A Mean Shift Vector-Based Shape Feature for Classification of High Spatial Resolution Remotely Sensed Imagery

Rongjun Qin, *Student Member, IEEE*

***Abstract*—The development of very high spatial resolution remote sensing sensors opens a new era for mapping the earth with submeter level of detail, whereas the increased resolution brings about difficulties for the land-cover classification in terms of intra-class variability and inter-class similarity. This paper presents a novel spatial feature, mean shift (MS) vector-based shape feature (MSVSF), to improve the classification accuracy of very high resolution (VHR) remote sensing imagery. MSVSF is a 3×1 feature vector extracted in per-pixel fashion. It describes the shape of a spectrally homogeneous area surrounding each pixel by measuring the two-dimensional (2-D) image deformation of its local area imposed by the MS vector. The proposed feature is particularly effective to discriminate objects with similar spectral response but different 2-D shapes, such as buildings and roads. Independent component analysis is adopted to extract spectral features and Support Vector Machine (SVM) classifier is adopted to classify the spectral and spatial features and several state-of-the-art spatial/structural features are compared to the proposed feature. A synthetic experiment demonstrates that the proposed feature has good capability to describe 2-D shapes with different scale, two real dataset experiments on QuickBird and IKONOS images show MSVSF has achieved better overall accuracy (OA) than the compared ones. In addition, the MSVSF feature is extended to the object-based classification (OBC), and the result shows that the MSVSF is effective to improve the classification accuracy on high resolution images of the urban area.**

***Index Terms*—Classification, IKONOS, local similarity area, mean shift vector, QuickBird, support vector machine (SVM), very high resolution (VHR) remotely sensed imagery.**

I. Introduction

A MAJOR challenge of land-cover classification is the increasing level of image details due to the availability of the very high resolution (VHR) of the advanced satellite sensors, e.g., IKONOS 1 m, WorldView 0.5 m, QuickBird 0.61 m, and Pleiades 0.5 m. Urban scenes become more complex and many different objects exist with similar spectral signatures. It is commonly agreed that the increased resolution does not facilitate the same level of improvement of the classification accuracy, especially for methods designed to classify the low-resolution satellite image. Therefore, it is necessary to explore new methods and incorporate the spatial features for dealing with the VHR images.

Manuscript received December 13, 2013; revised August 06, 2014; accepted September 01, 2014. This work was supported in part by Singapore-ETH Center for Global Environmental Sustainability (SEC) and in part by Singapore National Research Foundation (NRF), and in part by ETH Zurich.

The author is with Singapore ETH Center, Future Cities Laboratory, 138602 Singapore (e-mail: rqin@student.ethz.ch).

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/JSTARS.2014.2357832

A. Related Works

There have been a lot of efforts devoted to improving the classifiers and features: Tuia *et al.* [1] proposed a multikernel approach by combining different kernels of Support Vector Machine (SVM) to address different features. Very recently, Dópido *et al.* [2] developed a semisupervised learning framework for hyperspectral image classification, which assumed that the unlabeled training samples could be obtained by a limited number of available training samples, and they tried to improve the classification accuracy by increasing the number of self-learned training samples. However, this method was dependent on the assumption that the pixels with similar spectral signature belong to the same class. This might be possible for hyperspectral images, but not for multispectral images, since they contain many spectral ambiguities (e.g., roofs and roads, water and shadow). Hester *et al.* [3] adopted ISODATA (Iterative Self-Organizing Data Analysis Techniques) on the spectral response of the VHR imagery, and 89.3% overall accuracy (OA) was reported. In their work, the urban objects such as building roofs, roads, parking lots were categorized as a single class called impervious surface. This is meaningful for classification at landscape level, but not sufficient for urban mapping applications, since the impervious surfaces need to be elaborated into more detailed objects (e.g., roofs, roads, and pedestrian paths). Consequently, poor results were obtained considering these classes if mere spectral features were used [4]–[6] for the classification of urban objects.

Researchers had proved that introducing features describing the spatial pattern of the image content could consequently bring notable enhancements of the classification accuracy [4], [7]–[9]. Sirmacek and Unsalan [10] introduced the spatial pattern of the interest points on the building class, where only the panchromatic bands of the aerial/satellite images were used. Good detection accuracy was reported thanks to the spatial characteristics of building class. Gray Level Concurrent Matrix (GLCM) is also a spatial feature that extracts texture patterns over a window. Properties such as sum, average, contrast, and entropy are computed as features describing the local area of each pixel. It was first proposed by Haralick *et al.* [11] for image classification, and recapped later by Zhang [12] for building extraction. Recently, GLCM was adopted by Pacifici *et al.*

for classifying [13] through a neural network approach, and high-classification accuracy was reported.

Pesaresi and Benediktsson [14] adopted a set of morphological operators, together with structural elements of different size and shapes to build differential morphological profiles (DMP) for image segmentation, where its similar forms were proven to be particular useful for building detection applications [15], [16]. Later they extended the full DMP feature for IRS-1C and IKONOS image classification [17], and had obtained good OA in their experiments. Tuia *et al.* [18] did a study on finding the most relevant morphological operators and they highlighted that the opening and closing operators were the most suitable ones for classifying high spatial resolution image such as the QuickBird imagery. However, it needs a large number of training samples and high-computational load in the feature selection stage.

Shackelford and Davis [5] proposed the length–width extraction algorithm (LWEA) to extract the shape of a spectrally homogeneous area surrounding a centric pixel. It radiated searching lines that were spectrally similar to the centric pixel, and then extracted the longest and shortest paths of predefined searching orientation. The LWEA was viewed to be an effective feature to discriminate spectrally similar classes. Zhang *et al.* [6] improved the LWEA feature by summing up the length of all the paths, which formed a one-dimensional (1-D) feature named pixel shape index (PSI), and then they combined it with transformed spectral features to obtain better classification result. Later Huang *et al.* [19] incorporated the edge information into PSI, and assigned large weights on the edge pixels for the PSI computation, which improved the robustness of the feature extraction. Yoo *et al.* [20] developed an urban complexity indicator (UCI) based on the 3-D wavelet analysis, and it was further extended by Huang and Zhang [21] to a multiscale approach, in which the window size and the decomposition levels were taken into account to achieve a more robust performance in both urban and natural landscape areas.

A very recent study from Huang and Zhang [4] on classifying the VHR images had proved that, to achieve more reliable classification results, the joint use of spectral and spatial feature was necessary for further development of an integrated classification system. They mentioned that the urban layouts were complex and varied a lot among different places, so it was hard to find a single spatial feature working well under different conditions. Therefore, the complex system considering statistical analysis was developed by integrating spectral, spatial, and semantic information. However, such semantic information should be predefined on a case-by-case fashion and dependent on human operators and users with relevant expertise. Therefore, robust features working stable under different urban scenarios are in a high demand.

B. A New Spatial Feature

In sum, though the theoretical backgrounds of different spatial features vary, they can be generally divided into two groups concerning their feature extraction modes: 1) feature extraction based on a predefined window and 2) feature extraction on a local homogenous/similarity area (LSA). The LSA of a pixel is defined as a two-dimensional (2-D) connected image patch, which has similar spectral responses as this pixel. Features belong to the first group extract the statistics of the spectral response over the predefined window; typical examples are DMP and GLCM, and UCI features. Features in the second group extract the shape patterns of 2-D topological areas that are spectrally similar and spatially connected to the target pixels, and typical examples are LWEA and PSI. Features from both groups have advantages and deficiencies: Group 1 is robust for feature extraction, but is inaccurate near class boundaries, and the texture statistics may not properly describe the complex spatial patterns of the urban objects, as the shape of the objects sometimes varies with the perspective effects of VHR images [22], [23]. Group 2 is effective to discriminate classes that vary in 2-D shapes of the LSA, but the shape feature extraction process can be easily affected by inaccurate LSA due to noises.

Given the pros and cons of the both groups, a combination of these two should be considered as a tradeoff. In this paper, a novel spatial feature, mean shift vector-based shape feature (MSVSF), is developed, which integrates both of the feature extraction modes. The basic idea is to compute the mean shift (MS) vectors of each pixel, and then apply the MS vector on the LSA to measure the 2-D deformation of the area. The advantage of the idea lies in two aspects: 1) the MS vectors are computed by enrolling information over a predefined window of each pixel, such that the MS vectors of pixels in the LSA introduce information beyond the LSA; 2) the MSVSF measures the 2-D deformation of the LSA caused by the MS vector instead of the absolute shape of the LSA, which is assumed to be more robust.

