

## International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tres20>

### A multilevel decision fusion approach for urban mapping using very high-resolution multi/hyperspectral imagery

Xin Huang<sup>a</sup> & Liangpei Zhang<sup>a</sup>

<sup>a</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, 430079, PR China

Published online: 21 Nov 2011.

To cite this article: Xin Huang & Liangpei Zhang (2012) A multilevel decision fusion approach for urban mapping using very high-resolution multi/hyperspectral imagery, International Journal of Remote Sensing, 33:11, 3354-3372, DOI: [10.1080/01431161.2011.591444](https://doi.org/10.1080/01431161.2011.591444)

To link to this article: <http://dx.doi.org/10.1080/01431161.2011.591444>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## A multilevel decision fusion approach for urban mapping using very high-resolution multi/hyperspectral imagery

XIN HUANG\* and LIANGPEI ZHANG

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

(Received 4 June 2010; in final form 23 April 2011)

A novel multilevel decision fusion approach is proposed for urban mapping using very-high-resolution (VHR) multi/hyperspectral imagery. The proposed framework consists of three levels: (1) at level I, we first propose a self-dual filter for extracting structural features from the VHR imagery—subsequently, the spectral and structural features are integrated based on a weighted probability fusion; (2) level II extends level I by implementing the spectral–structural fusion in an object-based framework; and (3) at level III, the object-based probabilistic outputs at level II are used to identify unreliable objects, and shape attributes of these unreliable objects are then considered for refinement of classification. At this level, a decision-level object merging is used to improve the initial segmentation, since shape feature extraction is highly dependent on the quality of segmentation. Experiments were conducted on a Hyperspectral Digital Imagery Collection Experiment (HYDICE) DC Mall image and a QuickBird Beijing data set. The results revealed that the proposed approach provided progressively increasing accuracies when the multilevel features were gradually considered in the processing chain.

### 1. Introduction

With the increasing availability of very high-resolution (VHR) remotely sensed images, we can acquire a large amount of detailed ground information, for which there is the potential for many new remote-sensing applications in urban monitoring, human activity impact, precision agriculture and so on. Nevertheless, detailed structures are adequately presented in VHR images, resulting in the fact that the same classes have different spectral properties and different classes show similar spectral reflectance. Consequently, the increase of within-class variance and decrease of interclass variance lead to the inadequacy of traditional spectral-based and pixel-based classification approaches.

Consequently, in recent years, researchers have proposed different algorithms to improve the conventional classification approaches, which can be summarized into the following two categories.

1. *Exploitation of textural and spatial features.* The basic idea of this approach is to extract textural, structural or contextual features from VHR images for

---

\*Corresponding author. Email: [huang\\_w hu@163.com](mailto:huang_w hu@163.com)

complementing the spectral information and hence discriminating spectrally similar classes. The grey level co-occurrence matrix (GLCM) is a widely used textural measure. Puissant *et al.* (2005) examined the potential of GLCM textural features combined with spectral information to improve the classification accuracy of intra-urban land cover types. Due to the fact that the analysis window size of GLCM is difficult to determine for different image scenes, Huang *et al.* (2007b) proposed an adaptive algorithm for optimal window selection by considering the edge density within an analysis window. More recently, Pacifici *et al.* (2009) extracted the GLCM textures based on a series of directions and displacement values with multiple window sizes. The resultant high-dimensional features were then selected using a neural network pruning and saliency measurements.

Another effective textural measure refers to the wavelet transform. It is able to represent either micro-texture using wavelet decomposition of local windows or macro-texture based on the stationary wavelet transform performed on the whole image. Shah *et al.* (2010) presented a new feature set by combining the independent component analysis (ICA) and wavelet transformation for image information mining in geospatial data. Ouma *et al.* (2006) extracted multiscale features based on vertical, horizontal and diagonal wavelet sub-bands of multiple decomposition levels. The resultant textures were used to delineate urban trees of QuickBird images. Another study by Ouma *et al.* (2008) integrated the GLCM and wavelet textures for discrimination between different vegetation types. The wavelet textures were also utilized for classification of VHR urban images. Myint *et al.* (2004) compared different spatial measures for urban feature discrimination, and it was revealed that the wavelet textures gave the highest accuracies in comparison with GLCM, spatial autocorrelation and fractal features. Huang *et al.* (2008) proposed a wavelet-based multiscale classification method, where the wavelet decomposition was used to extract the multiscale spatial information in a series of concentric windows.

In comparison with the aforementioned textural measures, structural features have been found to be more effective for VHR imagery since they are able to describe shape and geometrical characters. Huang *et al.* (2007a) proposed a structural feature set based on the spectral coherence of groups of pixels along selected directions. Benediktsson *et al.* (2003) presented differential morphological profiles (DMPs) based on opening/closing-by-reconstruction with increasing sizes of structural elements (SEs). Subsequently, the DMPs were extended to the extended morphological profiles (EMPs) by constructing the morphological features based on the principal components of urban hyperspectral images (Benediktsson *et al.* 2005). More recently, Fauvel *et al.* (2008) improved the EMP approach by considering both spectral–structural features and building the profiles on kernel principal components (Fauvel *et al.* 2009). Other notable improvements to DMPs referred to the fuzzy model (Chanussot *et al.* 2006) and partial morphological reconstruction (Bellens *et al.* 2008).

2. *Object-based analysis.* The basic idea of the object-oriented classification is to group the spatially adjacent pixels into spectrally homogeneous objects and then conduct classification on objects as the minimum processing unit. The advantages of object-based processing include its ability to reduce local spectral variation, avoid salt–pepper effects of the pixel-based classification, mimic human perception in identifying objects and generate geometrical attributes.

Therefore, object-based image analysis has received much attention, especially for VHR image processing and applications.

Most of the existing object-oriented algorithms are based on the well-known eCognition commercial software (Trimble Company, Sunnyvale, CA, USA) and the embedded fractal net evolution approach (FNEA) (Benz *et al.* 2004). Yu *et al.* (2006) calculated the object-based spectral, textural and geometric features using eCognition software for detailed vegetation mapping. The resultant hyperdimensional features were statistically ranked according to importance and then classified using a classification and regression tree algorithm (CART). Chubey *et al.* (2006) utilized object-based segmentation to delineate homogeneous landscape components from IKONOS-2 data. Van Coillie *et al.* (2007) combined object-based features and genetic algorithm-based feature selection for forest mapping with IKONOS imagery. Zhou *et al.* (2009) evaluated the effectiveness of the object-based approach for classification of shaded areas in urban VHR images. In addition to the FNEA algorithm, researchers have presented other segmentation methods for object extraction from VHR images, such as the adaptive mean shift procedure (Huang and Zhang 2008), which is an unsupervised feature space clustering algorithm, watershed segmentation (Li and Xiao 2007) and the *J*-measure-based segmentation (Wang and Boesch 2007).

