

Remarks: The effectiveness of the MIL method, as a practical strategy for large-scale high-resolution urban mapping, was validated according to the satisfactory accuracies (Table 6).The future research plans concerning the MIL framework refer to the following points:

1) Improvement of the primitive indices, e.g., optimization of the MBI and MSI, introduction of new information index.
2) An object-based MIL allowing for a rule-based post-processing.
3) Attempt of more learning algorithms for the multiple information indices, such as active learning (Tuia et al., 2011), deep learning (Hinton et al., 2006).

Acknowledgement

This work was supported in part by the Natural Science Foundation of China (41101336 and 91338111), in part by the Program for New Century Excellent Talents in University of China (NCET-11-0396), and in part by the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (FANEDD) under Grant 201348.

References

Aguera, F., Aguilar, J.F., Aguilar, A.M., 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. ISPRS J. Photogramm. Remote Sens. 63 (6), 635–646.
Awrangjeb, M., Ravanbakhsh, M., Fraser, S.C., 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. ISPRS J. Photogramm. Remote Sens. 65 (5), 457–467.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65 (1), 2–16.
Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 44 (9), 2587–2600.
Dell'Acqua, F., Gamba, P., Lisini, G., 2009. Rapid mapping of high resolution SAR scenes. ISPRS J. Photogramm. Remote Sens. 64 (5), 482–489.
Fauvel, M., Chanussot, J., Benediktsson, J.A., 2012. A spatial-spectral kernel-based approach for the classification of remote-sensing images. Pattern Recogn. 45 (1), 381–392.
Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. ISPRS J. Photogramm. Remote Sens. 66 (1), 56–66.
Hinton, G.E., Osindero, S., Teh, Y., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18 (7), 1527–1554.
Huang, X., Zhang, L., 2008. An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 46 (12), 4173–4185.
Huang, X., Zhang, L., 2011. A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery. Photogramm. Eng. Remote Sens. 77 (7), 721–732.
Huang, X., Zhang, L., 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. 5 (1), 161–172.
Huang, X., Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 51 (1), 257–272.
Huang, X., Zhang, L., Li, P., 2007. Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. IEEE Geosci. Remote Sens. Lett. 4 (2), 260–264.
Hussain, M., Chen, D., Cheng, A., Wei, H., Stanley, D., 2013. Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J. Photogramm. Remote Sens. 80, 91–106.
Inglada, J., 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS J. Photogramm. Remote Sens. 62 (3), 236–248.
Jing, L., Hu, B., Noland, T., Li, J., 2012. An individual tree crown delineation method based on multi-scale segmentation of imagery. ISPRS J. Photogramm. Remote Sens. 70, 88–98.
Johnson, B., Xie, Z., 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J. Photogramm. Remote Sens. 66 (4), 473–483.
Liu, Y., Bian, L., Meng, Y., Wang, H., Zhang, S., Yang, Y., Shao, X., Wang, B., 2012. Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis. ISPRS J. Photogramm. Remote Sens. 68, 144–156.
Mathieu, R., Freeman, C., Aryal, J., 2007. Mapping private gardens in urban areas using object-oriented techniques and very high-resolution satellite imagery. Landscape Urban Plann. 81 (3), 179–192.
Ouma, O.Y., Tateishi, R., 2008. Urban-trees extraction from QuickBird imagery using multiscale spectex-filtering and non-parametric classification. ISPRS J. Photogramm. Remote Sens. 63 (3), 333–351.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Pesaresi, M., Benediktsson, J.A., 2001. A new approach for the morphological segmentation of high-resolution satellite imagery. IEEE Trans. Geosci. Remote Sens. 39 (2), 309–320.
Pesaresi, M., Gerhardinger, A., Kayitakire, F., 2008. A robust built-up area presence index by anisotropic rotation-invariant textural measure. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. 1 (3), 180–192.
Pingel, J.T., Clarke, C.K., MaBride, A.W., 2013. An improved simple morphological filter for the terrain classification of airborne LIDAR data. ISPRS J. Photogramm. Remote Sens. 77, 21–30.
Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y., 2008. SimpleMKL. J. Mach. Learn. Res. 9, 2491–2521.
Reis, S., Tasdemir, K., 2011. Identification of hazelnut fields using spectral and Gabor textural features. ISPRS J. Photogramm. Remote Sens. 66 (5), 652–661.
Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag, Berlin, Germany.
Sebari, I., He, D., 2013. Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. ISPRS J. Photogramm. Remote Sens. 79, 171–184.
Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M., 2010. Learning relevant image features with multiple-kernel classification. IEEE Trans. Geosci. Remote Sens. 48 (10), 3780–3791.
Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J., 2011. A survey of active learning algorithms for supervised remote sensing image classification. IEEE J. Sel. Topics Signal Process. 5 (3), 606–617.
Tzotsos, A., Karantzalos, K., Argialas, D., 2011. Object-based image analysis through nonlinear scale-space filtering. ISPRS J. Photogramm. Remote Sens. 66 (1), 2–16.
Waske, B., Benediktsson, J.A., 2007. Fusion of support vector machines for classification of multisensor data. IEEE Trans. Geosci. Remote Sens. 45 (12), 3858–3866.
Yang, G., Pu, R., Zhang, J., Zhao, C., Feng, H., Wang, J., 2013. Remote sensing of seasonal variability of fractional vegetation cover and its object-based spatial pattern analysis over mountain areas. ISPRS J. Photogramm. Remote Sens. 77, 79–93.
Yoo, H.Y., Lee, K., Kwon, B.D., 2009. Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery. Int. J. Remote Sens. 30 (23), 6219–6239.
Zhang, L., Huang, X., Huang, B., Li, P., 2006. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 44 (10), 2950–2961.
Zhang, X., Xiao, P., Song, X., She, J., 2013. Boundary-constrained multi-scale segmentation method for remote sensing images. ISPRS J. Photogramm. Remote Sens. 78, 15–25.