The rest of the paper is organized as follows. Section II recaps the concept of the mean shift vector and Section III introduces the proposed MSVSF feature. In Section IV, the independent component analysis (ICA) and SVM classifier are briefly introduced. Section V compares the OA with the state-of-the-art spatial features by fusing them with ICA feature into SVM classifier, and then extends the MSVSF feature into object-based classification (OBC). Section VI concludes the paper by discussing the pros and cons of the MSVSF feature.

II. Mean Shift Vector

MS analysis [24] has gained great success in the domain of image segmentation and edge-preserving filtering, but the applications of the MS vector have got surprisingly low attention. To the author's knowledge, there are only very few publications concerning the application of MS vector. A recent work adopts the MS vector to improve the gradient vector flow for medical image segmentation [25]. The MS vector is the shifting vector of each pixel in the MS process; it can be seen as the gradient vector of a pixel in its local region, providing the information beyond the single pixel itself. MS analysis is first proposed by Fukunaga and Hostetler [26] for density estimation; Cheng adopted [27] the MS analysis for mode seeking and clustering of the discrete data, and then it was recapped and adopted by Comaniciu and Meer [24] to low-level computer vision tasks such as edge-preserved filtering and image segmentation. It is a kernel density estimation approach: given a

set of data $\{x_i\}, \quad i = 1, 2, \ldots, n \in \mathbb{R}^d$, the density estimation function can be written as

$$f(x) = \frac{c}{nh^d} \sum_{i=1}^{n} K\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \tag{1}$$

where the kernel function $\mathrm{K(x)}$ models the correlation between dataset $\{\mathrm{x_i}\}$ and the density center x, and a typical Gaussian kernel is adopted in this paper as it models density as a smooth function. h is the bandwidth and c is the normalization parameter. The key idea of the MS approach is to compute the local maxima of the kernel function, which is located by finding the zero of the gradient namely $\nabla \mathrm{f} = 0$. To be more specific, the MS analysis over a multiband raster image is mainly divided into two domains: the spatial domain $\mathrm{x_s}$ and spectral domain $\mathrm{x_r}$, which is modeled as below

$$f(x) = \frac{c}{nh_s^2 h_r^p} \sum_{i=1}^{n} K\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right) \tag{2}$$

where $x_s = [t_y, t_x]$ represents the row (t_y) and column (t_x) index of each pixel, and $x_r = [r_1, \ldots, r_p]$ denotes the values of each multispectral band. h_s and h_r are the bandwidths of the spectral and spatial domain, respectively. n is the total number of samples used for computation. The spatial domain is a 2-D space represented by an image grid, and the dimensionality of spectral band domain p is dependent on the number of bands or the dimension of the selected spectral features extracted from the spectral bands. Differentiating (2) with the spatial vector x_s yields

$$\begin{aligned}\frac{\partial f(x)}{\partial x_s} &= \frac{c}{nh_s^4 h_r^p} \sum_{i=1}^{n} (x_s - x_{si}) K' \\ &\times \left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right) \\ &= \frac{2c}{nh_s^4 h_r^p} \left[\sum_{i=1}^{n} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)\right] \\ &\cdot \left[x_s - \frac{\sum_{i=1}^{n} x_{si} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}{\sum_{i=1}^{n} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}\right]\end{aligned} \tag{3}$$

where K' is the differentiation of K. The first term of (3) is proportional to the density estimated at $x = x_s \otimes x_r$ with kernel K'

$$\begin{aligned}f_{h_s,h_r,K'} &= \frac{2c}{nh_s^2 h_r^p} \sum_{i=1}^{n} K' \\ &\times \left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right).\end{aligned} \tag{4}$$

The second term is the MS in the spatial domain

$$m_{h_s,h_r,K'} = x_s - \frac{\sum_{i=1}^{n} x_{si} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}{\sum_{i=1}^{n} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}. \tag{5}$$

Term $f_{h_s,h_r,K'}$ can be assumed as a positive number [24]. By imposing $m_{h_s,h_r,K'} = 0$, the MS vector can be computed as

$$x_s = \frac{\sum_{i=1}^{n} x_{si} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}{\sum_{i=1}^{n} K'\left(\left\|\frac{x_s - x_{si}}{h_s}\right\|^2 + \left\|\frac{x_r - x_{ri}}{h_r}\right\|^2\right)}. \tag{6}$$

In practical implementation for image processing, the samples for computing the mean shift are usually restricted to a square window, and the length of its side is defined as $(2h_s + 1)$ to ensure the spatial component $\left\|\frac{x_s - x_{si}}{h_s}\right\|$ is within [0,1].

III. Mean Shift Vector-Based Shape Feature

To exploit the spatial characteristics of each pixel in the VHR remote sensing image, Shackelford and Davis's LWEA algorithm [5] searched for the spectrally similar and spatially connected pixels in each direction, and took the maximal length and the minimal length as a 2-D feature vector to depict the spatial signature of different classes. Similarly, PSI [6] counted the total number of spectrally similar pixels to the centric pixel in the multidirection scan lines, engendering a 1-D feature. Both of the aforementioned methods are based on extracting shape information from the LSA, which is approximated by multidirectional scanning lines. The proposed MSVSF imposes the MS vector on the local area to compute the deformation of the LSA induced by the MS vector. Most of the spatial/structural features extract information on the mere shape of the LSA; however, the MS vector provides semiglobal information: each pixel in the LSA enrolls information from its locally predefined window, and the MS vector of the LSA boundary pixels introduces information outside the LSA. Atypical example showing the advantage of the MS vector is in Fig. 1, where MS vectors in patches of different classes show different distributions: MS vectors have strong convergence to the central point of the roof areas, since roof areas are usually closed patches with clear boundaries; MS vectors in the road area has strong convergence to the central line; for the tree area they show relatively weak convergence to the central of the trees, and in the open ground they show no convergence. Such discrepancies of MS vector on different urban classes create a possible avenue to increase the classification accuracy for the VHR images. Like LWEA, PSI, GLCM features, the MSVSF feature is computed in a per-pixel fashion, and the computation of MSVSF mainly consists of the following three steps: 1) multiscale pixel-wise MS vector computation for VHR image; 2) LSA extraction with efficient region growing method; and 3) computation of the MSVSF. These steps will be introduced in detail in the following sections.

A. Multiscale Pixel-Wise MS Vector Computation

The MS vectors are computed for all the pixels in the VHR image, such that each pixel has a 2-D vector indicating the magnitude and the gradient orientation. It is worth noting that the bandwidths of the spatial and spectral domain h_s and h_r

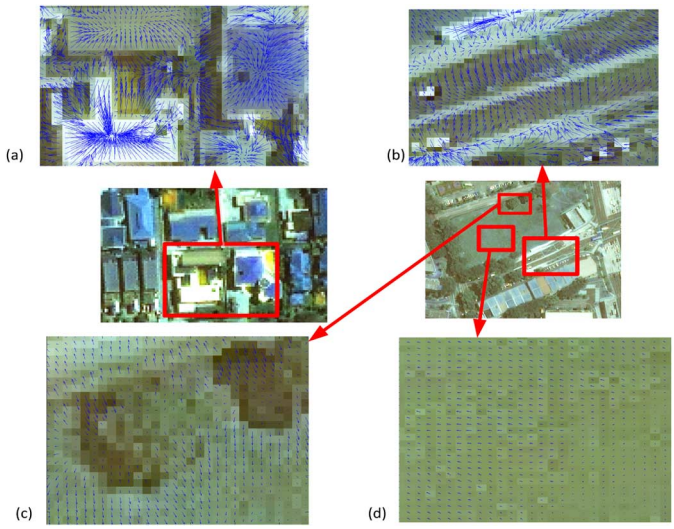

Fig. 1. Examples of MS vectors in different classes: (a) building roof; (b) roads; (c) trees; and (d) open ground.