In this article, a multilevel classification system is proposed by integrating the pixel-based structural features and the object-based shape features based on a novel decision-level probability fusion approach. The proposed system is able to yield multi-level outputs as the processing unit is updated progressively from the pixel level to the object level, and additional features are gradually considered in the decision process.

## 2. Methodology

The proposed multilevel processing chain is shown in figure 1.

1. At level I, a structural spectrum is proposed to represent the multiscale and multidirectional features based on the morphological transformation. Subsequently, the multispectral and the structural features are separately fed into two support vector machines (SVMs), respectively termed the spectral SVM and structural SVM. The spectral–structural decision fusion is carried out by taking the weighted probabilistic outputs of the two SVMs into consideration.
2. At level II, a mean shift segmentation is used to extract object boundaries. The object-based decision fusion is implemented by considering the probabilistic outputs of level I within the boundary of each object.
3. Level III focuses on the unreliable objects that have low probabilistic outputs at level II. A decision-level object merging is used to refine the initial segmentation of level II. Subsequently, shape attributes extracted from the refined boundaries are used to reclassify the unreliable objects. At this level, only the unreliable objects are of interest because shape features are not necessarily effective in interpreting all the objects.

### 2.1 Morphological structural spectrum

Mathematical morphology is an effective tool for extracting image components that are useful in the representation of structural features. Morphological features are

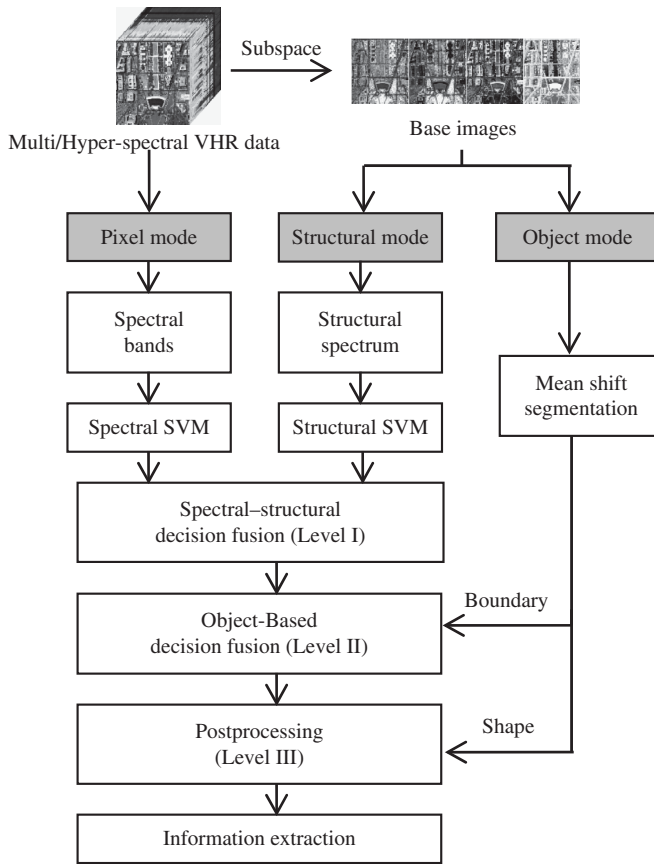


Figure 1. Flowchart of the multilevel decision fusion approach. VHR, very high resolution; SVM, support vector machine.

based on some basic operators, such as opening ( $\gamma$ ) and closing ( $\phi$ ), used to remove small bright (opening) or dark (closing) details while leaving the overall features relatively undisturbed. The morphological reconstruction is usually jointly used with the opening and closing, that is, opening-by-reconstruction ( $\gamma_R$ ) and closing-by-reconstruction ( $\phi_R$ ), since this family of filters has proved to have better shape preservation than classical morphological operators (Soille and Pesaresi 2002). These operators are applied to an image with a set of a known shape, namely a SE. In this study, the morphological transformation is used to construct the structural spectrum with a series of multiscale and multidirectional SEs. The proposed structural spectrum consists of the morphological centre transformation and the morphological profiles (MPs) that are described in the following steps.

**Step 1** A pair of opening-closing dual filters, for example, opening followed by closing (OFC) and closing followed by opening (CFO), are used in this study considering that they act simultaneously on the bright and dark structures:

$$\text{OFC}_R = \phi_R \gamma_R, \tag{1}$$

$$\text{CFO}_R = \gamma_R \phi_R, \tag{2}$$

where the OFC and CFO filters are implemented with the reconstruction form.

**Step 2** The linear SE is used in this study because it contains information about the maximum object dimensions, and it is able to represent multidirectional characteristics of structures in the VHR imagery. Hence, equations (1) and (2) can be extended to

$$\text{OFC}_R(d, s) = \phi_R(d, s) \gamma_R(d, s), \quad (3)$$

$$\text{CFO}_R(d, s) = \gamma_R(d, s) \phi_R(d, s), \quad (4)$$

where  $d$  and  $s$  indicate the direction and size of a linear SE, respectively.

**Step 3** The OFC and CFO filters process bright and dark structures much more equally; however, they are not self-dual because their results depend on the sequence of opening and closing. Hence, we propose to construct their self-dual filter based on the morphological centre transformation:

$$f_{\text{mc}}(d, s) = \zeta \langle \text{OFC}_R(d, s), \text{CFO}_R(d, s) \rangle, \quad (5)$$

where the function  $\zeta \langle \cdot \rangle$  stands for the morphological centre transformation. The self-dual filters are of interest because they are able to avoid dependence with regard to the local contrast of image structures (Soille and Pesaresi 2002) and preserve the shape of structures while simplifying them. Readers can refer to Soille and Pesaresi (2002) for a detailed description about the morphological centre.

**Step 4** MPs are defined as a series of feature images with varying parameters of SEs (e.g. sizes and directions). DMPs are defined as vectors where the measures of the slopes of the MPs are stored for every step of an increasing SE series (Benediktsson *et al.* 2003). In this study, the DMPs are built on the morphological centre filter with multiscale and multidirectional SEs:

$$\text{DMP} = \{ |f_{\text{mc}}(d, s) - f_{\text{mc}}(d, s - \Delta s)|, s \in [m, n], d \in \text{D} \}, \quad (6)$$

where  $\text{D}$  is the set of directions:  $\text{D} = (45^\circ, 90^\circ, 135^\circ, 180^\circ)$ , and  $m, n$  and  $\Delta s$  represent the minimum, maximum and interval values for the lengths of the linear SE.

The application of the DMPs to the multi/hyperspectral images is conducted on a series of dimensionally reduced components (e.g. the principal component analysis), namely base images. Under this circumstance, the DMPs are converted to the EMPs (Benediktsson *et al.* 2005):

$$\begin{aligned} \text{EMP} &= \{ \text{dist} (f_{\text{mc}}(d, s), f_{\text{mc}}(d, s - \Delta s)), s \in [m, n], d \in \text{D} \} \\ &\text{with } f_{\text{mc}}(d, s) = \left( f_{\text{mc}}^b(d, s) \right)_{b=1}^B, \end{aligned} \quad (7)$$

where  $f_{\text{mc}}^b(d, s)$  stands for the morphological centre filter of band  $b$ ;  $\text{dist}(\cdot)$  represents the distance (Euclidean distance in this study) between two vectors; and  $B$  is the number of spectral bands.

