An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery

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

***Abstract*—In recent years, the resolution of remotely sensed imagery has become increasingly high in both the spectral and spatial domains, which simultaneously provides more plentiful spectral and spatial information. Accordingly, the accurate interpretation of high-resolution imagery depends on effective integration of the spectral, structural and semantic features contained in the images. In this paper, we propose a new multifeature model, aiming to construct a support vector machine (SVM) ensemble combining multiple spectral and spatial features at both pixel and object levels. The features employed in this study include a gray-level co-occurrence matrix, differential morphological profiles, and an urban complexity index. Subsequently, three algorithms are proposed to integrate the multifeature SVMs: certainty voting, probabilistic fusion, and an object-based semantic approach, respectively. The proposed algorithms are compared with other multifeature SVM methods including the vector stacking, feature selection, and composite kernels. Experiments are conducted on the hyperspectral digital imagery collection experiment DC Mall data set and two WorldView-2 data sets. It is found that the multifeature model with semantic-based postprocessing provides more accurate classification results (an accuracy improvement of 1–4% for the three experimental data sets) compared to the voting and probabilistic models.**

***Index Terms*—Classification, feature extraction, high resolution, morphological, multifeature, object-based, semantic, support vector machines (SVMs), WorldView-2.**

I. Introduction

IN RECENT years, with the rapid development of space imaging techniques, remote sensors can provide high-resolution Earth observation data in both the spectral and spatial domains at the same time. Some airborne platforms, such as the hyperspectral digital imagery collection experiment (HYDICE), hyperspectral mapper, and reflective optics systems imaging spectrometer, provide multi/hyperspectral channels (100–300 bands) with a spatial resolution of 1–5 m. More interestingly, the recently launched WorldView-2 satellite [1], based on a spaceborne platform, is able to provide eight multispectral bands with a 2-m spatial resolution. This new type of high-resolution imagery contains detailed ground information in both the spectral and spatial domains; it therefore opens new avenues for remote sensing applications in urban mapping, forest monitoring, environment management, precision agriculture, and security and defense issues, etc.

However, it should be noted that higher resolutions do not naturally result in higher interpretation accuracies. On the one hand, the classification of hyperspectral images is subject to the so-called Hughes phenomenon, or the curse of dimensionality problem, due to the small ratio between the number of training samples and the number of features [2]. In other words, the high redundancy between hyperspectral bands with increasing spectral dimensionality might cause problems during data analysis, e.g., reduction of classification accuracy, particularly in the case of a small sample size [3]. On the other hand, the classification of high spatial resolution images suffers from uncertainty of the spectral information because the increase of the intra-class variance and decrease of the inter-class variance lead to a decrease of the separability in the spectral domain, particularly for the spectrally similar classes [4]. Taking into account the aforementioned two aspects, it is widely agreed that the accurate interpretation of multi/hyperspectral imagery with high spatial resolution relies on effective spectral-spatial joint feature extraction and classification. Accordingly, in recent years, a few studies have been reported on this topic. Dell'Acqua *et al.* presented a first assessment of spatial analysis algorithms for detailed urban classification of high-resolution hyperspectral data [5]. The experimental results verified the better performance of spatial classification compared to the pure spectral method. Benediktsson *et al.* proposed extended morphological profiles (EMP) for the classification of urban hyperspectral data with a high spatial resolution [6]. The principal components (PCs) of the hyperspectral imagery were used as base images for the construction of MPs, and, subsequently, the spectral PCs stacked with the EMP were input into a neural network classifier. Fauvel *et al.* [7] improved the method in [6] by concatenating the hyperspectral information and the MPs into one feature vector, since the original method did not fully utilize the spectral information in the data. Chini *et al.* [8]

Manuscript received September 24, 2011; revised January 16, 2012, March 24, 2012, and May 16, 2012; accepted May 26, 2012. Date of publication July 13, 2012; date of current version December 19, 2012. This work was supported by the National Natural Science Foundation of China under Grants 41101336 and 41061130553, the Program for New Century Excellent Talents in University of China under Grant NCET-11-0396, and the Research Fund for the Doctoral Program of Higher Education of China under Grant 20110141120072.

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

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

Digital Object Identifier 10.1109/TGRS.2012.2202912

TABLE I
RECENT LITERATURE ON MULTIFEATURE FUSION FOR THE CLASSIFICATION OF HIGH-RESOLUTION REMOTELY SENSED IMAGERY (DR = DIMENSIONALITY REDUCTION, NWFE = NONPARAMETRIC WEIGHTED FEATURE EXTRACTION, DBFE = DECISION BOUNDARY FEATURE EXTRACTION, RFE = RECURSIVE FEATURE ELIMINATION)

Literature	Classifier	DR Method	Multifeature
Benediktsson *et al.* [6]	Neural network	NWFE [18]	Spectral and EMP
Fauvel *et al.* [7]	SVM	DBFE [18]	Spectral and EMP
Bruzzone and Carlin [2]	SVM	NA	Spectral and multilevel object-based features
Tuia *et al.* [19]	SVM	RFE [20]	Spectral and morphological features
Tuia *et al.* [21]	SVM (composite kernels)	NA	Spectral and morphological features
Tuia *et al.* [22]	SVM (multi-kernels)	Correlation filter, and RFE	Spectral, morphological features, and multisource data
Tuia *et al.* [23]	Semisupervised SVM	NA	Spectral and morphological features
Pacifici *et al.* [24]	Neural network	Neural network pruning	GLCM textures
Chen *et al.* [25]	SVM	Bhattacharyya distance [26]	Spectral magnitude and shape features
Bau *et al.* [27]	Mahalanobis distance [28]	Mahalanobis distance [28]	Spectral and Gabor features
Huang *et al.* [29]	SVM	NA	Spectral and structural feature set
Huang *et al.* [30]	SVM	NA	Spectral and wavelet features
Huang and Zhang [31]	SVM	NA	Spectral and multiscale features

extended the MPs by implementing a series of anisotropic structural elements with a triangular shape. The anisotropic morphological features were then selected and classified using a multilayer perceptron neural network. Dalla Mura *et al.* [9] proposed to characterize the spatial information of high-resolution data by using a multilevel, multi-attribute approach (e.g., area, moment of inertia, and standard deviation) based on morphological attribute filters, leading to a more complete description of the scene and to a more accurate modeling of the spatial information than the traditional MPs. Subsequently, they extended the attribute profiles to the classification of hyperspectral images by implementing the transformation on the spectral independent components [10]. Huang and Zhang [11] conducted a comparative study of spatial approaches for urban mapping, using hyperspectral imagery with a high spatial resolution from Pavia City, northern Italy. Different spectral-spatial feature extraction and classification methods were implemented, including differential MPs (DMPs) [12], gray-level co-occurrence matrix (GLCM), pixel shape index (PSI) [13], object-based classification using the fractal net evolution approach [14], and the multiscale mean-shift procedure [15]. Results showed that the spectral-spatial approaches could effectively improve the mapping accuracy of pure spectral classification. In addition, the DMP and the multiscale mean-shift approach achieved better performance compared to other spectral-spatial methods. Recently, Tarabalka *et al.* [16] proposed the use of probability estimates obtained by the support vector machine (SVM) classification, in order to determine the most reliable classified pixels as seeds of spatial regions. Subsequently, the rest of the pixels were classified by constructing a minimum spanning forest on the reliable pixels. This spectral-spatial classification method was further improved by performing a multiple classification scheme on the selection of reliable pixels [17].

By summarizing the existing literature, it can be found that all the studies underline the important role of spatial information for the classification of high-resolution imagery. However, it should be recognized that although various spatial features are currently available for high-resolution image processing, such as morphological features [6]–[10], [12], structural feature set [29], PSI [13], wavelet-based texture [30], object-based features [32], and GLCM [4], [33], it is impossible to find one feature that is optimal for different image scenes. The traditional approach for addressing this issue is to use a vector stacking (VS) approach for the integration of multiple features, i.e., concatenate the multiple features and feed them into a classifier with a preprocessing of dimensionality reduction (see Table I). VS is frequently used for multifeature fusion as it is simple to carry out and is potential to enhance the separability between similar objects by forming a hyperdimensional multifeature space [31]. Furthermore, the VS approach is often jointly used with an SVM classifier since SVM is not constrained to prior assumptions on the distribution of input data, and it enables the weighting of the different features [34].

Although the existing studies show that "VS-SVM" is a feasible approach for multifeature fusion, it has several drawbacks. Calculation of the spatial features, such as the GLCM, DMP, and wavelet features, in most cases leads to hyperdimensional feature space since spatial features refer to different parameters such as sizes, scales, and directions. However, a recent study shows that the classification accuracy by an SVM varies as a function of the number of features used, and the accuracy may decline significantly with the addition of features [35]. Therefore, the VS approach does not necessarily result in the optimal performance for multifeature classification. Furthermore, although a feature selection algorithm is used as preprocessing for optimization of the hyperdimensional feature space, it is difficult to determine the appropriate number of the feature dimensionality.

In this context, we propose an SVM-based multiclassifier system combining a series of spectral and spatial features for high-resolution image classification. It is able to take advantage of multiple features and overcome the Hughes effect [36] and the over-fitting problem produced by the hyperdimensional

stacked feature space. The multiclassifier system has been applied to classification of hyperspectral images [3], [37], multisource data [38], and high-resolution urban images [39]. However, few studies use the multiclassifier system to simultaneously integrate the spectral, structural and semantic features at both pixel and object levels. The contribution of this study lies in a systematic combination of the spectral-spatial multifeatures coupled with a series of SVM classifiers, as described in the following three algorithms.

- Algorithm 1 (certainty voting): According to the decision results of the single-feature SVMs, the pixels in an image are separated into reliable and unreliable ones. The labels of reliable pixels are identified by majority voting of the SVMs, while the classification of unreliable pixels is performed by comparing the classification certainty degree [40] of the single-feature SVMs. This algorithm is written as C-voting in the following text.
- Algorithm 2 (probabilistic fusion): The certainty degree of each single-feature SVM is used as the weight of the probabilistic output of the SVM. Subsequently, the weighted probabilistic outputs of the SVMs are fused for the final classification. This algorithm is called P-fusion the following text.
- Algorithm 3 (object-based semantic approach): After segmentation of an image, image objects are divided into reliable and unreliable ones. The reliable objects are classified using the weighted probabilistic outputs of pixels that constitute the object, while the unreliable ones are identified based on a series of semantic rules. This algorithm is written as OBSA in the following text.