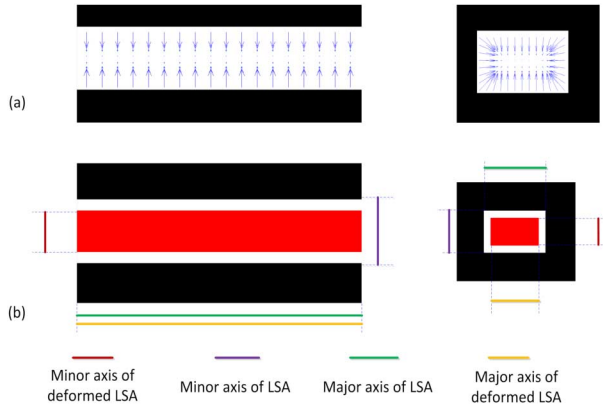

Fig. 2. Intuitive example of the MS vector (blue arrows, exaggerated for visual representation) and the major and minor axes for LSA (red region) and deformed LSA (white region): (a) the MS vector and (b) the visual illustration of major and minor axes.

are important for the MS vectors. h_r controls the sensitivity of spectral similarity of the centric pixel to the pixels in its local area, and h_s controls the weight of spatial proximity in the 2-D image space. For the MS vector calculation, h_r and h_s jointly control the magnitude and the orientation of the vector. A small h_r is able to depict small gradients but easy to be affected by noise, whereas a large h_r is more robust to noise but could ignore small but significant gradient. h_s controls the volume of information involved for the MS vector computation: for a large h_s, the computation involves more information in a local area, representing the gradient information in a large scale, and vice versa. The scales of objects vary a lot, e.g., shopping malls are much larger than the private house unit, and thus it is essential to compute the MS vector in a multiscale fashion. Based on (6), a multiscale approach is adopted to calculate a more robust MS vector

$$x_s = \frac{1}{m}\sum_{t=1}^{m}\frac{\sum_{i=1}^{n} x_{si}K'\left(\left\|\frac{x_s-x_{si}}{h_{st}}\right\|^2+\left\|\frac{x_r-x_{ri}}{h_r}\right\|^2\right)}{\sum_{i=1}^{n} K'\left(\left\|\frac{x_s-x_{si}}{h_{st}}\right\|^2+\left\|\frac{x_r-x_{ri}}{h_r}\right\|^2\right)} \quad (7)$$

where $\{h_{st}\},\ t=1,2,\ldots,m$ is a set of values for the spatial bandwidth, m can be a fairly small number to archive computational efficiency. Given a range of h_s as $[1, h_{s,max}]$, $\{h_{st}\}$ takes m equally spaced values within this range for the multiscale MS vector computation. In this paper, $m=3$ throughout all the experiments and $\{h_{st}\}=[1,5,10]$.

B. LSA Extraction With Region Growing Method

To extract the shape information from the LSA, the LWEA and PSI features radiate multiple scan lines in different directions to infer the shapes, while it can be easily affected by dot noises. To address this problem, LSA is extracted in a region-growing manner: taking the target pixel as the seed point, the algorithm iteratively accepts pixels into LSA if the spectral difference of the spatially connected pixels in a four-neighborhood is less than a predefined threshold T_h unless the sum of all the differences is larger the predefined upper bound T_b. To achieve a linear processing time, a stack can be used for the region growing process. The distances of the colors are measured with Euclidean distance in the RGB color space, and the color intensity is scaled to [0,1].

C. Computation of MSVSF

The MSVSF measures the deformation of the LSA before and after applying the MS vector to it. By shifting each pixel in the LSA based on its corresponding MS vector, the LSA is deformed as a new area. The deformation v is measured as the ratio of the major axis length and minor axis length of the LSA before and after applying the MS vector, which is computed as

$$v^T = \frac{S'}{S} = \frac{[M',N']}{[M,N]} = \left[\frac{M'}{M},\frac{N'}{N}\right] \quad (8)$$

where M, N denotes the length of the major and minor axes of the LSA, respectively, and M' and N' are the length of the major and minor axes of the deformed LSA. The major axis of the LSA is defined as the axis, which has the largest variance and the minor axis is the one orthogonal to the major axis. An intuitive example is shown in Fig. 2(a), where the MS vectors of a strip-shaped and rectangular areas are computed. Fig. 2(b) shows the deformed area in red, and since both of the two areas are upright rectangular area, the major and minor axes lie on either horizontal or vertical direction. It can be seen that the strip-shaped LSA deforms only in the vertical direction, whereas the rectangular LSA deforms in both direction. Moreover, the strip shaped LSA does not have horizontal deformation as the left and right side are open, where the MS vectors do not converge. Thus, it demonstrates that the deformation not only takes information within the LSA, but beyond the LSA.

The length of the major and minor axes can be calculated as the variance of the pixel coordinates within LSA along the corresponding axis. Let X denote the 2-D coordinates of the pixels within the LSA, and $\bar{X}$ be the zero mean vector of X, defined as $\overline{X} = X - \mathrm{mean}(X)$. " $\mathrm{mean}(\cdot)$" computes the mean value

for each dimension of X. $\bar{X}$ can be transformed to a coordinates system by a linear orthogonal transformation, where the two axes of the new coordinate system holds the largest and smallest variance of $\bar{X}$. The variance of the major and minor axes can be computed as the largest and second largest eigenvalue of $\bar{X}^T\bar{X}$ [28]. Since $\bar{X}^T\bar{X}$ is a 2×2 matrix, the variance of the major and minor axes equals to the two eigenvalues λ_1 and λ_2 of $\bar{X}^T\bar{X}$. Let X' be the 2-D coordinates of pixels within the deformed LSA, and $\bar{X}'$ be its zero mean vector, where

$$X' = X + MSV, \quad \bar{X}' = X' - \text{mean}(X') \tag{9}$$

where *MSV* is the MS vector for the corresponding pixel. Let ${\lambda_1}'$ and ${\lambda_2}'$ be the eigenvalues of $\bar{X}'^T\bar{X}'$. Thus, the deformation v can be computed as

$$v = \frac{[{\lambda_1}', {\lambda_2}']^T}{[\lambda_1, \lambda_2]^T} = \left[\frac{{\lambda_1}'}{\lambda_1}, \frac{{\lambda_2}'}{\lambda_2}\right]^T \tag{10}$$

v is defined as deforming ratio vector (DRV). It is scale invariant since it only computes the deformed ratio of a LSA. Building roofs normally have close boundaries, whereas roads are open in the road direction. Fig. 3 shows the 2-D plots of different features at 100 representative pixels per class. In Fig. 3(a), x-axis shows the PSI values (1-D) and different y values indicate different classes. It can be seen that in the x-axis there is a large overlap of PSI values of different classes. Fig. 3(b) shows the 2-D feature vector of the LWEA feature, where the scatter points of different classes mixed with each other. Fig. 3(c) shows the scatter plot of the DRV, where a part of building roofs and a part of roads overlapped with trees, but good separability of the building and roads can be observed. It can be seen that at the LSA of the building roofs, v is less than 1 in both of the major and minor axes, while at the LSA of the road, v is approximately 1 (no deformation) in the major axis and 0.5 in the minor axis. This shows promising evidence of the potentials to improve the classification accuracy.