An example of the morphological centre transformation is shown in figure 2. Figure 2(a) is the original image, and figures 2(b) and (c) are the OFC and CFO feature images, respectively. Figures 2(d) shows the morphological centre of figure 2(b)

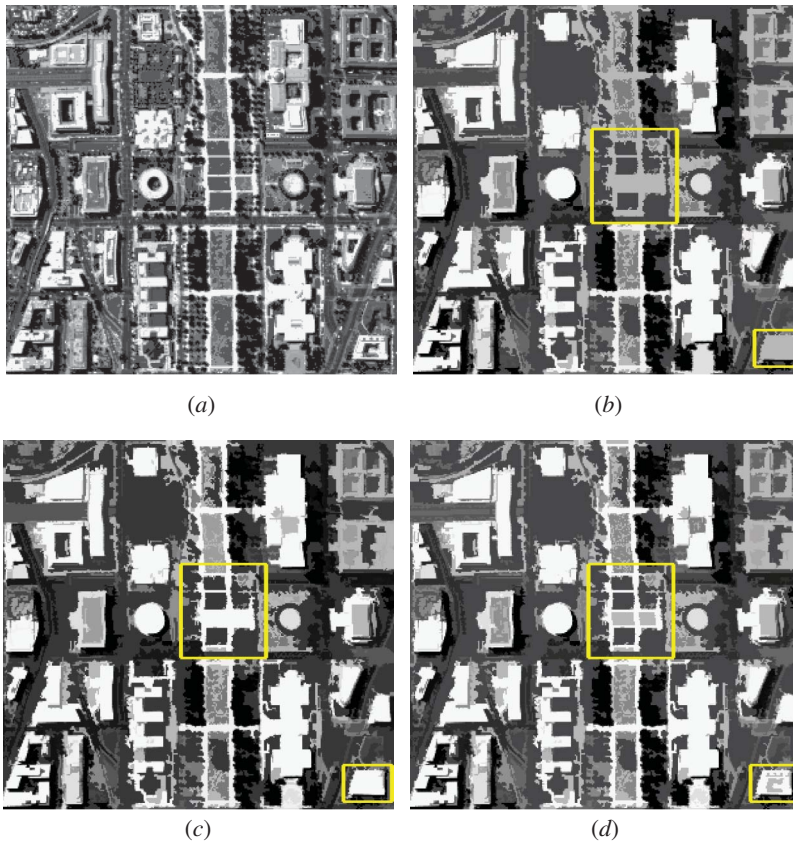


Figure 2. Demonstration of the morphological centre filter: (a) the original sample image, (b)  $OFC_R(d = 45, s = 15)$ , (c)  $CFO_R(d = 45, s = 15)$  and (d)  $f_{mc}(d = 45, s = 15)$ .

and (c). Comparing figure 2(b) with figure 2(c), it follows that their processing depends on whether one starts the filter with an opening or a closing. From figure 2(d), it can be seen that the morphological centre of the dual filters is able to effectively preserve the shape of structures when simplifying them (as highlighted in the two rectangular regions).

The radiance spectra and the structural spectra for some spectrally similar classes (e.g. water–shadows, grass–trees, trails–roofs–roads) in the Hyperspectral Digital Imagery Collection Experiment (HYDICE) data set are compared in figure 3. The  $x$ -axes of the radiance spectra and the structural spectra represent the dimensionality of the spectral and structural channels, respectively. Their  $y$ -axes represent the digital number (DN) value and feature value of the spectral and structural channels, respectively. From the figure, it can be found that in the VHR image, the spectral information alone is not adequate for discriminating the objects with similar spectral reflectance in spite of the availability of hundreds of channels. Nevertheless, it can be seen that the structural spectra has the potential to distinguish these spectrally similar classes, because it takes into account multiscale and multidirectional features with a series of anisotropic SEs.

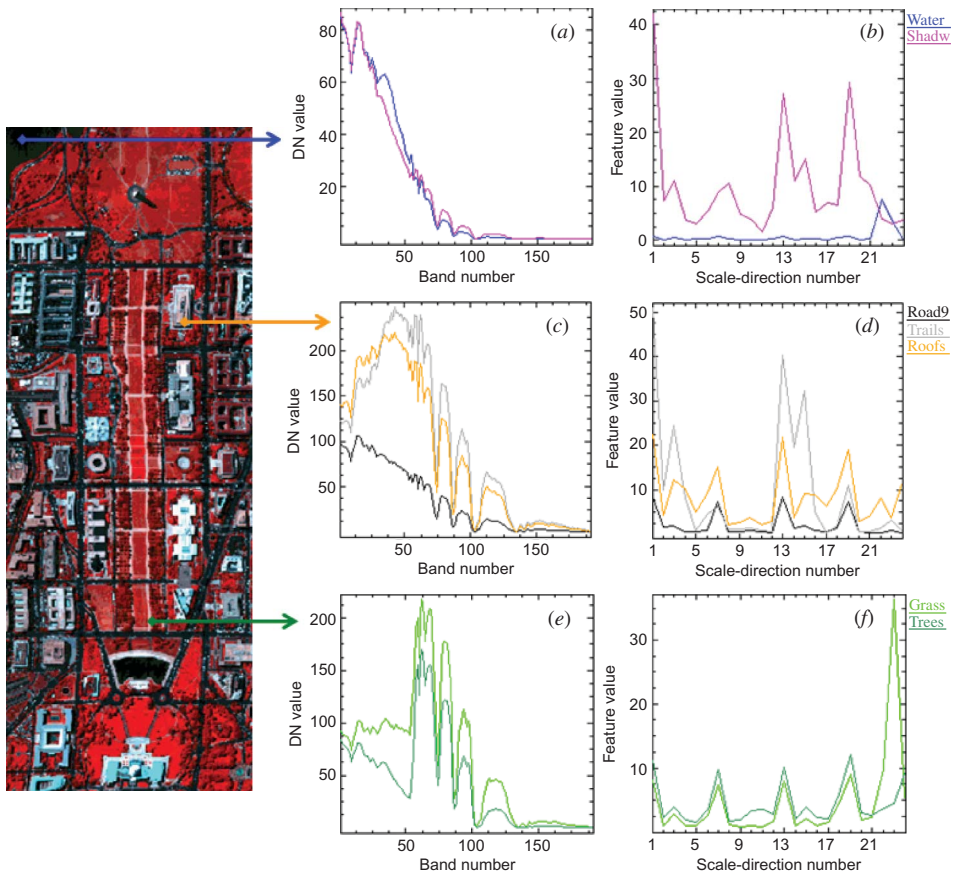


Figure 3. Comparison of the radiance spectra and the structural spectra for spectrally similar classes in the Hyperspectral Digital Imagery Collection Experiment (HYDICE) DC Mall data set: (a), (c) and (e) show the hyperspectra of water–shadow, roads–trails–roofs and grass–trees, respectively; (b), (d) and (f) show their structural spectra, respectively. DN, digital number.