The C-voting and P-fusion algorithms can be implemented at both pixel and object levels, while the OBSA is only at the object level since the semantic rules are generated based on objects. The algorithms are tested on three multispectral data sets with high spatial resolution: the HYDICE DC Mall data set and the WorldView-2 Hangzhou and Hainan images. The spatial features chosen for construction of the multifeature SVM ensemble include the GLCM [41], DMPs [6], [12], and urban complexity index (UCI) [42], which are described in Section II. The proposed multifeature SVM ensemble is introduced in Section III. The experimental data sets are described in Section IV, followed by the experimental results in Section V and the discussion in Section VI. The last section concludes the paper.

II. Spectral-Spatial Multifeature Extraction

A. Spectral Feature Extraction

In this paper, PC analysis [18], [43] is used for spectral feature extraction from multi/hyperspectral images, considering that it is simple and fast to implement. Furthermore, information contained in the multi/hyperspectral images can be represented by several spectral PCs. Although other spectral feature extraction techniques can also be employed, such as non-negative matrix factorization [44], independent component analysis [45], and decision boundary feature extraction [18], the discussion about different feature extraction algorithms is beyond the scope of this study. In our experiments, four and three PCs are extracted from the HYDICE and WorldView-2 data sets, respectively, containing over 99% of the information of the images.

B. Gray-Level Co-Occurrence Matrix (GLCM)

A GLCM is employed in this study considering that it is a standard technique for texture extraction [41] and has proved to be effective in enhancing the classification of high-resolution images [33], [46]. The texture function of GLCM can be expressed as $f_{\mathrm{GLCM}}(b, m, w, d)$, which contains several parameters: base image b, window size w, texture measure m, and direction d. These parameters in this study are defined as follows.

1) Base image: The texture measures are extracted from the PCs of the multi/hyperspectral images.
2) Texture measures: Contrast (CON), representing the gray-level difference between neighboring pixels, is used in this study. It is one of the most efficient measures for the discrimination between built-up and non-built-up areas [47]. Furthermore, multidirectional contrast has the potential to discriminate between roads and buildings [33]. Contrast is calculated by

$$\mathrm{CON} = \sum_{i} \sum_{j} (i-j)^2 \cdot P(i.j) \tag{1}$$

where $P(i, j)$ indicates the joint probability of occurrence of the pairs of gray levels i and j separated by a given distance and direction within the moving window. It should be underlined that other measures, such as homogeneity, entropy, and dissimilarity, can be also considered for the GLCM texture extraction. However, in this study, the measure of contrast obtained the best results, and it was shown that adding other measures in the texture vector did not give a better performance than using the contrast measure individually.
3) Window sizes: A single window size is not reasonable because the high-resolution image always shows multiscale characteristics. Therefore, in this study, the GLCM textures are computed using a series of analysis windows (see Table II).
4) Directions: Most of the existing studies averaged the textural measures derived from different directions in order to obtain a rotation-invariant measure, which was criticized by Pesaresi *et al.* [33]. In this paper, different directions (see Table II) are considered because the directionality of texture has the potential to discriminate between isotropic and anisotropic structures.

C. Differential Morphological Profiles (DMPs)

MPs [6], [7] perform a series of morphological openings and closings with a family of structuring elements (SEs) of increasing size. The opening and closing are basic morphological operators, used to remove small bright (opening) or dark (closing) details while leaving the overall features relatively undisturbed. These operators are applied to a gray image with

TABLE II
PARAMETERS OF SPATIAL FEATURES (GLCM, DMP, AND UCI).

Datasets	Features	Parameters	No. of Dimension
HYDICE (DC Mall)	GLCM	Base image: b=(PC1, PC2, PC3, PC4); Measure: m=Contrast; Window: w=(3, 7, 11); Direction: d=(45°, 90°, 135°, 180°).	48
	DMP	Base image: b=(PC1, PC2, PC3, PC4); SE: λ=(2, 4, 6, 8, 10); Morphological operators: opening and closing by reconstruction.	40
	UCI	Decomposition level: l=1; Window: w=(4, 8, 16).	3
WorldView-2 (Hangzhou and Hainan)	GLCM	Base image: b=(PC1, PC2, PC3); Measure: m=Contrast; Window: w=(5, 9); Direction: d=(45°, 90°, 135°, 180°).	24
	DMP	Base image: b=(PC1, PC2, PC3); SE: λ=(3, 5, 7, 9); Morphological operators: opening and closing by reconstruction.	24
	UCI	Decomposition level: l=1; Window: w=(4, 8, 16).	3

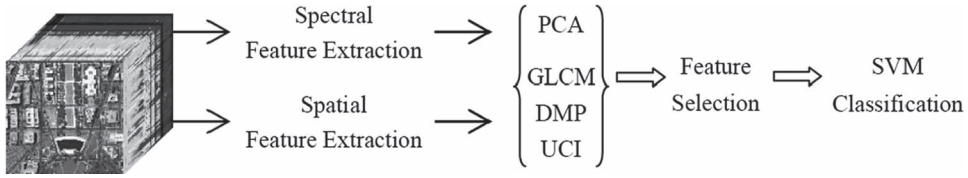

Fig. 1. Flowchart of the vector stacking multifeature fusion.

a series of SE. In addition, morphological operators are usually implemented using a reconstruction filter because this family of filters have better shape preservation and introduce less shape noise than the classical morphological filters [6].

Let $\gamma^{\mathrm{SE}}(I)$ and $\phi^{\mathrm{SE}}(I)$ be the morphological opening and closing by reconstruction with SE for an image I. MPs are defined using a series of SE with increasing sizes

$$\mathrm{MP}_\gamma = \left\{\mathrm{MP}_\gamma^\lambda(I) = \gamma^\lambda(I), \forall\lambda \in [0,n]\right\} \quad (2)$$

$$\mathrm{MP}_\phi = \left\{\mathrm{MP}_\phi^\lambda(I) = \phi^\lambda(I), \forall\lambda \in [0,n]\right\}$$

$$\text{with} \quad \gamma^0(I) = \phi^0(I) = I \quad (3)$$

where λ represents the radius of the disk-shaped SE. Subsequently, DMPs are defined as vectors where the measures of the slopes of the MPs are stored for every step of an increasing SE series

$$\mathrm{DMP}_\gamma = \left\{\mathrm{DMP}_\gamma^\lambda(I) = \left|\mathrm{MP}_\gamma^\lambda(I) - \mathrm{MP}_\gamma^{\lambda-1}(I)\right|,\ \lambda \in [1,n]\right\} \quad (4)$$

$$\mathrm{DMP}_\phi = \left\{\mathrm{DMP}_\phi^\lambda(I) = \left|\mathrm{MP}_\phi^\lambda(I) - \mathrm{MP}_\phi^{\lambda-1}(I)\right|,\ \lambda \in [1,n]\right\}. \quad (5)$$

In the experiments, DMP_γ and DMP_ϕ are always concatenated into a DMP vector in order to represent both bright and dark features in an image: $\mathrm{DMP} = \{\mathrm{DMP}_\gamma, \mathrm{DMP}_\phi\}$. The key parameters of the DMP include the base images and the radius of the disk-shaped SE. The parameter values for different data sets are listed in Table II.

D. Urban Complexity Index (UCI)

Most of the existing textural and structural features focus on the spatial domain alone, but few algorithms refer to feature extraction from the joint spectral-spatial domains. The recently developed UCI [42] based on 3-D wavelet transform (3-D-WT) processes a multi/hyperspectral image as a cube, and it is able to simultaneously describe the variation information in the joint spectral-spatial feature space. A 3-D-WT decomposes an image cube I by a tensor product

$$I^{(x,y,z)} = (L^x \oplus H^x) \otimes (L^y \oplus H^y) \otimes (L^z \oplus H^z) = \begin{cases} L^xL^yL^z \oplus L^xL^yH^z \oplus L^xH^yL^z \\ \oplus L^xH^yH^z \oplus H^xL^yL^z \oplus H^xL^yH^z \\ \oplus H^xH^yL^z \oplus H^xH^yH^z \end{cases} \quad (6)$$

where $\oplus$ and $\otimes$ denote the space direct sum and tensor product, respectively. L and H represent the low- and high-pass filters along the x, y, and z axis, respectively. The x and y directions stand for the spatial coordinates of an image, and z is the spectral axis. One-level 3-D-WT decomposes an image cube into eight subbands, which can be separated into three categories:

- Approximation: LLL
- Subbands of spectral variation $(^{**}H)$: LLH, LHH, HLH, HHH
- Subbands of spatial variation $(^{**}L)$: LHL, HLL, HHL

where the $^{**}H$ components represent spectral variation, since the high-pass filter is used along the spectral direction, while the $^{**}L$ components stand for spatial variation as they represent high-frequency information in the spatial domain and

low-frequency information in the spectral domain. Yoo *et al.* [42] proposed a UCI based on energy parameters of 3-D wavelet subbands. It was shown that the UCI was able to discriminate between complex urban and natural classes. The UCI is defined as the sum of all L components $(^{**}L)$ divided by the sum of all H components $(^{**}H)$

$$\text{UCI} = \frac{E(\text{HLL}) + E(\text{LHL}) + E(\text{HHL})}{E(\text{LLH}) + E(\text{LHH}) + E(\text{HLH})}. \tag{7}$$

The function $E(f)$ denotes energy of the subband f

$$E(f) = \sum_i \sum_j \sum_m \left(f(i,j,m)\right)^2 \tag{8}$$

where i, j, m stand for the coordinates of x, y, z directions in a subband, respectively. The basic idea of UCI is that natural features (e.g., water, forest, grass, and soil) have relatively smaller spatial changes than spectral changes, while urban areas (e.g., buildings, roads) have more variability in the spatial domain than the spectral domain. Consequently, according to (7), urban structures have larger UCI values than natural structures since the former contain more spatial variation information. A previous study showed that the first decomposition level ($l = 1$ with l being the number of decomposition levels for wavelets) gave the highest classification accuracies due to the fact that the first level contained the majority of the energy of the wavelet coefficients [42]. Therefore, the parameters of the UCI only refer to the window size w. In this experiment, three window sizes are considered: $w = (4, 8, 16)$, according to the spatial resolution and the characteristics of the information classes in the images.