In practice, it is effective for DRV to discriminate buildings and roads, and the DRV is robust to small errors of LSA, but it still suffers from large artifacts caused by either noise or failure due to the extraction of LSA on complex urban scenes with fixed parameters. Therefore, to improve the robustness, the elongation of the LSA $Elong = \frac{N}{M}$ is incorporated as a third element concatenating to the DRV, which is quite effective in discriminating buildings and roads [29], constituting the MSVSF ($\otimes$ is the concatenation operator)

$$\text{MSVSF} = v_1 \otimes Elong. \tag{11}$$

Fig. 4(b)–(d) shows the values of the three components of MSVSF of an image patch in Fig. 4(a). The different patterns of the buildings and roads can be clearly observed: since buildings have clearer boundaries on each side, it has larger deformation on both the major and minor axes (resulting in a small value). The area of road has smaller deformation in the along-road direction and larger deformation in the cross road direction, which results in large value in the first component and small value in the second component of MSVSF. Different from the third component (elongation), the first and second components

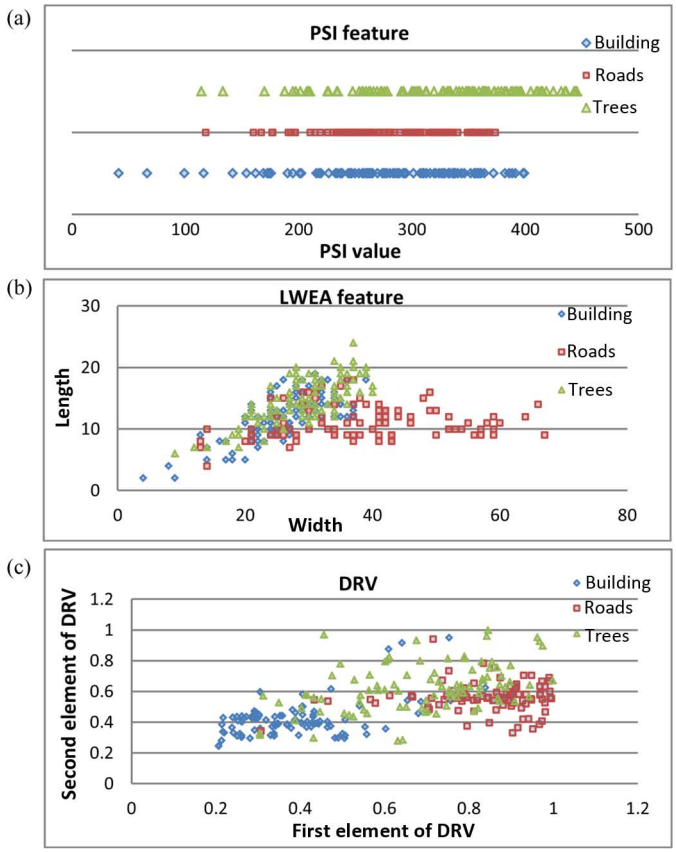

Fig. 3. 2-D plots of values of different spatial features: PSI, LWEA, and DRV. The same window size and threshold of similarity measurement are adopted for computing the above statistics. (a) PSI value of building, roads of trees; (b) LWEA values of building, roads and trees; and (c) DRV values of building, roads, and trees.

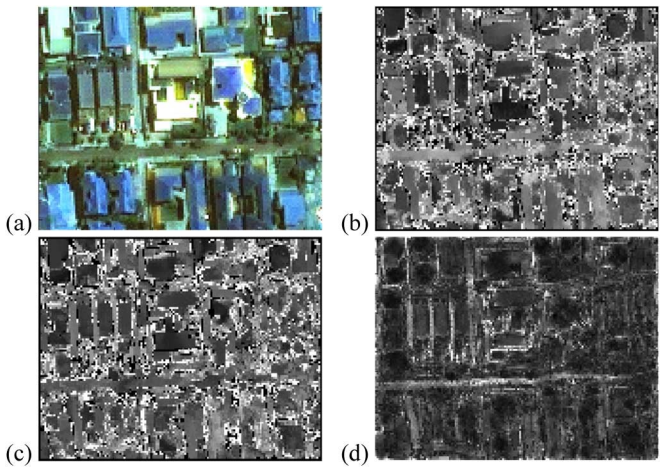

Fig. 4. Pixel-wise visual representation of the proposed MSVSF features: (a) original image patch; (b) the first component of MSVSF, showing the deformation ratio of the major axis; (c) the second component of MSVSF, showing the deformation ratio of the minor axis; and (d) the elongation value (values are properly scaled for visual effects).

of MSVSF provide clearer boundary between the buildings and their surroundings, and such characteristics provide more support for the classifier to distinguish object boundary and hence to reduce misclassifications.

IV. Spectral Feature Extraction and SVM Classifier

It is understood well that mere spatial/structural information is not sufficient for the VHR remote sensing image classification. These features normally need to be combined with the spectral signatures, which are the driving forces for remote sensing image classification. The most common way is to use the original spectral responses in each band. However, the signal tails of the each multispectral band may mix with each other, which affect the classification accuracy of the classifiers. Therefore, the signals from different bands need to be decorrelated using signal processing techniques. There were many spectral features proposed in the past, such as ICA [6], principle component analysis (PCA) [30], and nonnegative factorization (NMF) [31], mainly for the purpose of dimension reduction and robust representation of the spectral response.

A. Independent Component Analysis

In this study, the ICA transformation is adopted to perform the signal decorrelation, which is proven to be better than PCA and RGB feature for many classification applications [6], [32]. ICA models the observed signals X with linear mixture of statistically independent signals D

$$X = WD \tag{12}$$

where W stands for the coefficient matrix, the goal is to find the signal D, while W is also unknown. The ICA problem is very close to the blind source separation (BSS), the difference is that the components of D are assumed to be random variables, where noise term can be also added to the model, and the solution maximize the *Non-Gaussianity* of $W^T X$ by iterative optimization process [33].

In the context of the four-band multispectral images, the original signal X is a $m \times 4$ vector, where m is the number of the pixels. The transformed signal $D = [d_1, d_2, d_3, d_4]$ has the same dimension, and will be concatenated to the spatial feature vector $S = [s_1, \ldots, s_p]$ for classification, each element of the feature vector is normalized to [0,1] for putting into the SVM classifier

$$\text{Feature vector} = \left[\bar{d_1}, \overline{d_2}, \overline{d_3}, \overline{d_4}, \overline{s_1}, \ldots, \overline{s_p}\right]$$

where

$$\overline{d_j} = \frac{d_j - d_{\min}}{d_{\max} - d_{\min}}, \overline{s_j} = \frac{s_j - s_{\min}}{s_{\max} - s_{\min}} \tag{13}$$

where $d_{\min}$ and $d_{\max}$ represent the minimum and maximum of the corresponding ICA components over all the pixels in the image; $s_{\min}$ and $s_{\max}$ are the minimum and maximum of the corresponding spatial feature components.

B. SVM Classifier

SVM is widely used in many classification and machine learning applications [34], [35]. It tries to find a hyper plane that is the farthest from the multidimensional training samples from different classes. It tries to learn the form $f(\boldsymbol{x}) = \langle \boldsymbol{w} \cdot \boldsymbol{x} \rangle + b$ from the sample data $\{\boldsymbol{x_i}, y_i | i = 1, \ldots, N\}$, to maximize the margin between the hyper plane and the closest sample data. $\boldsymbol{x_i}$ is an n-dimensional feature vector as shown in (13), and $y_i = \pm 1$. The maximization can be achieved by minimizing $\|w\|^2/2$ will under the constraint $y_i(\langle \boldsymbol{w} \cdot \boldsymbol{x_i} \rangle + b) \geq 1$. The problem can be transformed to a dual problem by using the Lagrangian formulation

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j}^{N} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \right\}$$
$$\text{subject to} \quad \sum_{i=1}^{N} \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, \ldots, N \tag{14}$$

where

$$\boldsymbol{w} = \sum_{j=1}^{N} \alpha_j y_j \boldsymbol{x}_j \tag{15}$$

where $\{\alpha_i | i = 1, \ldots, N\}$ is the Lagrangian multiplier. For details of the SVM classifier, see [34], [36], [37]. $\langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle$ is called the kernel of SVM, which measures the similarity between two feature vectors, and it can be replaced with other functions for the similarity measurement [1], [38]. In this paper, the Gaussian radial basis function (RBF) is employed, as it is proven to be effective in many classification applications [39]. The RBF is formed as

$$K(\boldsymbol{x_i}, \boldsymbol{x}) = \exp\left(-\gamma \|\boldsymbol{x_i} - \boldsymbol{x}\|^2\right) \tag{16}$$

γ controls the weight similarity measurement, and is set to be 1 in the subsequent experiments, as the feature vectors are normalized. Since SVM is a binary classifier, and there are two main strategies to turn the binary classification into multiclass, "one against one" (OAO), and the "one against all" (OAA) [40], where OAO classifies each possible pair of class and then keeps the most common label for each pixels, while OAA classifies each class against the rest, choosing the label with largest confidence for each pixel. OAA is adopted in this paper, as it is said to be better when the number of class is small (less than 10) [41].

In sum, the classification workflow can be summarized in Fig. 5.