## 2.2 Pixel-level decision fusion (level I)

A decision fusion scheme is proposed to integrate the spectral and structural information based on probabilistic outputs of classifiers. The spectral and structural features are separately interpreted in two SVMs, termed the spectral SVM and structural SVM. Afterwards, a weighted fusion algorithm is used for classification of each pixel. SVM is of interest due to its swift learning pace and adaptability to the high-dimensional feature space. Moreover, it is not constrained to prior assumptions on the distribution of input data and is hence well suited for interpretation of complex features, such as spectral–spatial joint classification (Huang *et al.* 2007a), object-based classification (Huang *et al.* 2008) and hyperdimensional stacking (Bruzzone and Carlin 2006). Most of the existing applications for the SVM refer to a crisp output that only predicts a class label. However, in this study, the probabilistic outputs of spectral SVM and structural SVM are exploited for integrating the spectral and structural decision information. The algorithm consists of two steps.

**Step 1** The spectral and structural features are separately fed into two SVMs. The SVM probabilistic output is estimated by combining the pairwise class probabilities (Wu *et al.* 2004). We assume  $P_{\text{spe}}(c|x)$  and  $l_{\text{spe}}(x)$  as the probabilistic output (soft) and the classification prediction (crisp) of the spectral SVM, respectively. Similarly,  $P_{\text{str}}(c|x)$  and  $l_{\text{str}}(x)$  are the probability and the classification label of the structural SVM, respectively. The relationship between the probabilistic output and the class label for pixel  $x$  can be expressed as

$$l_{\text{spe}}(x) = \arg \max_{c \in \{1, 2, \dots, C\}} (P_{\text{spe}}(c|x)), \quad (8)$$

$$l_{\text{str}}(x) = \arg \max_{c \in \{1, 2, \dots, C\}} (P_{\text{str}}(c|x)), \quad (9)$$

where  $c$  stands for an information class.

**Step 2** The spectral–structural decision rule is defined as

$$l(x) = \arg \max_{c \in \{1, 2, \dots, C\}} (p(c|x)), \quad (10)$$

with

$$p(c|x) = \frac{w_{\text{spe}}(c)P_{\text{spe}}(c|x) + w_{\text{str}}(c)P_{\text{str}}(c|x)}{w_{\text{spe}}(c) + w_{\text{str}}(c)}, \quad (11)$$

where  $w(c)$  is the weight of the spectral (or structural) probability, and it is defined as the  $F$ -measure of the producer's and user's accuracies (Shah *et al.* 2010):

$$w(c) = \frac{2 A_p(c) A_u(c)}{A_p(c) + A_u(c)}, \quad (12)$$

where  $A_p(c)$  and  $A_u(c)$  are the producer's and user's accuracies of class  $c$ , respectively. In this article, the reference samples are divided into three parts: training, validation and test samples. The training and test samples are used for training and evaluation of classification, while the validation samples are used for calculation of  $w(c)$ .

Equation (10) is actually a pixel-level classification combining the decision information of both spectral SVM and structural SVM. It should be noted that equation (10) refers to a crisp classification result of the spectral–structural fusion, while the output of equation (11) refers to a soft one. The soft output  $p(c|x)$  provides the probability that a pixel  $x$  is assigned to a class  $c$ . The probabilistic output in equation (11) can be further employed in an object-based decision fusion.

### 2.3 Object-based decision fusion (levels II and III)

The object-based decision fusion consists of two levels:

1. At level II, the pixel-level decision fusion is extended to an object-based decision fusion by considering the probabilistic outputs within the boundary of

each object. The objects having low probabilistic outputs are considered as unreliable and are processed in the next level.

2. At level III, shape attributes of the unreliable objects are used for improving the classification of level II.

The algorithm is detailed as follows.

**Step 1** Segmentation: an adaptive mean shift segmentation algorithm (Huang and Zhang 2008) is used to extract boundaries of objects. The mean shift algorithm is selected in this study since recently it has received increasing attention for segmentation of remotely sensed data (Chaabouni-Chouayakh and Datcu 2010), and in our previous work it showed better performance than the widely used FNEA segmentation (Huang and Zhang 2008).

**Step 2** Object-based decision fusion at level II is implemented by considering the pixel-level probabilistic outputs of all the pixels within the object boundary:

$$l(\text{obj}) = \arg \max_{c=1,2,\dots,C} (p(c|\text{obj})), \quad (13)$$

where  $l(\text{obj})$  is the winning label of the object and  $p(c|\text{obj}) = \sum_{x \in \text{obj}} p(c|x)$  indicates the sum of probabilistic outputs of all the pixels for class  $c$  in the object.

**Step 3** Object merging: a decision-level object merging algorithm is used to merge the spatially adjacent objects with the same class. The following processing is repeated until no adjacent object pairs can be further merged:

**IF** (obj1 is adjacent to obj2) **AND**  $l(\text{obj1}) = l(\text{obj2})$   
**THEN** obj1 and obj2 are merged as one object

**Step 4** Post-processing: this step only focuses on the unreliable objects that satisfy the following condition:

$$p(\hat{c}|\text{obj}) = \max_c (p(c|\text{obj})) < t, \quad (14)$$

where  $\hat{c} = l(\text{obj})$  is the winning label of obj. The threshold  $t \in (0,1)$  for identification of the unreliable objects should be chosen according to the characteristics of different images. Shape attributes (e.g. length–width ratio) of the unreliable objects are used for their reclassification. It should be noted that shape information is used only for objects where it can bring improvements since a blind consideration of shape information to all objects could decrease the overall classification results. Therefore, in this study, the shape attributes are only considered for the unreliable objects. Specifically, in our experiments, the length–width ratio is used for discrimination between roofs and roads. Road objects generate large values since they are relatively elongated and narrow, while roof objects result in small values since they are rectangular and compact.

A graphical example of the processing chain is demonstrated in figure 4. It is clearly shown that: (1) the segmentation after the object merging is more effective in representing the actual shape of the objects than the initial segmentation; (2) the unreliable objects are determined according to the probabilistic outputs of all the pixels in each object; and (3) the shape attribute of the unreliable objects is then used for refinement of classification.