E. Multifeature Vector Stacking

In spite of the availability of multiple spatial features, it is difficult to determine which one is optimal for a specific image scene. Furthermore, a combination of multiple features may yield better classification performance than their individual use. Therefore, researchers have proposed to integrate multiple features for image interpretation. Traditionally, the most widely used multifeature fusion approach is to concatenate multiple features into one vector and then interpret the vector via a classifier, e.g., SVMs. Before introduction of the new multifeature fusion methods, the traditional VS-SVM algorithm is carried out in this subsection. Considering that the performance of the SVM classifier may be sensitive to the dimension of feature space [31], [35], a SVM-specific feature selection method, SVM recursive feature elimination (SVM-RFE) [20] is used for optimization for VS-SVM classification. The SVM-RFE utilizes the objective function as a feature-ranking criterion to produce a list of features ordered by discrimination ability. The flowchart is shown in Fig. 1. Parameters of the three kinds of spatial features (GLCM, DMP, and UCI) are listed in Table II.

Fig. 2 shows the relationship between accuracies of the VS-SVM fusion and the dimensionality of the multifeature space. In this experiment, the highest accuracies among the five classification results generated by different starting training samples are used to delineate the curves. In the figure, the dotted lines, dashed lines, and dash-dot lines denote the accuracies of the single-feature classification for DMP, UCI, and GLCM, respectively. By analyzing the figure, we can obtain the following observations.

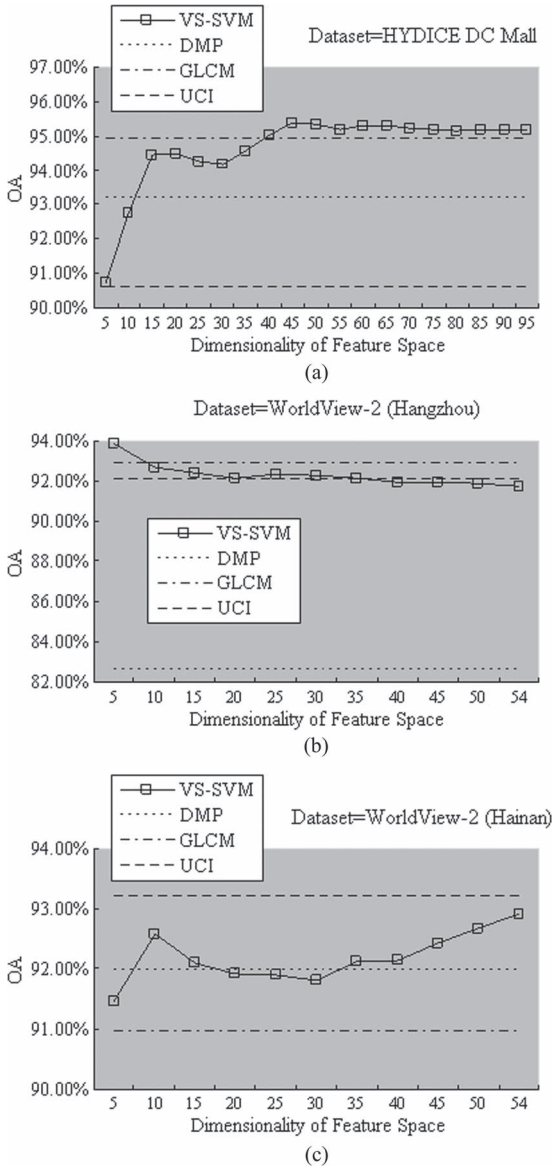

Fig. 2. Relationship between classification accuracy (overall accuracy) and the dimensionality of feature space for the VS-SVM multifeature fusion (the SVM-RFE is used for feature selection).

1) It is difficult to find a single feature that is optimal for different image scenes. For instance, although the GLCM measure gives the highest accuracies for the HYDICE DC Mall and the WorldView-2 Hangzhou data sets, it yields the worst performance for the WorldView-2 Hainan data set.
2) Compared to the highest accuracy achieved by the single-feature classification (i.e., the DMP, UCI, or GLCM feature is individually interpreted by a SVM classifier), the VS fusion provides slightly higher accuracy in the HYDICE DC Mall and the WorldView-2 Hangzhou data sets [Fig. 2(a) and (b)], but lower accuracy in the WorldView-2 Hainan data set [Fig. 2(c)].

3) By observing the trend of the accuracy curves, it is difficult to adaptively determine the optimal dimension of the hybrid multifeature space.

Based on the above analysis, it can be stated that although the VS-SVM is effective in combining multiple features, it is worthwhile to study other fusion methods in order to take advantage of different features and improve the classification results.

III. Multifeature SVM Ensemble

Before introduction of the proposed multifeature SVM ensemble algorithms, two concepts about the SVM are defined.

1) SVM probabilistic output: The output of a SVM can be a classification map that contains class labels for each pixel (or object), or a probability map that contains probability estimates for each pixel (or object) to belong to the assigned class. In this paper, the one-versus-all approach [48] is used for the multiclass SVM soft output. The one-versus-all approach builds K SVMs (K is the number of information classes), each of which is able to separate one class from all the others. For each pixel x, the SVM answers with a decision value $d_k(x)$ that indicates the distance between the pixel x and the separating hyperplane of class k. $p_k(x)$, the probability of pixel x belonging to class $k(k = 1, \ldots, K)$, is calculated by transforming the SVM decision value $d_k(x)$ based on a sigmoid function [49]

$$p_k(x) = \frac{1}{1 + \exp\left(A_k \cdot d_k(x) + B_k\right)} \tag{9}$$

where A_k and B_k are estimated for the SVM of class k by minimizing the mean square error on the training data between the original label and the output of the sigmoid function.

2) Certainty of SVM classification: The multiclass SVM probabilistic outputs $(p_1(x), \ldots, p_k(x), \ldots, p_K(x)$, $k = 1, 2, \ldots, K,)$ are able to reflect the classification certainty for each SVM. In this paper, the specificity measure [40] is used to calculate the certainty of SVM classification

$$S(x) = \sum_{k=1}^{K-1} \left[\hat{p}_k(x) - \hat{p}_{k+1}(x)\right] \cdot \frac{1}{k} \tag{10}$$

where $\hat{p}_1(x), \ldots, \hat{p}_k(x), \ldots, \hat{p}_K(x)$ represent the multiclass probabilistic outputs in a descending order. A larger value of $S(x)$ signifies that the SVM classification for pixel x is more reliable.

Based on the aforementioned concepts, the proposed three multifeature fusion algorithms are described as follows.

Algorithm 1: C-voting

Step 1: Single-feature SVM classification. The spectral PCs concatenated with a kind of spatial feature (e.g., GLCM, DMP, or UCI) are fed into a SVM for classification, resulting in the crisp (class label) and soft (probabilistic) outputs for each single-feature SVM.

Step 2: Pixel-based C-voting. Results of the multiple SVMs are utilized to determine the reliable (x_{r}) and unreliable pixels (x_{un}). The reliable pixels are defined as the ones that all the single-feature SVMs give the same label, and the other pixels are defined as unreliable. The reliable pixels are classified by

$$C(x_{\mathrm{r}}) = \arg\max_{k=\{1,\ldots,K\}} V_x(k)$$
$$\text{where} \quad V_x(k) = \sum_{f=1}^{F} I\left(C(x_f) = k\right) \tag{11}$$

where I is the indicator function, $C(x_{\mathrm{r}})$ is the class label of the reliable pixel x, $C(x_f)$ is the class label of the fth SVM, $V_x(k)$ is the number of votes that pixel x receives for the class k, and F is the number of kinds of spatial features, i.e., the number of SVMs. The unreliable pixels are classified according to the certainty measure

$$C(x_{\mathrm{un}}) = C(x_{\hat{f}}) \quad \text{with} \quad \hat{f} = \arg\max_{f=\{1,\ldots,F\}} S_f(x) \tag{12}$$

where $C(x_{\mathrm{un}})$ is the class label of the unreliable pixel x, $S_f(x)$ is the classification certainty for pixel x with the feature f, and $\hat{f}$ is the optimal feature that has the largest specificity measure among all the F SVMs. In (12), the certainty measure is used to resolve the classification conflict between the multiple SVM classifiers. In this way, the multiple spectral-spatial SVMs are fused by minimizing the classification uncertainty.

Step 3: Object-based C-voting. In order to take advantage of the spatial smoothness of segmentation and hence reduce the salt and pepper effect of the pixel-based processing, the pixel-based C-voting algorithm is extended to its object-based version via segmentation-based majority voting

$$C(obj) = \arg\max_{k=\{1,\ldots,K\}} V_{obj}(k)$$
$$\text{where} \quad V_{obj}(k) = \sum_{x \in obj} I\left(C(x) = k\right) \tag{13}$$

where $C(obj)$ is the class label of the object *obj*, and $V_{obj}(k)$ is the number of times class k is detected within the obj. In this paper, an adaptive mean-shift procedure [15] is used to generate objects from an image. Mean shift is an efficient spatial feature extraction approach that is capable of delineating arbitrarily shaped clusters due to its nonparametric nature [50]. The processing flow of the C-voting algorithm is presented in Fig. 3(a).

Algorithm 2: Probabilistic fusion (P-fusion)

Step 1. Single-feature SVM classification.