V. Experiment and Analysis

A. Synthetic Experiments

To test the proposed feature in an ideal situation, a synthetic experiment is carried out for different spatial features. A synthetic class map is used as the image, where pixels belonging to the same class are painted with the same color [Fig. 6(a)]. The purpose is to exclude the uncertainty of the LSA extraction, and focus only on the spatial features' capability of separating classes with different 2-D shapes.

As shown in Fig. 6(a), the synthetic image consists of four classes: building, road, tree, and grass. Buildings are simplified

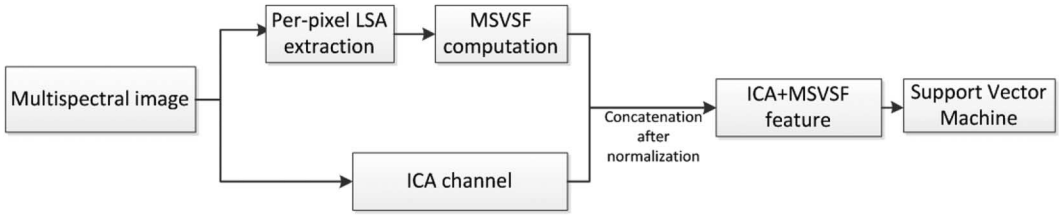

Fig. 5. Workflow of the classification process.

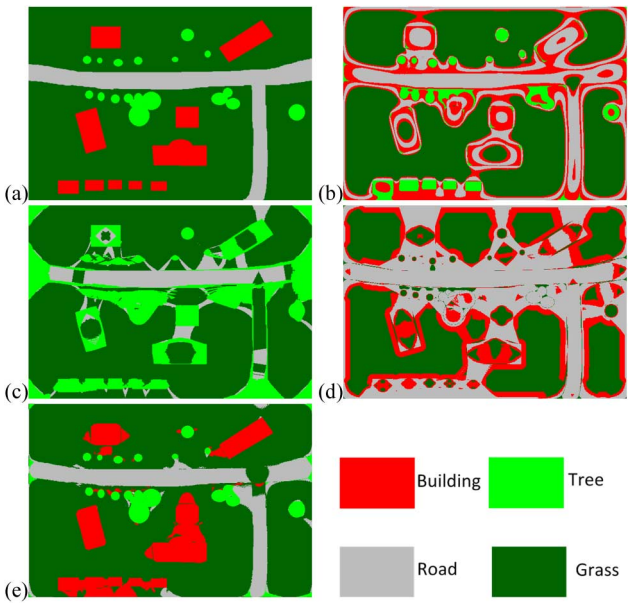

Fig. 6. Synthetic experiments on the spatial features: (a) original synthetic images; (b) classification results with PSI (OA: 64.67%); (c) classification results with LWEA (OA: 74.39%); (d) classification results with elongation (OA: 51.64%); and (e) classification results with MSVSF (OA: 92.54%).

as rectangles with different lengths, widths, and scales; trees are approximated as isolated and grouped blobs with different scales; roads are abstracted as strip shapes and the grass is described as open and homogeneous area. Due to the characteristics of the synthetic data, only features related to LSA will be tested (PSI, LWEA, Elongation, and MSVSF). Features relying on texture patterns such as GLCM will certainly produce bad results on the synthetic data, as there are no texture patterns for each class in this experiment. The synthetic image contains 639×437 pixels, and the radius of line scanning procedure for extracting LSA for PSI and LWEA are 50 pixels to ensure the extracted LSAs are meaningful.

Hundred pixels for each class are randomly selected as the training sample, and only spatial features are used. Fig. 6(b)–(e) shows the classification results of the tested features. PSI misclassified most of the building pixels, and LWEA missed almost all the building pixels. This is due to the fact that the shape measurement of PSI and LWEA is scale variant. While for the elongation feature of the LSA, it misclassified the trees completely as the grasses. This is because the grass area is an open area, the LSA of which is homogenous, while the trees are approximate as round blobs, which have homogenous shape as well. The result of elongation feature also demonstrates that elongation alone cannot produce usable results. The results of MSVSF have shown good separability between different shapes, and have also achieved a high OA (shown in Fig. 6).

TABLE I
STATISTICS OF THE REFERENCE DATA AND SAMPLES

No. of pixels	QuickBird dataset	IKONOS dataset
Building	81 090	83 386
Tree	15 745	43 008
Road	23 039	23 779
Grass	22 463	25 445
Shadow	3347	4992
Water	6466	N/A
Ground	4806	3294
Total	156 956	183 904
Training samples	2100	1800

B. Experiments on Real Datasets

The synthetic experiment designed for the spatial features based on LSA helps to understand the validity of the MSVSF, but it needs to be tested in real dataset. Therefore, experiments are performed on IKONOS four band (red, green, blue, and near infrared) pan-sharpened dataset and QuickBird four band pan-sharpened imagery, and both of the datasets are ortho-ready products. The IKONOS dataset maps an area of a developed country in a tropical area, and the QuickBird dataset maps an underdeveloped country. The ground sampling distances (GSD) of both dataset are 1 and 0.6 m, respectively. In the experiments, the radiometric value is scaled to 8-bit to extract the spatial features for computational convenience.

Two test sites are selected from each of the datasets under different urban scenarios and complexities: 1) an urban area with an underdeveloped country (QuickBird dataset, with 802×555 pixels) and 2) a Central Business District (CBD) area of the developed country (IKONOS dataset, with 819×716 pixels). The reference data are manually sketched by carefully inspecting the data sources and Table I shows the number of pixels of the reference data for each class in the two experiments. The training samples are generated by randomly sampling the reference data: for each experiment, 300 pixels from each class are randomly selected from the training sample ($\sim 1.3\%$ of the reference pixels for the QuickBird dataset and $\sim 0.098\%$ for the IKONOS dataset), and the rest of the pixels are used for testing the classification accuracy. A fivefold cross-validation (CV) is performed on the training samples to evaluate the robustness of the proposed features. The reference data are shown in the second row in Fig. 7. The confusion matrices are computed, the resulting OA and Kappa coefficient (KC) are compared with

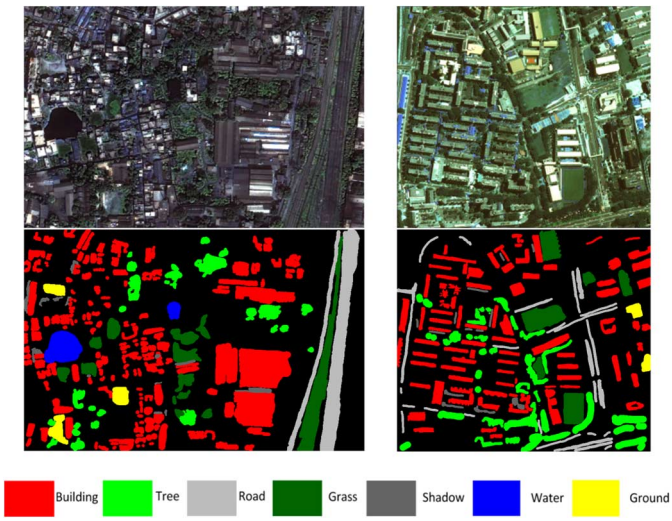

Fig. 7. Left column: QuickBird dataset, Urban area of the Bangladesh, and the hand-drawn reference data. Right column: IKONOS dataset, CBD area of Singapore, and the hand-drawn reference data.

TABLE II
SPATIAL FEATURES AND PARAMETERS TO BE TESTED (SYMBOLS ARE ADOPTED FROM THE ORIGINAL PAPER)

Features	Parameters in the experiments	Original paper
PSI	$T_1 = 100$ $T_2 = 50$ $D = 20$	[6]
LWEA	$L_{max} = 50$ $TH_{ED} = 20$ $N_{dir} = 20$	[5]
GLCM	window size: 21×21; used feature: homogeneity, contrast, similarity, entropy	[11, 13]
DMP	Maximal radius of disk shaped element: 20	[17]
MSVSF	$T_h = 20$ $T_b = 100$	

the state-of-the-art spatial features. To evaluate the spatial features particularly, spatial and spectral features are fused with a simple vector concatenation method and trained by SVM classifier with RBF kernel (with $\gamma = 1$) as described in Section IV. The window size and pixel similarity threshold are set to be the same for all the spatial features in the comparative studies. To make the comparative studies more reliable, the following features are tested and the parameters of each feature are shown in Table II, the parameters are optimized to ensure that their LSA (this is mainly for PSI and LWEA features) are as similar as possible.