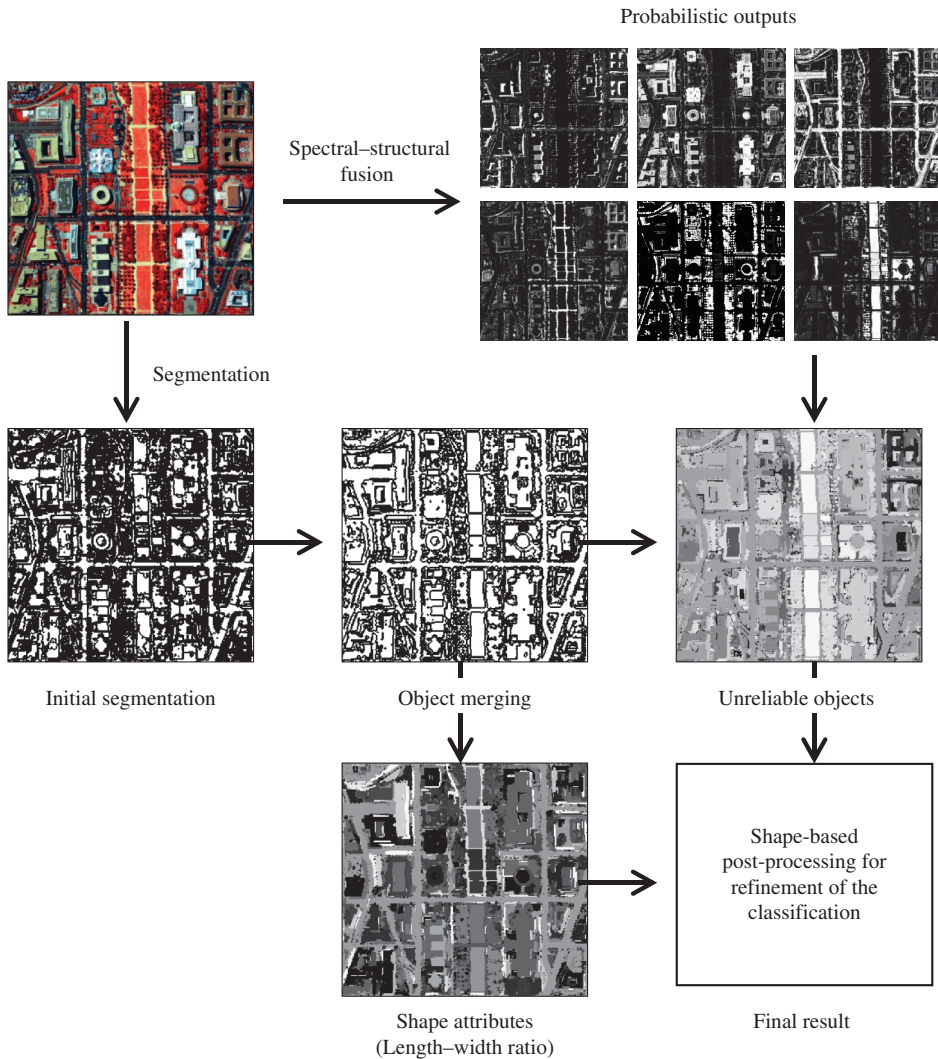


Figure 4. A graphical example of the processing chain. In the images of the probabilistic outputs, a large brightness value stands for a high possibility that a pixel belongs to a class. For the image showing the unreliable objects, the grey level represents the value of  $p(\hat{c} | \text{obj})$  defined in equation (14), where a large brightness value means a high probability and a small uncertainty for the classification. As for the feature image of length-width ratio, a large brightness stands for a large ratio.

### 3. Experiments

In this section, two VHR data sets are used to validate the effectiveness and robustness of the proposed multilevel decision fusion approach. The HYDICE DC Mall image is a standard data set for evaluation of segmentation and classification algorithms. The other data set includes a pan-sharpened QuickBird multispectral image in Beijing, the capital of China.

### 3.1 HYDICE DC Mall data set

The HYDICE airborne hyperspectral data flight over the Mall in Washington, DC was used in this experiment. Two hundred and ten bands were collected in the 0.4–2.4  $\mu\text{m}$  region of the visible and infrared spectrum. The water absorption bands were then deleted, resulting in 191 channels. This data set contains 1280 scan lines with 307 pixels in each scan line. In this experiment, the non-negative matrix factorization (NMF) transformation (Lee and Seung 1999) was used to extract three-dimensional base images from the hyperspectral channels and yields the morphological structural spectra. Training, validation and test samples are shown in table 1.

Accuracies of the multilevel decision fusion are listed in table 2, where the class-specific accuracies are the average of the producer's accuracy (PA) and user's accuracy (UA). The kappa coefficient and the overall accuracy (OA) are computed based on the confusion matrix. The average accuracy (AA) is also reported since it is relatively independent of the number of samples for each class (Benediktsson *et al.* 2003). Based on table 2, we have the following observations.

1. At level I. The spectral and structural features were integrated using weight decision fusion (equation (10)). At this level, the decision fusion increased the overall accuracies from 91.0% (spectral SVM) and 93.8% (structural SVM) to 95.2%.

Table 1. Numbers of the reference samples (HYDICE data set).

Class	Number of training pixels	Number of validation pixels	Number of test pixels
Road	110	333	3 334
Grass	111	308	3 075
Water	110	288	2 882
Trail	114	103	1 034
Tree	110	205	2 047
Shadow	111	109	1 093
Roof	111	587	5 867
Total	777	1 933	19 332

Note: HYDICE, Hyperspectral Digital Imagery Collection Experiment.

Table 2. Accuracies of the multilevel decision fusion (HYDICE data set).

Class	Spectral SVM	Structural SVM	Level I	Level II	Level III
Road	91.4	91.6	92.3	91.9	97.9
Grass	97.1	98.0	97.8	98.5	98.5
Water	94.8	99.8	99.2	99.9	99.9
Trail	76.1	82.4	87.7	93.4	95.4
Tree	97.2	98.0	97.8	98.9	98.9
Shadow	84.2	91.4	94.6	96.5	96.5
Roof	90.2	92.4	94.5	95.2	99.3
OA (%)	91.0	93.8	95.2	96.0	98.6
AA (%)	90.1	93.4	94.8	96.3	98.1
$\kappa$	0.891	0.925	0.941	0.951	0.983

Note: HYDICE, Hyperspectral Digital Imagery Collection Experiment; SVM, support vector machine; OA, overall accuracy; AA, average accuracy.

2. At level II. In comparison with the level I, OA and kappa at level II increased by 0.8% and 1.0%, respectively. The increments were achieved by extending the spectral–structural decision fusion from the pixel level to the object level.
3. At level III. The length–width ratio was used for classification of the unreliable objects ( $t = 0.46$ ). The rule is defined as  
 FOR ( $\hat{c} = \text{roads or trails}$ ) AND ( $p(\hat{c} | \text{obj}) < t$ )  
 IF Ratio  $< 2.5$   
 THEN  $\hat{c} = \text{roads}$ .

The objective of this shape-based rule is to avoid the misclassifications between roads, trails and roofs. In table 2, it can be seen that the post-processing at level III gave very promising results (OA = 98.6% and  $\kappa = 0.983$ ). In comparison with level II, accuracies of the man-made structures were improved by 6.0%, 2.0% and 4.1 % for roads, trails and roofs, respectively.