Step 2. Pixel-based P-fusion. The soft outputs of the multiple spectral-spatial SVMs are integrated, and the

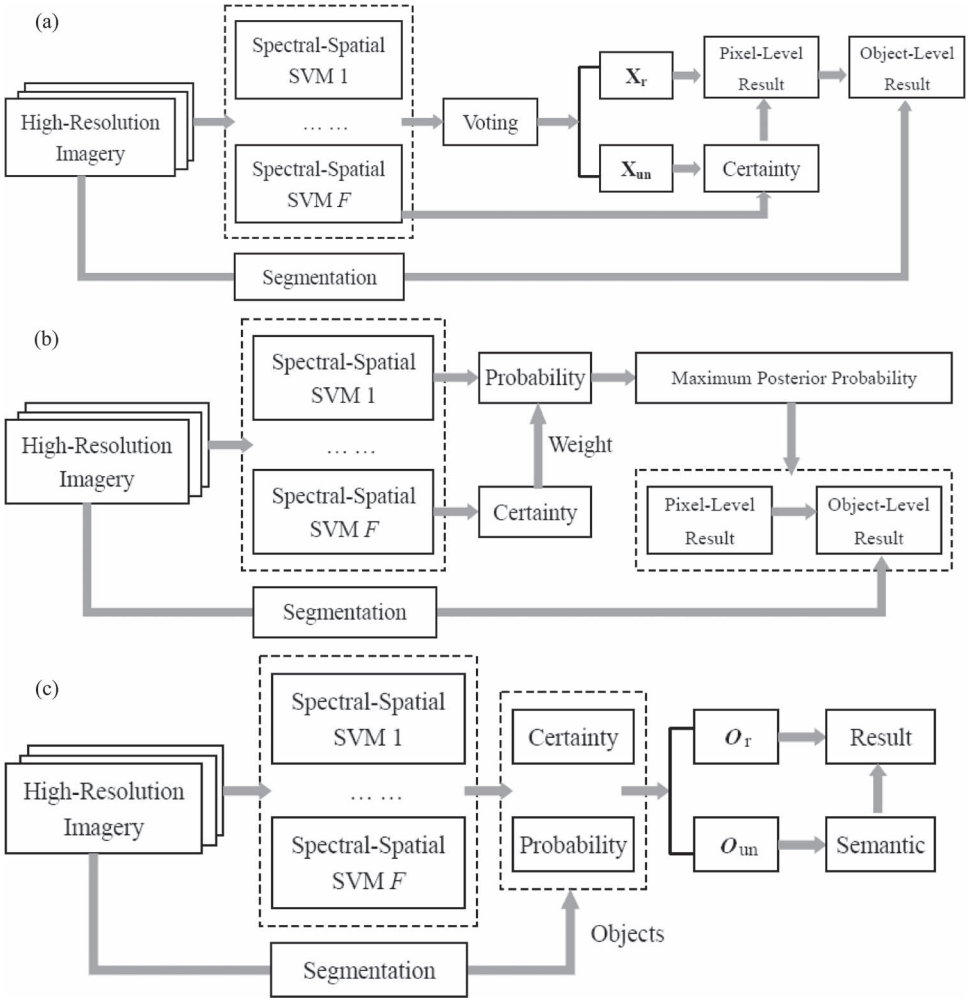

Fig. 3. Processing chains of C-voting (a), P-fusion (b), and OBSA (c) for the multifeature SVM ensemble.

final result is determined by the maximum posterior probability

$$C(x) = \arg\max_{k=\{1,\ldots,K\}} \left\{ \frac{1}{F} \sum_{f=1}^{F} S_f(x) \cdot p_f^k(x) \right\} \quad (14)$$

where $p_f^k(x)$ represents the probabilistic value of pixel x for the class k with the feature f. The specificity measure is used as the weight of the probabilistic value in order to reduce the influence of unreliable information and enhance the relative weight of reliable information.

Step 3. Object-based P-fusion. The pixel-based P-fusion is extended to the object-based result by majority voting. An adaptive mean shift is also used for the segmentation. The P-fusion algorithm is shown in Fig. 3(b).

Algorithm 3: OBSA

Step 1. Single-feature SVM classification.

Step 2. Adaptive mean-shift segmentation.

Step 3. Object-based probabilistic outputs. The probabilistic outputs for each object O are calculated by averaging the probabilistic values of all the pixels within the object

$$p^k(O) = \frac{\sum_{x \in obj} \sum_{f=1}^{F} S_f(x) \cdot p_f^k(x)}{N \times F} \quad (15)$$

where $p^k(O)$ is the probabilistic output of object O for class k, and N is the number of pixels in the object.

Step 4. Reliable and unreliable objects. The objects are divided into reliable (O_{r}) and unreliable (O_{un}) ones, according to the weighted probabilistic outputs

$$O_{\mathrm{r}} : p_{\max}(O) \geq T, \quad \text{and} \quad O_{\mathrm{un}} : p_{\max}(O) < T,$$
$$\text{with} \quad p_{\max}(O) = \max\left\{p^k(O), \quad k = \{1, \ldots, K\}\right\}$$

where $p_{\max}(O)$ is the maximum probabilistic output for an object O. It can also be viewed as the probabilistic value of the winning label of the object. An object is defined as a reliable one when its probabilistic value of the winning label is larger than a threshold T. The threshold is used to control the proportion of objects on which the semantic rules are imposed. It should be noted that a blind application

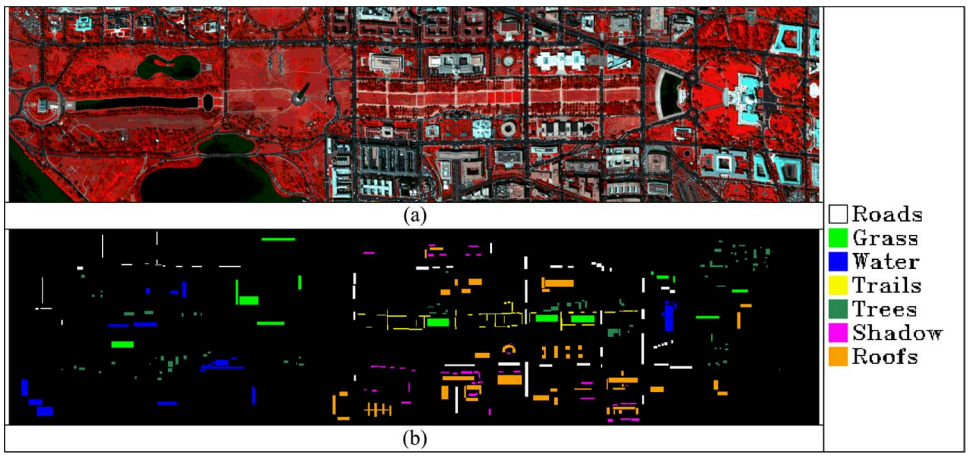

Fig. 4. Test image of HYDICE DC Mall: (a) with bands 60, 27, and 17 for red, green, and blue colors, respectively, and the ground truth reference (b).

of the semantic processing to all objects in the image could decrease the overall accuracy due to the inaccuracy of the segmentation. An appropriate range of T is between 0.1 and 0.5. A small value signifies that a small fraction of the objects are defined as unreliable and chosen for the semantic postprocessing. In this paper, T is set to 0.3.

Step 5. Classification of reliable and unreliable objects. The reliable objects are classified by the maximum posterior probability

$$C(O_{\mathrm{r}}) = \arg\max_{k=\{1,\ldots,K\}} p^k(O_{\mathrm{r}}) \tag{16}$$

while the unreliable objects are classified by the following semantic rules:

Rule 1) An unreliable object of roof is reclassified as a road or soil when the following conditions are satisfied:
- O is an unreliable object of roof: $p_{\max}(O) < T$, and $C(O) =$ roof;
- The relative border of the object to road or soil is larger than $T_{\mathrm{B}}(T_{\mathrm{B}} = 10\%)$;
- The distance between the object to its nearest shadow object is larger than zero (roofs are always adjacent to shadows).

Rule 2) An unreliable object of road or soil is reclassified as a roof when the following conditions are satisfied:
- O is an unreliable object of road or soil: $p_{\max}(O) < T$, and $C(O) =$ road or soil;
- The relative border of the object to roof is larger than T_{B};
- The distance between the object to its nearest shadow object is equal to zero since roofs are always adjacent to shadows.

Rule 3) An unreliable object of water is assigned to shadow when the following conditions are met:
- O is an unreliable object of water: $p_{\max}(O) < T$, and $C(O) =$ water;
- The distance between the object to its nearest shadow object is equal to zero.

TABLE III
NUMBERS OF THE TRAINING AND TEST SAMPLES (HYDICE DC MALL)

Information Classes	No. of Training Samples	No. of Test Samples
Roads	100	3,334
Grass	100	3,075
Water	100	2,882
Trails	100	1,034
Trees	100	2,047
Shadow	100	1,093
Roofs	100	5,867
Total	700	19,332

Rule 4) An unreliable object of shadow is assigned to water when the following conditions are met:
- O is an unreliable object of shadow: $p_{\max}(O) < T$, and $C(O) =$ shadow;
- The distance between the object to its nearest water object is equal to zero.

The processing chain of the OBSA is presented in Fig. 3(c). The semantic rules focus on the unreliable objects that cannot be correctly identified by the feature extraction and classification techniques. Rules 1 and 2 are used to resolve the misclassification between roads, roofs, and soil, while rules 3 and 4 are related to water and shadow.

IV. STUDY AREAS AND DATA SETS

The well-known HYDICE airborne hyperspectral data set from the Washington DC Mall (191 bands with 3.0-m spatial resolution) is used for evaluation of algorithms, considering that it is a standard test image for the classification of urban hyperspectral data. This image contains 1280 scan lines with 307 pixels in each scan line. The test image and its reference are shown in Fig. 4. The challenges for classifying this image are: 1) discrimination between roads, trails, and roofs, since they are made of similar materials and have similar spectral properties; 2) discrimination between water and shadow, as they have very similar spectral reflectance. The training and test samples (Table III) are chosen according to the ground reference provided in [18]. It should be underlined that the training samples

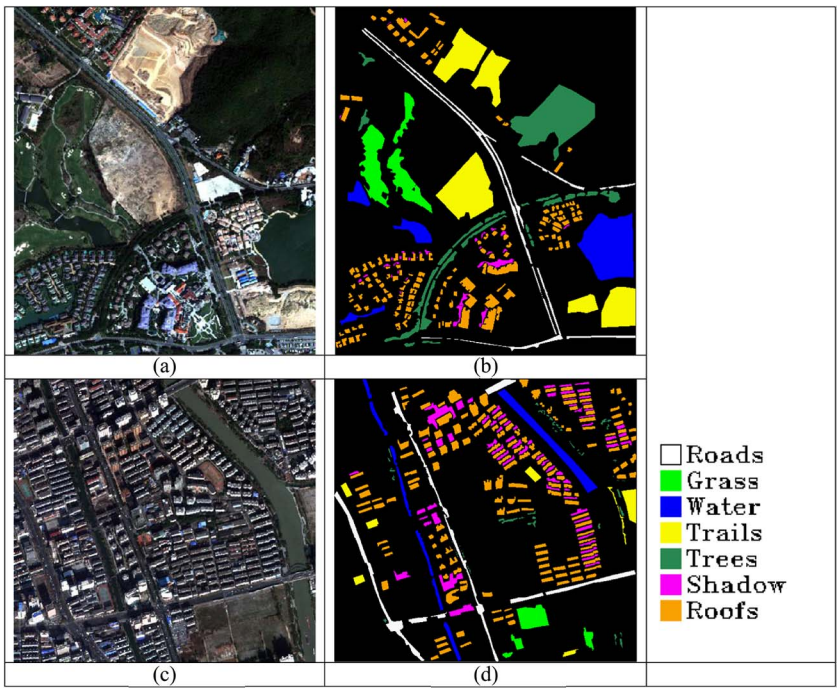

Fig. 5. WorldView-2 test data sets: (a) and (b) are the test image and ground reference of the Hainan data set (rural area), and (c) and (d) are the test image and ground reference of the Hangzhou data set (dense urban), respectively. The test images are displayed with a composite of red, green, and blue bands.