1) QuickBird Dataset of Urban Area in Bangladesh: The underdeveloped areas usually lack regular urban layout, where residential houses, industrial and business area mixed with each other, so that building scales vary a lot. The challenge of this dataset is that roofs with the similar spectral signatures connected together, forming large homogeneous areas, which are similar to open grounds.

Although the spatial feature alone may not produce usable results for real datasets, it is important to examine its sole performance on classifying a real dataset. Therefore, the experiment is first conducted with spatial feature only. Table III shows the OA of classification using the spatial feature alone. It can be seen MSVSF outperforms the compared spatial features,

TABLE III
OA OF SPATIAL FEATURES ONLY ON THE QUICKBIRD DATASET

Features	PSI	LWEA	GLCM	DMP	MSVSF	%
	12.43	29.21	17.31	19.33	36.32	

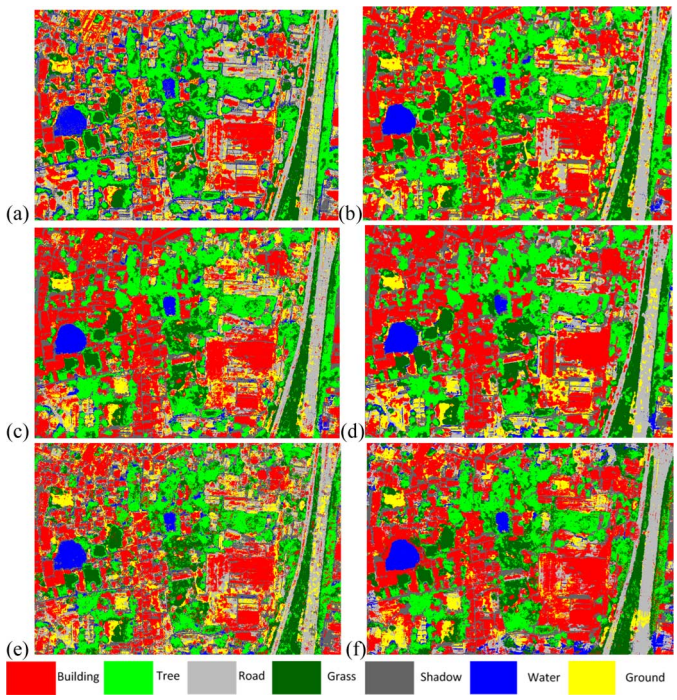

Fig. 8. Classification map of QuickBird dataset: (a) spectral feature ICA; (b) PSI + ICA; (c) LWEA + ICA; (d) ICA + GLCM; (e) DMP + ICA; and (f) ICA + MSVSF.

TABLE IV
CLASSIFICATION ACCURACY OF DIFFERENT SPATIAL FEATURES FUSED WITH ICA (QUICKBIRD DATASET) (%)

(%)	Spectral only	PSI	LWEA	GLCM	DMP	MSVSF
Building	56.66	68.98	70.23	69.09	62.14	73.73
Tree	73.32	78.66	79.95	82.39	73.90	80.36
Road	69.76	69.48	76.98	74.03	66.49	86.35
Grass	75.88	78.84	82.54	85.12	79.17	82.14
Shadow	74.78	91.57	90.26	87.99	81.12	93.67
Water	82.91	94.62	92.54	94.84	86.75	97.16
Ground	39.35	60.30	70.89	68.33	59.70	79.61
CV	70.31	76.73	79.92	78.63	70.94	84.31
OA	63.96	72.73	75.34	74.91	67.76	79.04

which proves its validity in recognizing the spatial patterns in the remotely sensed images.

The ICA components are then concatenated to the spatial feature vector, engendering a $(4 + p)$ feature vector, where p is the dimension of the spatial features. Results are shown in Fig. 8 and Table IV.

Fig. 8(a) shows that the spectral information alone produces poor results for such a challenging dataset, which mainly occur in building/road and the water/shadow area. By incorporating the spatial and spectral features, the accuracies increase to a notable level. From the classification result shown in Fig. 8(b)–(f), it can be seen that the misclassification occurs at

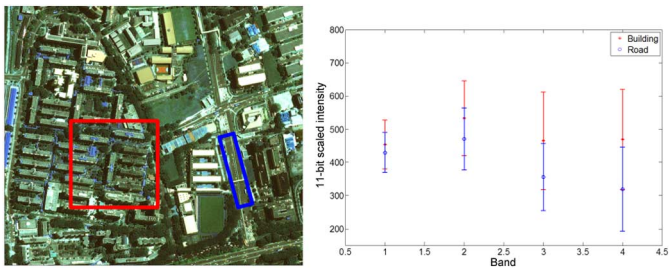

Fig. 9. Spectral response of the buildings and roads in the original 11-bit image. Left: the RGB view of the multispectral image, red rectangle indicates the buildings, and blue rectangle indicates the roads, the spectral range of building and road in different bands (band 1: Red; band 2: green; band 3: blue; and band 4: near infrared).

connected roofs, which has similar spectral response as roads. The 2-D shapes of these connected roofs are also similar to that of the roads and grounds, which makes them difficult to be distinguished by spatial features. This is also shown by the classification accuracy (Table IV) obtained from the comparison of different spatial features. Shape features extracted from the LSA can resolve most of the ambiguity between the spectrally similar classes. Table IV shows that among the spatial features tested in this experiment, MSVSF has obtained the best result in terms of OA, and this is due to the improvement of the classification accuracy of the building roofs and roads. The classification accuracy based on MSVSF of trees is slightly worse than that of the GLCM feature. For the rest of the classes, MSVSF significantly improves the classification accuracy, contributing to an OA of 84.31% and an improvement around 5%–8% than the current spatial features in this experiment. The PSI and LWEA do not perform well due to the large scale difference of the buildings, as they are scale-dependent features.

2) IKONOS Dataset: CBD Area in Singapore: The IKONOS dataset maps a CBD area in Singapore, and large buildings vary in scales and shapes. One of the challenges is that the large buildings with strip shapes (in red rectangle) in the left part of the image have very similar spectral response as the road (Fig. 9), where there are large overlaps between the spectral responses of building and roads in each band (Fig. 9, right). The joint spectral–spatial similarity between the building and road class creates the extreme cases to deviate the buildings and roads. Fig. 10 shows the classification results of the tested spatial features and Table V shows the classification accuracy. The classification result with the spectral feature shows that the OA is 73.74%, which is mainly contributed by the high accuracy on the grass and shadow, while most of the strip-shaped building roofs are misclassified as the roads and grounds. The road and building class can be separated by employing the spatial features, but for some of the strip-shaped buildings as marked in Fig. 9(a), the misclassification rate is still very high. Compared to the result using merely the spectral feature, PSI and LWEA have obtained higher OA, but with only slight improvement on the buildings. MSVSF gains 13% improvement of the buildings, 4.7% improvement of the roads, and 9% improvement of the ground, which contributes 7.8% improvement to the OA.

Fig. 10. Classification map of IKONOS dataset: (a) spectral feature ICA; (b) PSI + ICA; (c) LWEA + ICA; (d) ICA + GLCM; (e) DMP + ICA; and (f) ICA + MSVSF.