The results of the proposed multilevel fusion and the pixel-level classification are compared in figure 5. Their confusion matrices are provided in table 3. Subsets of the classification maps are displayed in figure 6. Figure 6(a) shows the pixel-level spectral

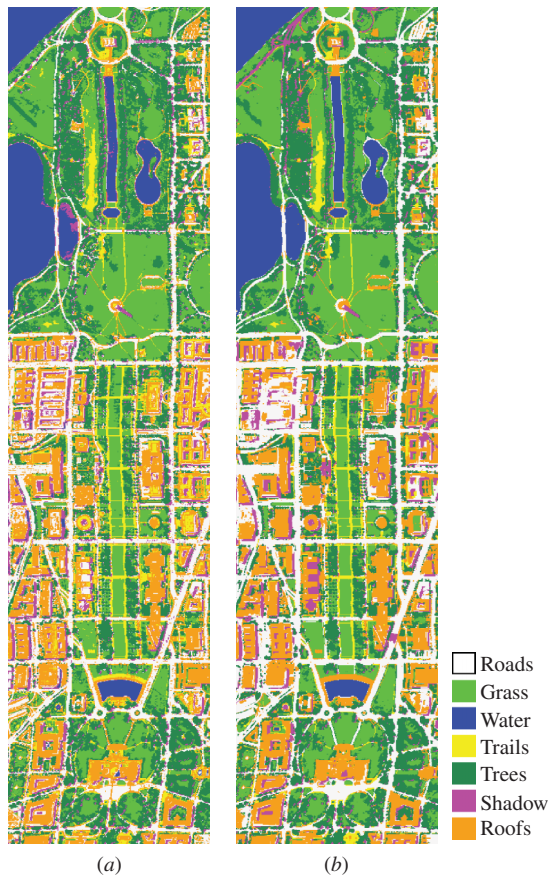


Figure 5. Classification results of the airborne hyperspectral data over the Washington DC Mall: (a) support vector machine (SVM)-based pixel-level interpretation and (b) the multilevel decision fusion.

Table 3. Confusion matrices of the HYDICE data set for (a) SVM-based pixel-level classification and (b) the multilevel decision fusion.

(a) Confusion matrix for SVM-based pixel-level classification (figure 5(a))									
	Road	Grass	Water	Trail	Tree	Shadow	Roof	Total	UA (%)
Road	3 096	0	0	3	3	4	330	3 436	90.1
Grass	4	2 953	0	41	7	0	2	3 007	98.2
Water	0	0	2 663	0	0	72	0	2 735	97.3
Trail	10	45	0	978	0	0	663	1 696	57.6
Tree	19	76	0	0	2 032	1	7	2 135	95.1
Shadow	103	0	219	0	5	1 016	4	1 347	75.4
Roof	102	1	0	12	0	0	4 861	4 976	97.6
Total	3 334	3 075	2 882	1 034	2 047	1 093	5 867	19 332	
PA (%)	92.8	96.0	92.4	94.5	99.2	92.9	82.8		
OA = 91.0%, $\kappa = 0.891$									
(b) Confusion matrix for the multilevel decision fusion (figure 5(b))									
	Road	Grass	Water	Trail	Tree	Shadow	Roof	Total	UA (%)
Road	3 277	0	0	5	2	52	27	3 363	97.4
Grass	4	3 043	0	44	9	0	2	3 102	98.1
Water	0	0	2 875	0	0	0	0	2 875	100.0
Trail	0	29	0	981	0	0	13	1 023	95.9
Tree	22	0	0	0	2 034	1	8	2 065	98.5
Shadow	10	0	6	0	2	1 036	3	1 057	98.0
Roof	21	3	1	4	0	4	5 814	5 847	99.4
Total	3 334	3 075	2 882	1 034	2 047	1 093	5 867	19 332	
PA (%)	98.3	99.0	99.8	94.9	99.4	94.8	99.1		
OA = 98.6%, $\kappa = 0.983$									

Note: HYDICE, Hyperspectral Digital Imagery Collection Experiment; SVM, support vector machine; OA, overall accuracy; PA, producer's accuracy; UA, user's accuracy.

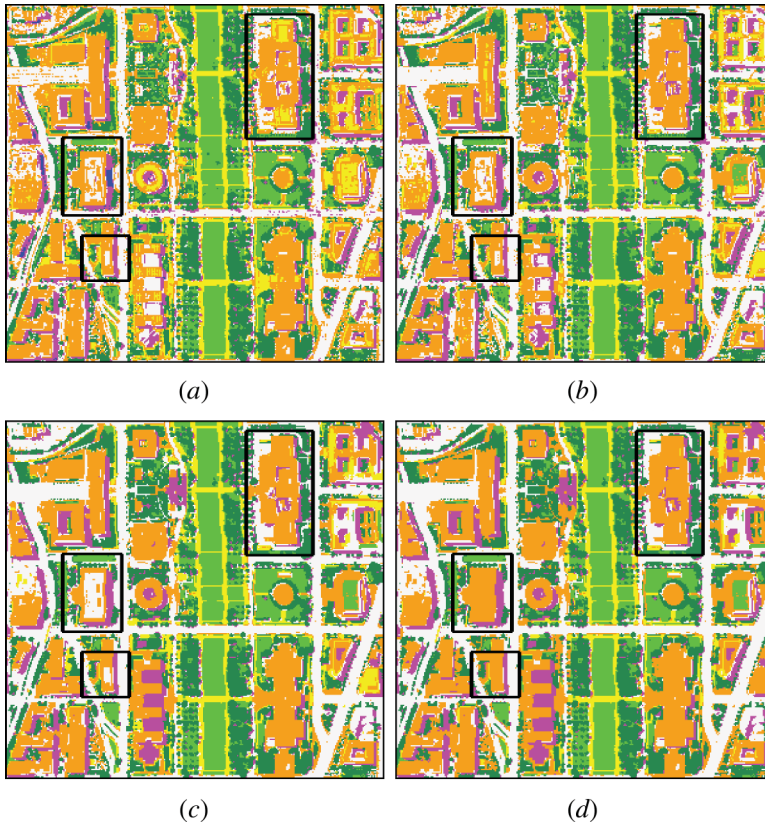


Figure 6. Subsets of classification maps of the Hyperspectral Digital Imagery Collection Experiment (HYDICE) DC Mall image: (a) pixel-level classification using the hyperspectral bands with the support vector machine (SVM) classifier; (b), (c) and (d) classification results of levels I, II and III, respectively.

classification using the hyperspectral bands with a SVM classifier, and figure 6(b)–(d) are the classification maps of the proposed method for levels I, II and III, respectively. Some misclassifications are highlighted with several rectangular regions. The pixelwise classification (figure 6(a)) clearly shows the pepper–salt effect and misclassifications between spectrally similar classes. The spectral–structural classification in figure 6(b) revises some errors of water–shadow and trails–roofs. The object-based decision fusion (figure 6(c)) is able to reduce the uncertainties and the pepper–salt effect of pixelwise classification. The performance of the shape-based post-processing can be observed in figure 6(d), where some misclassifications for trails–roofs and roads–roofs in figure 6(a)–(c) are corrected.