TABLE IV
NUMBERS OF THE TRAINING AND TEST SAMPLES (WORLDVIEW-2 DATA SETS)

Datasets	WorldView-2 Hainan		WorldView-2 Hangzhou	
Information Classes	No. of Training Samples	No. of Test Samples	No. of Training Samples	No. of Test Samples
Roofs	50	11,148	50	23,685
Roads	50	5,100	50	8,800
Soil	50	18,319	50	3,229
Grass	50	7,417	50	3,359
Shadow	50	1,427	50	9,486
Trees	50	14,086	50	1,228
Water	50	11,209	50	7,237
Total	350	68,706	350	57,024

(100 pixels for each class) are randomly selected from the training sets (300 pixels for each class), and all the experiments are repeated five times with different starting training sets. In this way, the mean and standard deviation of the results for the five runs are reported in the experiments.

The WorldView-2 imagery is of interest as it is a new-generation satellite image that is able to provide rich spectral and spatial information at the same time (8 bands with 2.0-m spatial resolution). The WorldView-2 images are shown in Fig. 5, where (a) and (b) are the test image and the ground truth reference of the Hainan data set (rural area with 520 $\times$ 600 pixels), and (c) and (d) are the test and reference images of the Hangzhou data set (dense urban with 606 $\times$ 567 pixels), respectively. The ground truth reference images are generated by field campaign and visual interpretation of the study areas. The numbers of training and test samples for the two WorldView-2 data sets are shown in Table IV. The training samples (50 pixels for each class) are randomly selected from the training sets (300 pixels for each class). All the experiments are repeated five times with different starting training sets. The mean and standard deviation of the results are reported for the assessment of the classification algorithms. The challenge for the classification of the WorldView-2 data sets is to distinguish the spectrally similar classes such as soil-roads-roofs, trees-grass, and water-shadow.

V. RESULTS

The class-specific accuracies of the HYDICE DC Mall, WordView-2 Hainan, and Hangzhou data sets with different classification algorithms are presented in Tables V–VII, respectively. The results are reported based on five runs with different starting training sets, and the mean and standard deviation of the accuracies are presented in the tables.

The general comments regarding the results are summarized as follows:

1) Concerning the single-feature classification, the GLCM obtained the best results for DC Mall $(\text{OA} = 94.4\%)$ and WorldView-2 Hangzhou $(\text{OA} = 92.8\%)$ data sets, while the UCI gave the best result for the Hainan data set $(\text{OA} = 92.4\%)$.
2) For the multifeature classification at the pixel level, the P-fusion algorithm slightly outperformed the C-voting in all the three experiments. In addition, the proposed C-voting and P-fusion algorithms outperformed the traditional VS-SVM for the two WorldView-2 data sets, but the VS-SVM gave higher accuracy in the DC Mall experiment.
3) At the object level, the OBSA algorithm gave the most accurate results in all the experiments. Compared to the object-based C-voting and P-fusion, the accuracy improvements achieved by the OBSA were 1.5%–5%

TABLE V
CLASS-SPECIFIC ACCURACIES AND OVERALL ACCURACIES (OA) (%) FOR DIFFERENT CLASSIFICATION ALGORITHMS (HYDICE DC MALL)

Classes	Spectral	Single-Feature			Multifeature (Pixel Level)			Multifeature (Object Level)			
		DMP	GLCM	UCI	VS-SVM	C-voting	P-fusion	VS-SVM	C-voting	P-fusion	OBSA
Roads	92.7 ± 0.1	95.8 ± 0.8	91.1 ± 0.3	91.8 ± 0.3	95.1 ± 0.4	91.9 ± 0.3	91.9 ± 0.3	97.2 ± 0.4	90.8 ± 0.1	90.8 ± 0.1	96.7 ± 0.0
Grass	96.7 ± 0.2	96.9 ± 0.8	98.9 ± 0.4	98.1 ± 0.2	95.9 ± 0.3	98.9 ± 0.2	98.9 ± 0.1	95.1 ± 1.0	98.9 ± 0.1	99.0 ± 0.1	99.0 ± 0.1
Water	85.5 ± 4.0	99.9 ± 0.0	96.9 ± 0.2	98.1 ± 0.4	99.4 ± 0.1	99.3 ± 0.2	99.2 ± 0.1	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
Trails	49.1 ± 2.5	77.8 ± 1.4	91.3 ± 2.5	61.9 ± 3.8	97.9 ± 0.4	76.5 ± 6.2	81.7 ± 5.2	98.3 ± 0.8	81.5 ± 8.9	84.6 ± 5.7	98.2 ± 0.5
Trees	97.1 ± 0.1	97.6 ± 0.6	98.2 ± 0.1	97.3 ± 0.2	97.1 ± 0.5	98.3 ± 0.1	98.3 ± 0.1	96.3 ± 1.3	98.2 ± 0.2	98.3 ± 0.2	98.2 ± 0.2
Shadow	73.8 ± 4.4	86.9 ± 1.7	91.1 ± 0.3	93.3 ± 1.4	84.1 ± 1.2	97.1 ± 0.4	96.9 ± 0.1	79.6 ± 0.5	96.9 ± 0.0	97.0 ± 0.0	97.0 ± 0.0
Roofs	75.5 ± 2.8	89.9 ± 0.3	92.5 ± 0.5	83.3 ± 2.0	92.6 ± 0.2	89.1 ± 2.1	90.9 ± 1.5	92.9 ± 0.1	90.5 ± 2.6	91.7 ± 1.5	98.9 ± 0.0
OA	82.7 ± 1.7	93.4 ± 0.1	94.4 ± 0.3	89.7 ± 0.9	94.8 ± 0.1	93.3 ± 1.1	94.1 ± 0.8	94.8 ± 0.3	93.9 ± 1.4	94.4 ± 0.8	98.5 ± 0.0

TABLE VI
CLASS-SPECIFIC ACCURACIES AND OVERALL ACCURACIES (OA) (%) FOR DIFFERENT CLASSIFICATION ALGORITHMS (WORLDVIEW-2 HAINAN)

Classes	Spectral	Single-Feature			Multifeature (Pixel Level)			Multifeature (Object Level)			
		DMP	GLCM	UCI	VS-SVM	C-voting	P-fusion	VS-SVM	C-voting	P-fusion	OBSA
Buildings	62.0 ± 2.0	81.2 ± 0.6	81.3 ± 1.3	85.3 ± 1.2	87.2 ± 0.8	86.3 ± 0.4	86.6 ± 0.3	87.4 ± 0.8	87.3 ± 1.1	87.5 ± 0.9	92.6 ± 1.2
Roads	71.0 ± 0.9	82.9 ± 2.6	80.8 ± 1.2	81.1 ± 1.7	84.6 ± 1.2	86.6 ± 1.6	87.3 ± 1.3	85.4 ± 1.3	90.5 ± 2.3	90.5 ± 2.2	91.8 ± 2.1
Grass	95.6 ± 1.6	89.9 ± 3.0	94.3 ± 2.9	96.8 ± 1.9	87.5 ± 2.8	96.5 ± 1.0	96.8 ± 1.2	86.5 ± 2.9	96.8 ± 1.1	97.0 ± 1.2	96.7 ± 1.4
Trees	96.7 ± 0.8	94.4 ± 0.8	95.7 ± 2.3	97.5 ± 0.2	93.4 ± 0.7	97.0 ± 0.7	97.1 ± 0.8	92.7 ± 0.7	96.6 ± 0.7	96.7 ± 0.7	96.5 ± 0.8
Soil	85.3 ± 1.9	94.3 ± 0.5	93.1 ± 0.7	95.3 ± 0.9	96.8 ± 0.2	95.0 ± 0.4	95.2 ±0.4	97.3 ± 0.3	95.2 ± 0.3	95.4 ± 0.4	98.2 ± 0.4
Water	99.1 ± 0.0	98.9 ± 1.3	96.3 ± 0.1	95.7 ± 0.8	98.9 ± 0.8	99.4 ± 0.5	99.1 ± 0.4	99.4 ± 1.1	99.9 ± 0.0	99.9 ± 0.0	99.9 ± 0.0
Shadow	83.2 ± 1.9	75.3 ± 3.1	63.5 ± 4.2	63.2 ± 2.8	70.1 ± 2.0	78.6 ± 3.3	76.8 ± 3.4	67.1 ± 2.4	76.1 ± 4.3	76.1 ± 3.7	76.2 ± 2.4
OA	86.0 ± 1.0	91.3 ± 0.4	90.5 ± 1.2	92.4 ± 0.7	92.4 ± 0.3	93.9 ± 0.3	94.0 ± 0.3	92.4 ± 0.2	94.4 ± 0.5	94.5 ± 0.5	96.0 ± 0.6

TABLE VII
CLASS-SPECIFIC ACCURACIES AND OVERALL ACCURACIES (OA) (%) FOR DIFFERENT CLASSIFICATION ALGORITHMS (WORLDVIEW-2 HANGZHOU)