TABLE V
CLASSIFICATION ACCURACY USING DIFFERENT SPATIAL FEATURES FUSED WITH ICA (IKONOS DATASET)

(%)	Spectral only	PSI	LWEA	GLCM	DMP	MSVSF
Building	61.54	64.72	67.49	70.11	70.55	74.46
Tree	84.47	91.25	93.26	90.31	87.62	90.09
Road	70.29	74.66	73.98	73.43	64.30	74.97
Grass	94.87	94.32	94.92	95.13	93.73	94.23
Shadow	91.27	90.06	91.17	90.97	87.58	90.04
Ground	77.66	83.97	85.12	89.34	84.34	86.22
CV	80.42	82.75	82.59	82.32	78.59	83.07
OA	73.74	77.34	79.11	79.64	77.65	81.55

C. Parameter Analysis

There are several tunable parameters for computing the MSVSF features, which can be divided into two groups due to the mode of extracting spatial features: 1) parameters of MS vector calculation: the spectral bandwidth h_r and the maximal spatial bandwidth h_s (as a multiscale approach is used for the MS vector computation); 2) parameters of LSA extraction: T_b, T_h. As mentioned in Section III-A, h_r mainly concerns to what extent the spectral responses of two pixels can be evaluated as to be similar. Fig. 11(a) shows the relationship between the parameter h_r and the KC of the classification results on the reference data and CV on the training samples. It can be seen that these two curves are highly correlated. There are inconsistencies between these two curves, but not significant. Since the results on the reference data reflect the final classification,

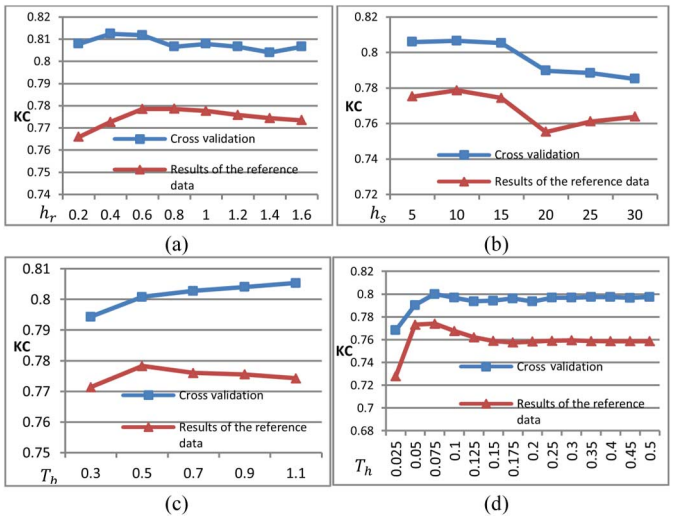

Fig. 11. Parameter analysis of MSVSF feature for the IKONOS dataset: (a) relationship between h_r and KC; (b) relationship between w and KC; (c) relationship between T_b and KC; and (d) relationship between T_h and KC (add the results of the CV for all the curves).

the following analysis is based on this curve. h_r varies from 0.2 to 1.6, with maximal difference of KC around 0.02, and it reaches the peak when $h_r = 0.6$. As the value of h_r increases after the peak, the MS vector is less sensitive to spectral difference, and the MS vectors of different pixels tend to be the same, which causes decreased KC. Fig. 11(b) shows that when h_s changes from 5 to 30, it reaches the peak at $h_s = 10$, and then decrease when w continuously increases until 20. This is because as h_s becomes larger, it may include redundant information that affects the orientation of the MS vectors, which causes a decrease of classification accuracy. When h_s change from 20 to 30, it increases slightly, which is contributed by the improved accuracy of the open ground. T_b and T_h controls the shape and for LSA extraction; a small T_b leads to a small LSA, which might be only a part of local homogeneous area, and a large T_b results in a large LSA. If T_b go beyond a certain limit, the searching process may go cross some spectral change. Fig. 11(c) shows that, for the IKONOS dataset, 0.5 is the optimal value for T_b, and KC slightly decreases as T_b increases from 0.5, when T_b is smaller than 0.3, the extracted LSA may be too small to depict the local shape difference. T_h is the acceptance threshold. In the region growing process, a small value leads to a small LSA. It is relatively robust to the final classification accuracy. Fig. 11(d) shows that a very small T_h (i.e., 0.025) will cause a significant decrease of KC, because the resulting small LSA does not contain enough information to distinguish between different classes. As T_h increases, the region growing process tends to accept more adjacent pixels into LSA, but the LSA extraction tends to be controlled by threshold T_b that constrains the sum of the spectral difference over the LSA. The choice of h_s is dependent on the resolution of the image and the other parameters can be estimated based on the noise level of the data. These parameters are relatively robust to the final classification results, and the CV has a high agreement to the final classification results. A trial-and-error approach may help

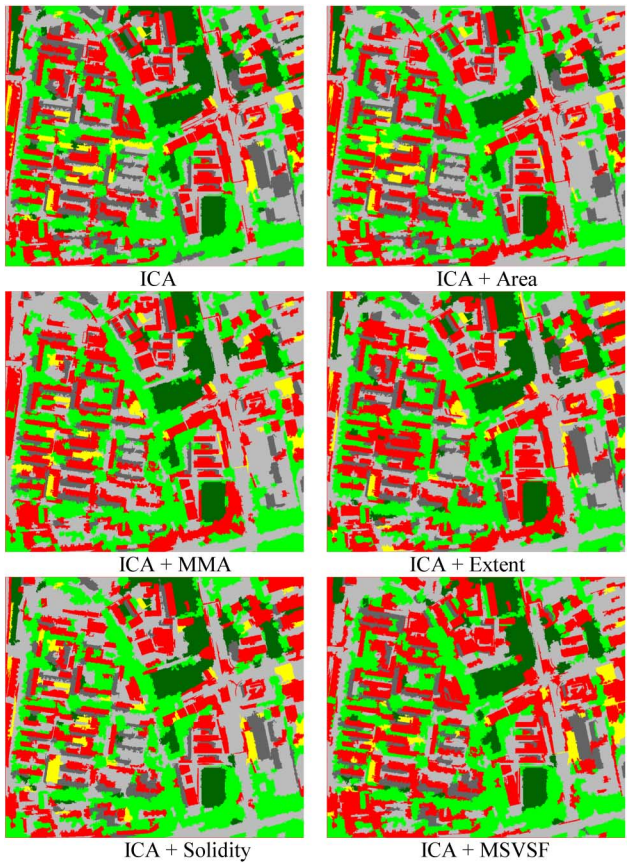

Fig. 12. OBC using different shape features (refer the color legend to Fig. 10).

to define the range of the parameter value, and the K-fold CV could be used to choose the optimal value for each parameter.

D. Extension of MSVSF on OBC

Per-pixel classification has the advantage of robustness, while object-based approaches can properly improve the classification accuracy [4], [9], as well as reduce the computational load for processing large volume of images [42], [43]. After segmenting the image into unclassified object primitives, the classification algorithms take these segments as the basic unit for labeling. Since the MSVSF feature extracts information based on the LSA, the MSVSF feature can be computed by treating each image segments as LSA. Therefore, to test the validity of the MSVSF feature, the classic mean-shift segmentation [24] is first applied on the dataset and then the converged spectral response of each segment is stored as their spectrum feature. By fusing these two features into SVM classifier, the classification result of the IKONOS dataset is presented in Fig. 12 and Table VI. Since the segmentation technique is still in research phase, adopting the segmentation algorithms on the complex roof textures and road patterns may not always generate meaningful shapes [44], thus spatial features extracted merely from segments may cause erroneous classification. The advantage of the MSVSF is that MS vectors are extracted from the original image, and it introduces information not only in the LSA, but beyond the LSA. The result of the OBC with MSVSF is compared to several typical shape measurements:

TABLE VI
OBC ACCURACY WITH DIFFERENT SHAPE MEASUREMENT

(%)	Spectral only	Area	MMA	Extent	Solidity	MSVSF
Building	70.62	70.26	74.38	79.04	73.61	80.86
Tree	96.85	80.99	80.62	80.46	94.71	95.14
Road	78.99	89.97	90.53	84.83	88.37	88.52
Grass	98.68	97.68	97.52	98.68	97.68	97.69
Shadow	89.88	89.50	68.81	89.36	82.05	91.79
Ground	66.88	66.88	74.89	69.40	76.17	90.50
CV	78.38	64.34	64.87	79.14	79.83	81.79
OA	82.17	79.58	80.99	82.95	84.06	87.99

1) area of the segments (number of pixels), which is the direct extension of PSI; 2) the length of the major and minor axes of the segments (MMA), which is the direct extension of LWEA; 3) extent; and 4) solidity, which are computed as

$$\text{Extent} = \frac{\text{Area}}{\text{Major Axis Length} \times \text{Minor Axis Length}}$$
$$\text{Solidity} = \frac{\text{Area}}{\text{Area of Convex hull}}. \quad (17)$$

Tables VI shows that the MSVSF in OBC gains more than 6% improvement in OA compared to using spectral feature alone. It shows that the Area and MMA feature even result in a loss of classification accuracy compared to the result of using spectral response alone. Therefore, inappropriate spatial features on the fragmental and the irregular segments will lead to a negative effect. Extent, Solidity, and MSVSF contribute positive effects to the classification results. Among these tested features, MSVSF achieves better results, which proves the proposed MSVSF feature is among the state-of-the art spatial measurements for OBC.