### 3.2 QuickBird Beijing data set

In this experiment, a pan-sharpened QuickBird image of Beijing with a spatial resolution of 0.61 m is used to validate the proposed multilevel model. The study area is shown in figure 7(a). It is a typical urban landscape in China, including dense residential and apartments with gardens. The training, validation and test samples are

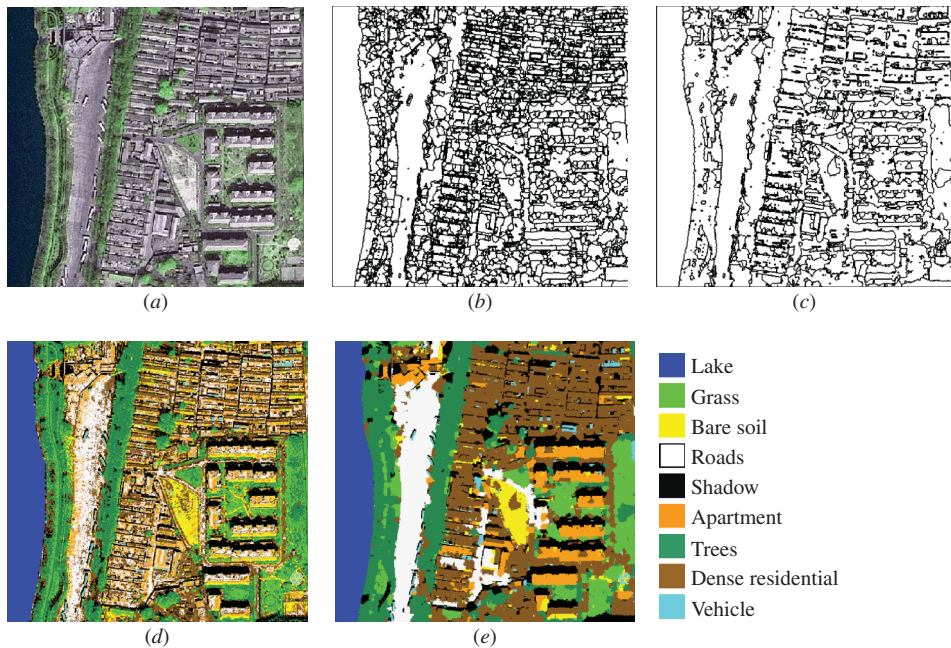


Figure 7. (a) QuickBird colour image of the study area in the northwest of Beijing, the capital city of China, (b) mean shift segmentation result and (c) refined segmentation map with the decision-level object merging. (d) and (e) Classification maps for the pixel-level spectral classification and the proposed multilevel fusion approach, respectively.

Table 4. Numbers of the reference samples (Beijing QuickBird data set).

Class	Number of training pixels	Number of validation pixels	Number of test pixels
Lake	60	300	6 986
Grass	60	300	1 654
Bare soil	60	300	2 806
Road	60	300	6 603
Shadow	60	300	2 062
Apartment	60	300	7 630
Tree	60	300	4 691
Dense residential	60	300	7 414
Vehicle	60	300	592
Total	540	2 700	40 438

listed in table 4. The challenge for classifying this area is to discriminate between the spectrally similar classes such as bare soil, roads, residential area, apartment and vehicle.

Accuracies of the multilevel decision fusion are reported in table 5. From this table, a similar conclusion is obtained as the HYDICE experiment. In the QuickBird experiment, the accuracy increments achieved by the multilevel framework are more significant. The results are very promising since the OA increased from 65.5% (spectral classification) to 78.9%, 87.2% and 93.3% for decision fusion at levels I, II and

Table 5. Accuracies of the multilevel decision fusion (Beijing QuickBird data set).

Class	Spectral SVM	Structural SVM	Level I	Level II	Level III
Lake	100.0	99.1	100.0	100.0	100.0
Grass	73.1	48.0	78.6	93.4	93.4
Bare soil	60.9	79.4	79.6	91.0	91.0
Road	62.5	65.9	71.8	81.7	92.3
Shadow	88.7	92.2	90.8	96.2	96.2
Apartment	33.1	68.7	67.4	74.0	90.8
Tree	86.4	83.1	90.7	97.2	97.2
Dense residential	50.1	70.2	72.0	89.4	89.4
Vehicle	57.7	74.8	74.5	85.7	85.7
OA (%)	65.5	75.1	78.9	87.2	93.3
AA (%)	68.1	75.7	80.6	89.8	92.9
$\kappa$	0.600	0.710	0.754	0.850	0.922

Note: SVM, support vector machine; OA, overall accuracy; AA, average accuracy.

III, respectively. Once again, in table 5, it is seen that the accuracy of each class increased when structural (level I), boundary (level II) and shape (level III) features were gradually included in the processing chain.

The segmentation and classification maps are displayed in figure 7, where figure 7(b) is the mean shift-based initial segmentation, and figure 7(c) is the refined segmentation after the decision-level object merging. The classification maps for the pixel-level spectral classification and the proposed multilevel decision fusion are compared in figures 7(d) and (e). Their confusion matrices are provided in table 6. The comparison in table 6 reveals that classification for all the information classes are improved with the multilevel method. This observation is further supported by the visual inspection in figure 7.

#### 4. Conclusion

In this study, we presented a novel multilevel model for effectively integrating spectral, structural and geometrical features for classification of the VHR multi/hyperspectral imagery. The model is progressively enhanced from a low level (pixel) to a high (object) level when the multilevel features are gradually considered in the processing chain. Moreover, the model is adaptively extended by considering the probabilistic outputs of each level. The main contributions of this study are summarized as follows.

1. At level I. A morphological centre filter was used to construct the self-dual operators and extend the algorithm of MPs. Furthermore, a weight probability decision fusion was proposed to integrate spectral and structural information.
2. At level II. The spectral–structural decision fusion was extended from the pixel level to the object level by considering probabilistic outputs within the boundary of each object. The results showed that this extension improved the accuracies of level I by 0.8% and 8.3% for the HYDICE and QuickBird data sets, respectively.
3. At level III. This level only focused on the unreliable objects of level II. The shape attribute of the unreliable objects was extracted for refining the result of level II. The accuracy increments achieved by the shape-based post-processing were 2.6% and 6.1% for the HYDICE and QuickBird data sets, respectively.

Table 6. Confusion matrices for (a) SVM-based pixel-level classification and (b) the proposed multilevel decision fusion (QuickBird data set).