Classes	Spectral	Single-Feature			Multifeature (Pixel Level)			Multifeature (Object Level)			
		EMP	GLCM	UCI	VS-SVM	C-voting	P-fusion	VS-SVM	C-voting	P-fusion	OBSA
Buildings	64.3 ± 7.6	82.0 ± 0.7	94.1 ± 0.3	92.2 ± 0.4	94.6 ± 0.2	93.7 ± 0.3	93.8 ± 0.2	95.7 ± 0.3	94.6 ± 0.2	94.5 ± 0.3	98.4 ± 0.3
Roads	79.3 ± 0.7	75.5 ± 0.7	85.6 ± 0.8	85.9 ± 0.5	81.9 ± 0.3	85.5 ± 0.2	85.9 ± 0.2	84.6 ± 0.4	89.8 ± 0.6	89.5 ± 0.7	96.9 ± 0.8
Grass	89.7 ± 1.5	88.9 ± 0.9	96.9 ± 1.4	94.0 ± 1.9	94.0 ± 0.5	95.3 ± 1.2	95.7 ± 1.3	94.4 ± 0.2	96.3 ± 1.7	96.7 ± 1.9	98.9 ± 0.1
Trees	93.0 ± 0.9	92.3 ± 1.2	88.2 ± 2.8	85.2 ± 3.3	90.2 ± 1.5	90.3 ± 1.2	90.8 ± 1.5	90.1 ± 1.3	93.5 ± 1.1	94.3 ± 1.4	97.3 ± 1.6
Soil	28.5 ± 11.0	50.8 ± 2.3	83.6 ± 1.3	80.7 ± 2.7	84.0 ± 0.3	81.1 ± 1.7	83.0 ± 1.8	88.1 ± 1.2	84.7 ± 3.0	86.9 ± 2.7	97.5 ± 1.3
Water	91.2 ± 0.9	95.8 ± 1.7	96.9 ± 0.4	95.5 ± 0.4	96.2 ± 1.3	98.2 ± 0.2	98.2 ± 0.0	97.7 ± 2.1	99.1 ± 0.0	99.2 ± 0.1	99.4 ± 0.2
Shadow	91.8 ± 0.2	90.5 ± 1.3	95.5 ± 0.4	96.4 ± 0.0	89.9 ± 1.3	96.0 ± 0.1	96.6 ± 0.1	90.6 ± 2.1	97.4 ± 0.1	97.7 ± 0.1	98.2 ± 0.3
OA	72.4 ± 4.1	82.5 ± 0.4	92.8 ± 0.3	91.6 ± 0.3	91.3 ± 0.3	92.7 ± 0.2	93.1 ± 0.1	92.7 ± 0.5	94.5 ± 0.1	94.6 ± 0.1	98.2 ± 0.4

due to the introduction of semantic rules. It should be mentioned that the object-based C-voting and P-fusion methods increased the accuracies by at most 2.0% compared to their pixel-based versions by courtesy of the spatial smoothness of segmentation.

VI. DISCUSSION

A. McNemar's Test

In order to evaluate the statistical significance in accuracy for the different classification algorithms, including the VS-SVM,

TABLE VIII
MCNEMAR'S TEST FOR THE THREE EXPERIMENTS (N = NO SIGNIFICANCE, S+ = POSITIVE SIGNIFICANCE, S− = NEGATIVE SIGNIFICANCE, SF = SINGLE-FEATURE). THE VS-SVM, C-VOTING, AND P-FUSION ALGORITHMS ARE IMPLEMENTED AT THE OBJECT LEVEL

	P-fusion	VS-SVM	OBSA	Optimal SF
C-voting	12N, 3S−	11S+, 1N, 3S−	15S−	11S+, 1N, 3S−
P-fusion		11S+, 2N, 2S−	15S−	11S+, 3N, 1S−
VS-SVM			15S−	4S+, 10N, 1S−
OBSA				15S+
Optimal SF				

C-voting, P-fusion, and OBSA, McNemar's test [51] is utilized in all three experiments. McNemar's test is based on the standardized normal test statistic

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{17}$$

where f_{12} indicates the number of samples classified correctly by classifier 1 and incorrectly by classifier 2. The difference in accuracy between classifiers 1 and 2 is viewed to be statistically significant if $|Z| > 1.96$ using 5% of significance. In our experiments, the following three cases are defined according to McNemar's test:

1) No significance between classifiers 1 and 2: $-1.96 \leq Z \leq 1.96$;
2) Positive significance: $Z > 1.96$;
3) Negative significance: $Z < -1.96$.

McNemar's test is presented in Table VIII, where five runs with different training samples of the three data sets generate 15 test results for each pair of classifiers. From the test, we can obtain the following conclusions regarding the proposed three multifeature SVM algorithms:

1) C-voting versus P-fusion (12N, 3S−): The P-fusion algorithm is slightly better than the C-voting algorithm, showing that probabilistic fusion is more appropriate for the integration of multiple spectral-spatial SVMs than crisp voting.
2) C-voting versus VS-SVM $(11\text{S}+, 1\text{N}, 3\text{S}-)$ and P-fusion versus VS-SVM $(11\text{S}+, 2\text{N}, 2\text{S}-)$: It is shown that the proposed C-voting and P-fusion algorithms are more effective for multifeature fusion than the traditional VS-SVM. It should be noted that the VS-SVM can also be viewed as effective for multifeature classification since it achieved comparable or better results than the optimal single-feature classification in most of the experiments (4S+ and 10 N).
3) OBSA: The object-based semantic approach provided significantly higher accuracies than the other algorithms. It is shown that although sophisticated classification techniques (such as multifeature C-voting and P-fusion) are able to yield satisfactory results for high-resolution image classification, the introduction of semantic rules can further improve the overall accuracies by 2–4% compared to other object-based multiple classifier systems, i.e., the object-based VS-SVM, C-voting, and P-fusion methods.

B. Comparison With the Multikernel SVM

A multikernel SVM classification was also implemented in this study for comparison. The multikernel learning (MKL) has been proved to be effective for classification of multisource data than the single or simple kernel classifier [21], [22], [52]. In this paper, the multikernel SVM proposed by Tuia *et al.* [22] was used to classify the multiple spectral-spatial features. Specifically in this experiment, each group of features corresponds to 4 RBF kernels with a series of bandwidth values [0.1, 0.25, 0.35, 0.5]. As a result, the composite kernel is built on 16 kernels for the four kinds of features including spectral, DMP, GLCM, and UCI. Readers can refer to [22] for details about the multikernel optimization and classification. The results are compared to the C-voting, P-fusion, and VS-SVM (Table IX). It can be seen that the MKL-SVM gave higher accuracies than the simple VS-SVM in the WorldView-2 experiments but a slightly lower accuracy in the HYDICE data set. It is also found that the C-voting and P-fusion achieved comparable or better results than the MKL-SVM in the three data sets in terms of accuracies.

C. Evaluation of the Knowledge-Based Rules

In order to evaluate the effectiveness of the knowledge-based rules, we compare the models with and without rule-based postprocessing. The model without post-processing steps uses the maximum probability to classify both reliable and unreliable objects

$$C(O) = \arg\max_{k=\{1,\ldots,K\}} p^k(O) \tag{18}$$

where O represents an arbitrary object in an image. The overall accuracies are compared in Table X.

From the table, it can be seen that rule-based postprocessing increased the overall accuracies by 3.9%, 3.1%, and 1.3% for the HYDICE DC Mall, WorldView-2 Hangzhou, and Hainan data sets, respectively. It is shown that the semantic approach is appropriate for the interpretation of unreliable objects, and it can be used as a postprocessing of the multifeature classification system.

D. Visual Inspection

For a visual comparison, a series of subset images extracted from the HYDICE DC Mall data set are shown in Fig. 6. Comments on the figure are summarized as follows:

1) Spectral classification: Misclassifications between the spectrally similar classes such as roads, roofs, and trails are obvious, and the salt and pepper effects can be clearly observed.
2) Single-feature classification: UCI and DMP refer to errors that some buildings are wrongly identified as trails, while the window size effects can be seen from the GLCM classification (edges of the water are wrongly classified as shadow) due to the moving window of the texture calculation.
3) VS-SVM (object level): The multifeature stacking approach is potential to discriminate spectrally similar

TABLE IX
COMPARISON BETWEEN THE MKL-SVM, VS-SVM, C-VOTING, AND P-FUSION

Methods	HYDICE DC Mall	WorldView-2 Hangzhou	WorldView-2 Hainan
VS-SVM	94.8 ± 0.1	91.3 ± 0.3	92.4 ± 0.3
MKL-SVM	94.2 ± 0.8	92.3 ± 0.3	93.1 ± 0.7
C-voting	93.3 ± 1.1	92.7 ± 0.2	93.9 ± 0.3
P-fusion	94.1 ± 0.8	93.1 ± 0.1	94.0 ± 0.3

TABLE X
ACCURACY COMPARISON BETWEEN THE OBJECT-BASED MULTIFEATURE MODELS WITH AND WITHOUT RULE-BASED POSTPROCESSING

	HYDICE DC Mall	WorldView-2 Hangzhou	WorldView-2 Hainan
Without post-processing	94.6 ± 0.9	95.1 ± 0.3	94.7 ± 0.3
With post-processing	98.5 ± 0.0	98.2 ± 0.4	96.0 ± 0.6