VI. CONCLUSION

In this study, a novel spatial feature MSVSF is proposed. It is based on the MS vector and measures the 2-D deformation of the LSA imposed by MS vector. SVM classifier with Gaussian Radial basis function was adopted to classify four-band VHR images with the proposed feature. A synthetic experiment and two experiments on real dataset have been conducted and the proposed MSVSF feature has achieved satisfactory results in terms of accuracy and robustness. The contribution of this study lies in the following aspects.

1) The MSVSF feature extraction is a novel combination of two traditional modes of extracting spatial features: 1) feature extraction from pre-defined window; 2) topological feature extraction from LSA, which acts as a tradeoff between the pros and cons of the two modes.
2) The MSVSF measures the 2-D deformation of LSA instead of the real shape of the LSA, such that it is relatively robust to LSAs that are not well extracted due to inappropriate threshold.
3) One synthetic experiment and two experiments on real dataset under different urban scenarios are conducted: 1) urban area of under-developed country; 2) CBD area of developed country. The synthetic experiment shows that MSVSF outperforms similar spatial features in classifying different 2-D shapes with different scales. Results in the real experiment show that the MSVSF feature gains 7%–16% improvement of OA compared to results using the spectral information alone. Significant improvements are obtained to discriminate among the classes of the impervious surface.
4) MSVSF is extended to the OBC approach. The experiment shows better performance compared to some of the state-of-the-art features.

The parameters of MSVSF are discussed, it turns out the most sensitive parameter is the spatial bandwidth h_s for computing MS vector, which is dependent on the resolution of data. A value of 10 achieved the highest KC for images with such high spatial resolution with the tested dataset. Compared with the window size, T_b, T_h, and h_r are more stable and can be easily applied to other dataset.

The proposed MSVSF feature is effective for classifying the VHR imagery. However, in this paper, a vector concatenation with simple normalization approach is used to integrate the spectral feature and spatial feature, and this is for simplifying the comparison of different features. It is understood that such an approach is not an optimal solution for fusing the spectral and spatial features [4], and more appropriate integration methods will be investigated to improve the classification accuracy with MSVSF. Moreover, the elongation feature in MSVSF is used for improving the robustness of the proposed method, while it may not be sufficient to support the urban objects with concave and more complex shapes, thus more features in this respect are also planned to improve the proposed feature.

The 2-D deformation measurement is only one property of the MS vector, more properties such as the histogram of the MS vectors and the pattern regularity of MS vectors flows may be useful for the VHR image classification. In the future work, the author plans to explore more properties of the MS vector.

ACKNOWLEDGMENT

The authors would like to thank the Global Land Cover Facility (GLCF) in University of Maryland for providing the QuickBird image and DigitalGlobe for providing the IKONOS image for the experiments. They would also like to thank the anonymous reviewers for their precious comments and suggestions.

REFERENCES

[1] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multisource composite kernels for urban-image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 88–92, Jan. 2010.
[2] I. Dópido *et al.*, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.
[3] D. B. Hester, H. I. Cakir, S. A. Nelson, and S. Khorram, "Per-pixel classification of high spatial resolution satellite imagery for urban land-cover mapping," *Photogramm. Eng. Remote Sens.*, vol. 74, p. 463, 2008.
[4] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[5] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1920–1932, Sep. 2003.

[6] 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.

[7] W. Y. Yan, A. Shaker, and W. Zou, "Panchromatic IKONOS image classification using wavelet based features," in *Proc. IEEE Toronto Int. Conf. Sci. Technol. Humanity (TIC-STH)*, 2009, pp. 456–461.

[8] B. Johnson and Z. Xie, "Classifying a high resolution image of an urban area using super-object information," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 40–49, 2013.

[9] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354–2363, Oct. 2003.

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

[11] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[12] Y. Zhang, "Optimisation of building detection in satellite images by combining multispectral classification and texture filtering," *ISPRS J. Photogramm. Remote Sens.*, vol. 54, pp. 50–60, 1999.

[13] 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, pp. 1276–1292, 2009.

[14] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[15] 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, pp. 721–732, 2011.

[16] R. Qin and W. Fang, "A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization," *Photogramm. Eng. Remote Sens.*, vol. 80, pp. 37–48, 2014.

[17] 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.

[18] 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.

[19] X. Huang, L. Zhang, and P. Li, "Classification of very high spatial resolution imagery based on the fusion of edge and multispectral information," *Photogramm. Eng. Remote Sens.*, vol. 74, pp. 1585–1596, 2008.

[20] 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, pp. 6219–6239, 2009.

[21] X. Huang and L. Zhang, "A multiscale urban complexity index based on 3D wavelet transform for spectral–spatial feature extraction and classification: An evaluation on the 8-channel WorldView-2 imagery," *Int. J. Remote Sens.*, vol. 33, pp. 2641–2656, 2012.

[22] R. Qin, J. Gong, H. Li, and X. Huang, "A coarse elevation map-based registration method for super-resolution of three-line scanner images," *Photogramm. Eng. Remote Sens.*, vol. 79, pp. 717–730, 2013.

[23] R. Qin, J. Gong, and C. Fan, "Multi-frame Image super-resolution based on knife-edges," in *Proc. IEEE Int. Conf. Signal Process. (ICSP)*, Beijing, China, 2010, pp. 972–975.

[24] D. Comaniciu 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.

[25] H. Zhou, G. Schaefer, M. E. Celebi, F. Lin, and T. Liu, "Gradient vector flow with mean shift for skin lesion segmentation," *Comput. Med. Imag. Graphics*, vol. 35, pp. 121–127, 2011.

[26] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, Jan. 1975.

[27] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.

[28] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.

[29] R. Qin, "Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 96, pp. 179–192, 2014.

[30] 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.

[31] 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.

[32] 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.

[33] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[34] L. Wang, *Support Vector Machines: Theory and Applications*, vol. 177. New York, NY, USA: Springer, 2005.

[35] E. Pagot and M. Pesaresi, "Systematic study of the urban postconflict change classification performance using spectral and structural features in a support vector machine," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 1, no. 2, pp. 120–128, Jun. 2008.

[36] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[37] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, pp. 121–167, 1998.

[38] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[39] 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.

[40] 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.

[41] J. Milgram, M. Cheriet, and R. Sabourin, ""One against one" or "one against all": Which one is better for handwriting recognition with SVMs?," in *Proc. 10th Int. Workshop Front. Handwriting Recognit.*, 2006 [Online]. Available: http://hal.archives-ouvertes.fr/docs/00/10/39/55/PDF/cr102875872670.pdf

[42] R. Qin and A. Gruen, "3D change detection at street level using mobile laser scanning point clouds and terrestrial images," *ISPRS J. Photogramm. Remote Sens.*, vol. 90, pp. 23–35, 2014.

[43] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, pp. 2–16, 2010.

[44] R. Qin, "An object-based hierarchical method for change detection using unmanned aerial vehicle images," *Remote Sens.*, vol. 6, pp. 7911–7932, 2014.

Rongjun Qin (S'14) received the B.S. degree in computational mathematics from Wuhan University, Wuhan, China, in 2009, and the M.S. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMRS), Wuhan University, in 2011. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing at ETH, Zurich, Switzerland.

Since 2011, he has been a Researcher with Singapore ETH Centre, Future Cities Laboratory, Singapore, under the Simulation Platform Module. His research interests include remote sensing image classification, UAV images processing, image dense matching, 3-D modeling, and change detection.

Mr. Qin serves as a Reviewer for several international journals including *ISPRS Journal of Photogrammetry and Remote Sensing and Photogrammetric Engineering and Remote Sensing*. His rewards include the First Prize of National Mathematic Modeling Competition in China, and several scholarship awards.