(a) Confusion matrix for SVM-based pixel-level classification (figure 7(d))										
Lake	Grass	Bare soil	Road	Shadow	Apartment	Trees	Dense residential	Vehicle	Total	UA (%)
Lake	6 986	0	0	0	0	0	0	0	6 986	100.0
Grass	1 432	2	0	0	24	931	7	0	2 396	59.7
Bare soil	29	2 057	11	0	1 697	0	421	19	4 234	48.5
Road	0	24	4 802	0	2 398	0	1 896	49	9 169	52.3
Shadow	0	18	24	2 036	101	71	321	12	2 583	78.8
Apartment	0	0	291	921	1 935	0	1 510	68	4 725	40.9
Tree	0	191	3	2	0	3 670	12	0	3 878	94.6
Dense residential	0	2	320	839	1 103	19	3 168	19	5 494	57.6
Vehicle	0	0	91	0	372	0	79	425	973	43.6
Total	6 986	1 654	2 806	6 603	7 630	4 691	7 414	592	40 438	
PA (%)	100.0	86.5	73.3	72.7	98.7	78.2	42.7	71.7		
OA = 65.5%, $\kappa = 0.600$										
(b) Confusion matrix for the multilevel decision fusion (figure 7(e))										
Lake	Grass	Bare soil	Roads	Shadow	Apartment	Trees	Dense residential	Vehicle	Total	UA (%)
Lake	6 986	0	0	0	0	0	0	0	6 986	100.0
Grass	1 644	0	0	0	47	166	25	0	1 882	87.4
Bare soil	0	2 552	6	0	189	0	54	1	2 802	91.1
Road	0	0	6 432	0	432	0	468	49	7 381	87.1
Shadow	0	6	1	2 019	27	19	56	7	2 135	94.6
Apartment	0	0	69	0	6 504	0	158	18	6 749	96.4
Tree	0	10	0	31	0	4 488	6	15	4 550	98.6
Dense residential	0	0	228	12	415	18	6 635	25	7 428	89.3
Vehicle	0	0	20	0	16	0	12	477	525	90.9
Total	6 986	1 654	2 806	6 603	7 630	4 691	7 414	592	40 438	
PA (%)	100.0	99.4	91.0	97.4	97.9	95.7	89.5	80.6		
OA = 93.3%, $\kappa = 0.922$										

Note: SVM, support vector machine; OA, overall accuracy; PA, producer's accuracy; UA, user's accuracy.

Future work will be related to the use of some standard techniques in the processing chain. Some other methods for the spatial features (e.g. textural bands), classifiers (e.g. the maximum likelihood classification and neural networks) and segmentation (e.g. the FNEA and  $J$ -measure-based approach) can be also used in the proposed multilevel system. Hence, it will be interesting to evaluate and compare their performance in future experiments.

### Acknowledgement

This work was supported by the National Science Foundation of China (41061130553), the Fundamental Research Funds for the Central Universities (3101016) and the LIESMARS Special Research Funding.

### References

- BELLENS, R., GAUTAMA, S., MARTINEZ-FONTE, L., PHILIPS, W., CHAN, J.C.W. and CANTERS, F., 2008, Improved classification of VHR images of urban areas using directional morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 2803–2813.
- BENEDIKTSSON, J.A., PALMASON, J.A. and SVEINSSON, J.R., 2005, Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, pp. 480–491.
- BENEDIKTSSON, J.A., PESARESI, M. and ARNASON, K., 2003, Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 1940–1949.
- BENZ, U.C., HOFMANN, P., WILLHAUCK, G., LINGENFELDER, I. and HEYNEN, M., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58**, pp. 239–258.
- BRUZZONE, L. and CARLIN, L., 2006, A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2587–2600.
- CHAABOUNI-CHOUAYAKH, H. and DATCU, M., 2010, Coarse-to-fine approach for urban area interpretation using TerraSAR-X data. *IEEE Geoscience and Remote Sensing Letters*, **7**, pp. 78–82.
- CHANUSSOT, J., BENEDIKTSSON, J.A. and FAUVEL, M., 2006, Classification of remote sensing images from urban areas using a fuzzy possibilistic model. *IEEE Geoscience and Remote Sensing Letters*, **3**, pp. 40–44.
- CHUBEY, M.S., FRANKLIN, S.E. and WULDER, M.A., 2006, Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 383–394.
- FAUVEL, M., BENEDIKTSSON, J.A., CHANUSSOT, J. and SVEINSSON, J.R., 2008, Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 3804–3814.
- FAUVEL, M., CHANUSSOT, J. and BENEDIKTSSON, J.A., 2009, Kernel principal component analysis for the classification of hyperspectral remote-sensing data over urban areas. *EURASIP Journal on Advances in Signal Processing*, doi: 10.1155/2009/783194.
- HUANG, X. and ZHANG, L., 2008, An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 4173–4185.
- HUANG, X., ZHANG, L. and LI, P., 2007a, Classification and extraction of spatial features in urban areas using high resolution multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 260–264.

- HUANG, X., ZHANG, L. and LI, P., 2007b, An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 654–658.
- HUANG, X., ZHANG, L. and LI, P., 2008, A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform. *International Journal of Remote Sensing*, **29**, pp. 5923–5941.
- LEE, D.D. and SEUNG, H.S., 1999, Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, pp. 788–791.
- LI, P. and XIAO, X., 2007, Multispectral image segmentation by a multichannel watershed-based approach. *International Journal of Remote Sensing*, **28**, pp. 4429–4452.
- MYINT, S.W., LAM, N.S.N. and TYLER, J., 2004, Wavelets for urban spatial feature discrimination: comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 803–812.
- OUMA, Y.O., NGIGI, T.G. and TATEISHI, R., 2006, On the optimization and selection of wavelet texture for feature extraction from high-resolution satellite imagery with application towards urban-tree delineation. *International Journal of Remote Sensing*, **27**, pp. 73–104.
- OUMA, Y.O., TETUKO, J. and TATEISHI, R., 2008, Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery. *International Journal of Remote Sensing*, **29**, pp. 3417–3456.
- PACIFICI, F., CHINI, M. and EMERY, W.J., 2009, A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, **113**, pp. 1276–1292.
- PUISSANT, A., HIRSCH, J. and WEBER, C., 2005, The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing*, **26**, pp. 733–745.
- SHAH, V.P., YOUNAN, N.H., DURBHA, S.S. and KING, R.L., 2010, Feature identification via a combined ICA–wavelet method for image information mining. *IEEE Geosciences and Remote Sensing Letters*, **7**, pp. 18–22.
- SOILLE, P. and PESARESI, M., 2002, Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, pp. 2042–2055.
- VAN COILLIE, F.M.B., VERBEKE, L.P.C. and DE WULF, R.R., 2007, Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment*, **110**, pp. 476–487.
- WANG, Z. and BOESCH, R., 2007, Color- and texture-based image segmentation for improved forest delineation. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, pp. 3055–3062.
- WU, T.F., LIN, C.J. and WENG, R.C., 2004, Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, **5**, pp. 975–1005.
- YU, Q., GONG, P., CLINTON, N., BIGING, G. and SCHIROKAUER, D., 2006, Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 799–811.
- ZHOU, W., HUANG, G., TROY, A. and CADENASSO, M.L., 2009, Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: a comparison study. *Remote Sensing of Environment*, **113**, pp. 1769–1777.