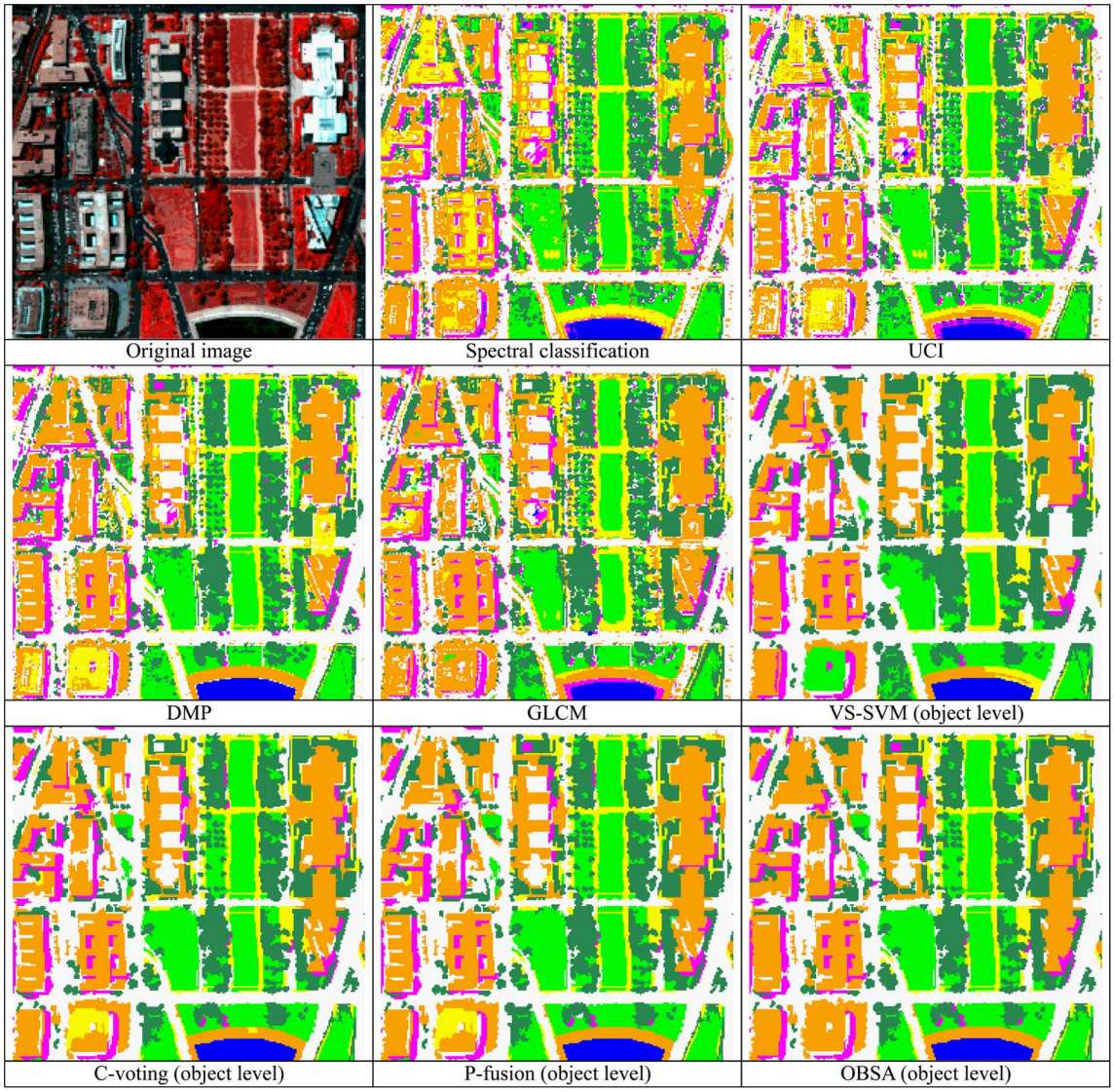

Fig. 6. Subset image classification maps for the HYDICE DC Mall data set (white = roads, orange = roofs, yellow = trails, light green = grass, sea green = trees, magenta = shadow).

classes that are not well separated by the single feature. However, the hyperdimensional and hybrid feature space may result in classification uncertainties. For instance, the building in the bottom-left corner is wrongly labeled as grass.

4) C-voting and P-fusion (object level): The object-based C-voting and P-fusion algorithms produce similar results. The performance of the multiple SVMs ensemble is influenced by the single-feature SVMs. When most of the SVMs give wrong classifications, the ensemble system often leads to wrong results. For instance, the building in the bottom-left corner is incorrectly classified as trails for C-voting and P-fusion since both DMP and UCI wrongly identify it as trails.

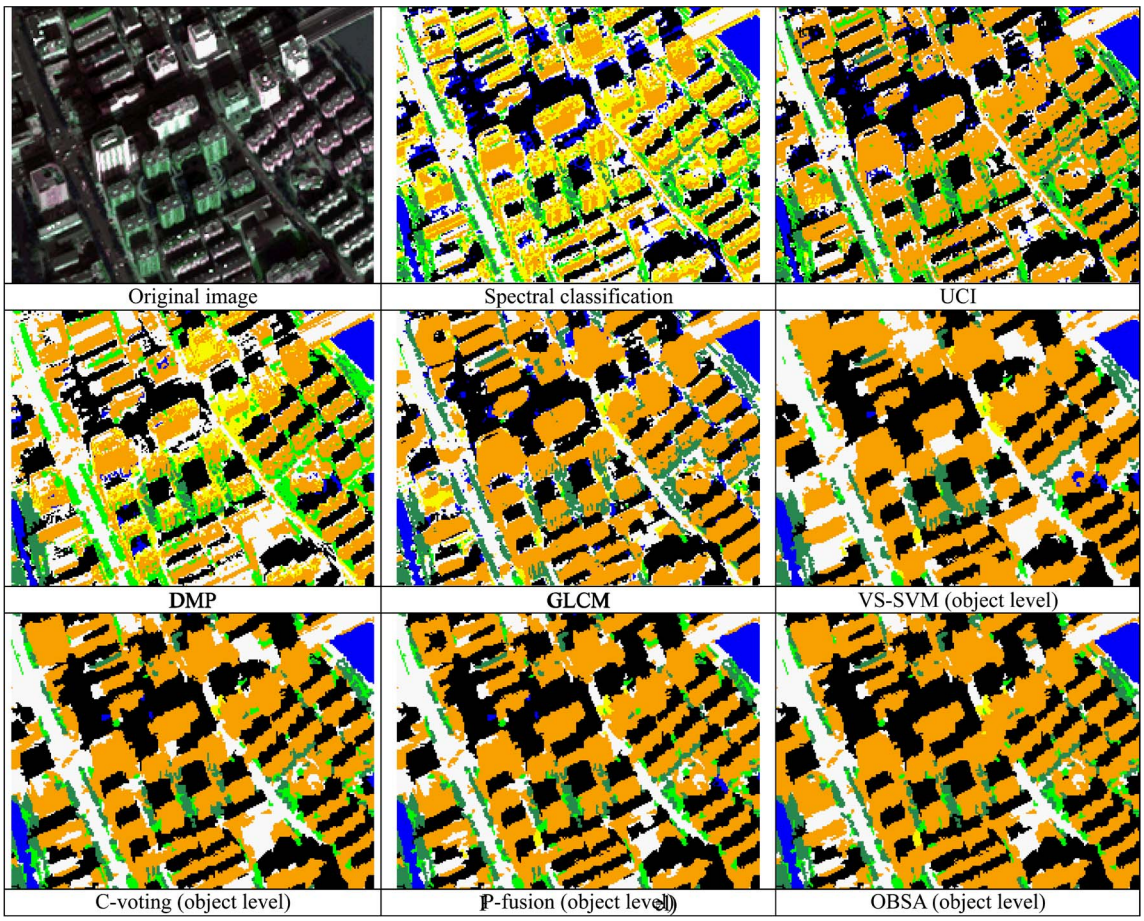

Fig. 7. Subset image classification maps for the WorldView-2 Hangzhou data set (white = roads, orange = roofs, yellow = soil, light green = grass, sea green = trees, black = shadow).

5) OBSA (object level): It can be seen that the misclassifications between roofs, roads and trails are significantly reduced due to the consideration of semantic rules.

Classification maps for a subset image of the WorldView-2 Hangzhou data set are compared in Fig. 7. It can be seen that the OBSA algorithm is able to correctly discriminate between roads, buildings, and soil, but the other algorithms cannot. Furthermore, the other algorithms fail to separate water and shadows, but the misclassifications can be effectively reduced by the semantic rules.

The classification maps for a subset image of the World View-2 Hainan data set are shown in Fig. 8. The challenges for classifying this image lie in the discrimination between soil and buildings. It can be clearly seen that the other algorithms do not give satisfactory results except for the OBSA.

VII. Conclusion

The objective of this article is to systematically study SVM-based multifeature ensemble methods for the classification of high-resolution remotely sensed imagery. The study is inspired by the fact that in recent years, researchers have developed a series of spatial and structural features, but it is difficult to find one feature that is appropriate for different image scenes. In this context, this study proposes three strategies: C-voting, P-fusion, and OBSA, implemented at both the pixel and object levels, for a combination of spectral, spatial, and semantic features. The algorithms were evaluated on three multispectral high-resolution data sets, and their performances were compared with the VS and MKL algorithms in experiments. The important conclusions are summarized as follows.

1) Both P-fusion and C-voting algorithms are effective for SVM-based multifeature fusion since in most cases they present significantly better results than the optimal single-feature classification. Furthermore, it is revealed that the probabilistic output is more appropriate than the uncertainty analysis for multifeature fusion, as the P-fusion algorithm outperforms the C-voting algorithm.
2) The VS-SVM algorithm has the potential to enhance the discrimination between spectrally similar classes by forming a hyperdimensional feature space. It gives comparable accuracies to the optimal single-feature classification. However, the proposed C-voting and P-fusion algorithms produce significantly better results than the VS-SVM in most of the experiments, as shown in Table VIII. In addition, it is revealed that the multikernel SVM is able to improve the results of simple VS-SVM for the multifeature classification.
3) Object-based C-voting and P-fusion improve the overall accuracies by 0.3–2.0% compared to their pixel-based versions. The accuracy increment can be attributed to the spatial smoothness of the segmentation.

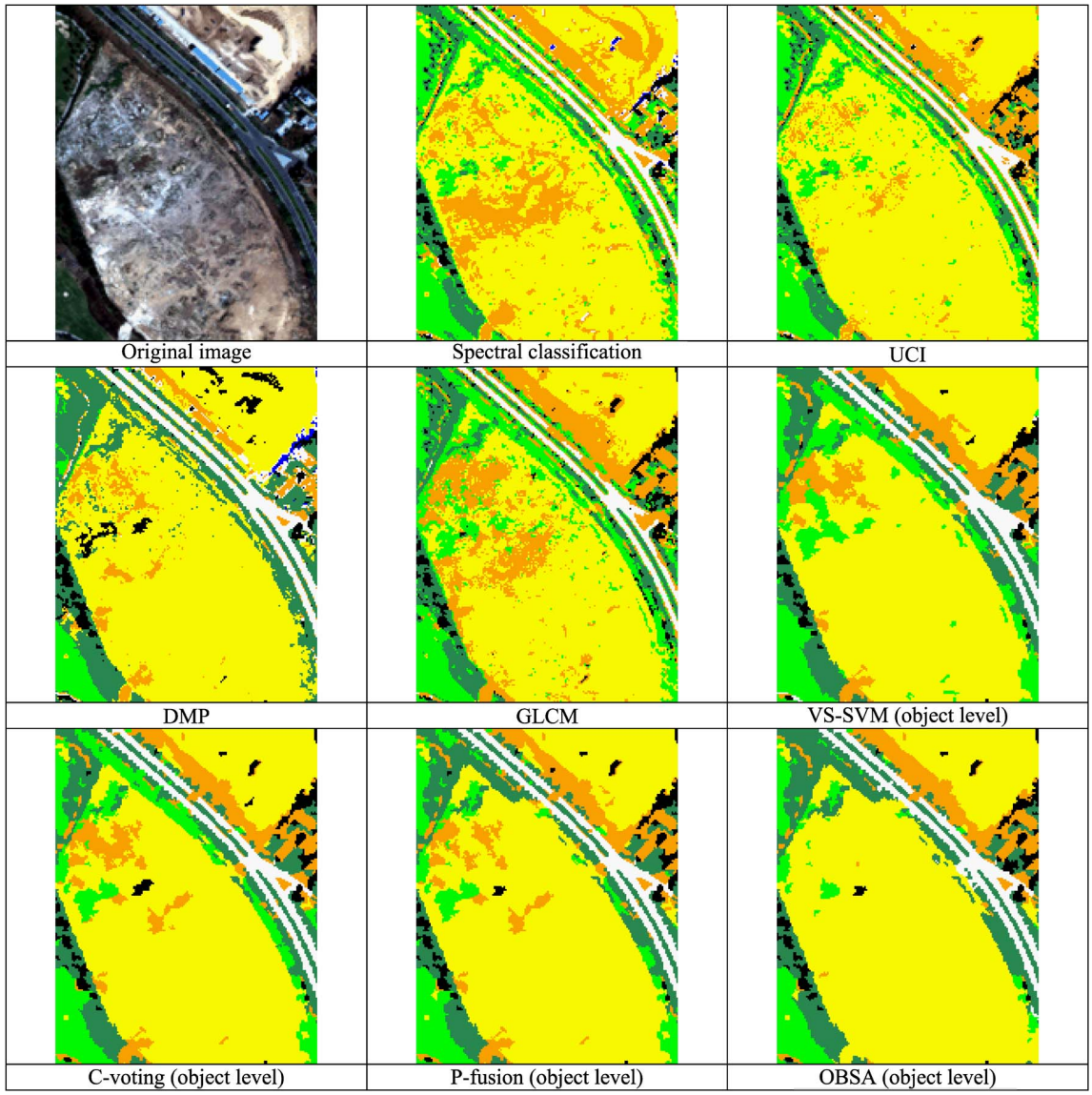

Fig. 8. Subset image classification maps for the WorldView-2 Hainan data set (white = roads, orange = roofs, yellow = soil, light green = grass, sea green = trees, black = shadow).

4) The basic idea of the OBSA is to use a series of semantic rules for a postprocessing of the multifeature SVM ensemble. It provides the most accurate results for both quantitative evaluation and visual inspection. In comparison to the model without postprocessing, the semantic analysis is able to achieve an accuracy improvement of 1–4%.

A limitation of the proposed methods is that training samples are needed for the optimization and learning of the SVMs. Therefore, in our future work, we plan to discuss the effects of the size of the training sets on the classification performance. In addition, an unsupervised version of the proposed multifeature model deserves studying, particularly for the recognition of a specific target, e.g., buildings [53].

Another limitation of the proposed methods lies in the construction of the knowledge-based rules. In this paper, the semantic analysis is used as a postprocessing of the multifeature system. Consequently, several simple rules are defined for the discrimination between the typical urban classes such as buildings-roads-soil, and water-shadow. The rules can be transferred to other image scenes by tuning the threshold parameters. In addition, the semantic rules are only applied to the unreliable objects when the ensemble classifier leads to large classification uncertainty. This is because the effectiveness of the semantic processing depends on the segmentation quality, and its blind application to all the objects could decrease the overall classification results. Therefore, future research should be related to the construction of the standard semantic rule library for high-resolution image interpretation.

It should be noted that a recently emerging field for image classification, active learning [54], is highly related to the proposed multifeature ensemble system. For instance, the entropy query-by-bagging algorithm [55] is close to C-voting as they both employ committee-based voting to evaluate the uncertainty of classification. The breaking ties algorithm [55] is similar to P-fusion as they both use the SVM-based posterior probability to evaluate the reliability of classification.

Active learning aims to choose the pixels in the candidate training pool in order to adapt the classification, while the methods presented in this study aim to choose the appropriate single-feature classifiers and reduce the uncertainty of multifeature fusion. Furthermore, although the proposed C-voting and P-fusion algorithms have proved effective for multifeature fusion, our experiments show that knowledge-based rules are more crucial for the accurate interpretation of high-resolution images. In our future research, we plan to introduce active learning into the multifeature ensemble model.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. A. Landgrebe (Purdue University, USA) for providing the HYDICE data set. We also appreciate the insightful suggestions from the anonymous reviewers.

REFERENCES

[1] [Online]. Available: http://worldview2.digitalglobe.com/

[2] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[3] S. Prasad and L. Mann Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, May 2008.

[4] X. Huang, L. Zhang, and P. Li, "An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 654–658, Oct. 2007.

[5] F. Dell'Acqua, P. Gamba, A. Ferari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[6] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[7] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[8] M. Chini, F. Pacifici, and W. J. Emery, "Morphological operators applied to X-band SAR for urban land use classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Cape Town, South Africa, Jul. 12–17, 2009, pp. IV-506–IV-509.

[9] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[10] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[11] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, Northern of Italy," *Int. J. Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, Jun. 2009.

[12] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[13] L. Zhang, X. Huang, B. Huang, and P. Li, "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2950–2961, Oct. 2006.

[14] T. Blaschke, "Object-based contextual image classification built on image segmentation," in *Proc. IEEE Workshop Adv. Tech. Anal. Remote Sens. Data*, Oct. 2003, pp. 113–119.

[15] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[16] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[17] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[18] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003.

[19] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[20] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, no. 1–3, pp. 389–422, Jan. 2002.

[21] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multi-source composite kernels for urban image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 88–92, Jan. 2010.

[22] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[23] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 65–74, Mar. 2011.

[24] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, Jun. 2009.

[25] J. Chen, C. Wang, and R. Wang, "Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2193–2205, Jul. 2009.

[26] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.

[27] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[29] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 260–264, Apr. 2007.

[30] X. Huang, L. Zhang, and P. Li, "A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform," *Int. J. Remote Sens.*, vol. 29, no. 20, pp. 5923–5941, Oct. 2008.

[31] X. Huang and L. Zhang, "Comparison of vector stacking, multi-SVMs fuzzy output, and multi-SVMs voting methods for multiscale VHR urban mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 261–265, Apr. 2010.

[32] Q. Yu, P. Gong, N. Clinton, G. Biging, and D. Schirokauer, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 7, pp. 799–811, Jul. 2006.

[33] M. Pesaresi, A. Gerhardinger, and F. Kayitakire, "A robust built-up area presence index by anisotropic rotation-invariant textural measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 3, pp. 180–192, Sep. 2008.

[34] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[35] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.

[36] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[37] B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2880–2889, Jul. 2010.

[38] M. Petrakos, J. A. Benediktsson, and I. Kanellopoulos, "The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 11, pp. 2539–2546, Nov. 2001.
[39] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, Oct. 2006.
[40] R. Yager, "On the specificity of a possibility distribution," *Fuzzy Set. Syst.*, vol. 50, no. 3, pp. 279–292, Sep. 1992.
[41] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 11, pp. 610–621, Nov. 1973.
[42] H. Y. Yoo, K. Lee, and B. D. Kwon, "Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery," *Int. J. Remote Sens.*, vol. 30, no. 23, pp. 6219–6239, Dec. 2009.
[43] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer-Verlag, 1999.
[44] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6775, pp. 788–791, Oct. 1999.
[45] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, Mar. 1999.
[46] Y. O. Ouma, J. Tetuko, and R. Tateishi, "Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery," *Int. J. Remote Sens.*, vol. 29, no. 12, pp. 3471–3456, Jun. 2008.
[47] M. Pesaresi, "Texture analysis for urban pattern recognition using fine-resolution panchromatic satellite imagery," *Geograph. Environ. Model.*, vol. 4, no. 1, pp. 47–67, May 2000.
[48] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.
[49] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*. Cambridge, MA: MIT Press, 1999, pp. 61–74.
[50] D. Comanicium and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[51] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.
[52] G. Camps-Valls, L. Gómez-Chova, J. Munoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.
[53] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral GeoEye-1 imagery," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 7, pp. 721–732, Jul. 2011.
[54] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, Sep. 2011.
[55] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

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

Currently, he is an Associate Professor at the LIESMARS, Wuhan University. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications. He has published more than 25 peer-reviewed articles in international journals such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, *Photogrammetric Engineering and Remote Sensing*, and *International Journal of Remote Sensing*.

Dr. Huang has served as a Reviewer for most of the international journals for remote sensing. He was the Recipient of the Top-Ten Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the best paper in image analysis and interpretation from the American Society for Photogrammetry and Remote Sensing. In 2011, he was the Recipient of the New Century Excellent Talents in University from the Ministry of Education of China. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society as a Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Liangpei Zhang (SM'08) received the B.S. degree in physics from Hunan Normal University, ChangSha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

Currently, he is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, as the Head of the Remote Sensing Division. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He has more than 240 research papers and 5 patents. Currently, he is the Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang regularly serves as a Co-Chair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and the geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence, International Journal of Image and Graphics, International Journal of Digital Multimedia Broadcasting, Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing*. He is a Fellow of the Institution of Electrical Engineers, Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme, Executive Member for the China Society of Image and Graphics, etc.