



A probabilistic framework for weighted combination of multiple-feature classifications of hyperspectral images

Reza Seifi Majdar¹ · Hassan Ghassemian²

Received: 6 July 2019 / Accepted: 27 August 2019 / Published online: 12 September 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Spatial information such as texture and shape features as well as spatial contextual information play a key role in representation and analysis of hyperspectral images. Spatial information improves the classification accuracy and addresses the common problem of pixel-wise classification methods, i.e. limited training samples. In this article, a new combination of spectral, texture and shape features, as well as, contextual information in the probabilistic framework is proposed. The texture features are extracted utilizing Gabor filters and the shape features are represented by morphological profiles. The spectral, texture and shape features are separately fed into a probabilistic support vector machine classifier to estimate the per-pixel probability. These probabilities are combined together to calculate the total probability on which three weights determine the efficacy of each one. Finally, the classification result obtained in the previous step is refined by majority voting within the shape adaptive neighbourhood of each pixel. Instead of the simple majority vote we applied the majority vote in the probabilistic framework on which the reliability of the labels in the region is also considered. Experiments on three hyperspectral images: Indian Pines, Pavia University, and Salinas demonstrate the efficiency of the proposed method for the classification of hyperspectral images, especially with limited training samples. Moreover, after comparing with some recent spectral–spatial classification methods, the performance of the proposed method is demonstrated.

Keywords Hyperspectral image · Spectral-spatial classification · Texture features · Shape features · Probabilistic SVM

Introduction

Because of the key role of the classification in hyperspectral imaging, there have recently done many efforts to design more accurate classifiers (Camps-Valls et al. 2014; Canty 2014). The support vector machine (SVM) classifier is one of the best classifiers which has shown good performance in hyperspectral image classification even with limited reference

data (vapnik 2000). However, in SVM and also other pixel-wise classifiers the information of each pixel is only used for classification on which the abundant spatial information of the hyperspectral, especially high resolution, images and the relationship between neighbouring pixels is not considered. Therefore, the salt-and-pepper error appears in the classification map (Fauvel et al. 2013). Furthermore, recent studies have shown that spatial information can be used to achieve better classification accuracy (Benediktsson and Ghamisi 2015). In addition to the spectral features, the texture, shape and size features can be applied to distinguish the pixels belonging to different classes better (Khodadadzadeh et al. 2014; Golipour et al. 2016; Wang et al. 2016a, b). An improved SVM classifier with multiple kernels, based on the spectral and spatial features, is proposed by Wang and Duan 2018. Since, the spectral kernel is constructed through each pixel's spectral features, and the spatial kernel is modeled by using the extended morphological profile method. In some works, Markov Random Field (MRF) has been used to model the spatial correlation between adjacent pixels. MRF intuitively represents the spatial information which used for

Communicated by: H. Babaie

✉ Reza Seifi Majdar
r.seifi.m@iauardabil.ac.ir

Hassan Ghassemian
ghassemi@modares.ac.ir

¹ Department of Electrical Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran

² Image Processing and Information Analysis Lab, Faculty of Computer and Electrical Engineering, Tarbiat Modares University, Tehran, Iran

classification. In these works, the classification results are refined using spatial contextual information and by minimizing a suitable energy function (solberg et al. 1996; Tarabalka et al. 2010a; Zhang et al. 2011; Wang et al. 2016a, b). In another work (Wang and Duan 2019), the authors proposed a combination of the techniques of algebraic multigrid (AMG), hierarchical segmentation (HSEG) and MRF for spectral-spatial classification of hyperspectral images. In this method, a novel segmentation method is developed by combining the AMG-based marker selection approach and the conventional HSEG algorithm to construct a set of unsupervised segmentation maps in multiple scales. On the other hand, an improved MRF energy function is proposed for multiscale information fusion. Some other represented works are based on the features extraction. Extended attribute profiles (EAP) and extended multi-attribute profiles (EMAP) have been used for extracting spatial features and modelling spatial information (Dalla Mura et al. 2010). The spatial features are generated based on the morphological attribute filters and multilevel analysis. An extinction profile (EP) is an impressive and recent feature extraction (FE) method for hyperspectral images (Fang et al. 2018a). In FE by EP, the spatial information and the geometrical characteristics are represented better than the previous spectral-special methods. In Fang et al. (2018b) a novel local covariance matrix representation method is represented on which the correlation between different spectral bands and the spatial-contextual information is described when conducting FE from hyperspectral images.

Another important category of spectral-spatial classification methods is the object-based classification (Li and Wan 2015; Zehtabian and Ghassemian 2015; Fang et al. 2015). After extracting objects from the image, a feature or a set of features assigned to each object. A classifier is applied to classify the image, object by object. All pixels of an object share the same label assigned to that object.

The other well-known schemes of this category are those which utilize texture features. Texture is one of the main features of the image which reveals the spatial information. Gabor filters, gray-level-co-occurrence matrices (GLCMs) and wavelet transformation are three main methods used to extract texture features (Guo et al. 2014; Wang et al. 2016a, b; Mirmehdi 2008). In Huo and Tang (2011) the spectral features and the extracted texture features are stacked in one feature vector and used for classification. Extended expression of the stacked features approach can be found in Mirzapour and Ghassemian (2015). Various combinations of spectral, texture and shape features are stacked together. Four different types of spectral features, i.e. original spectral data, features extracted by linear discriminant analysis (LDA); principal component analysis (PCA); and non-parametric weighted feature extraction (NWFE) (Jia et al. 2013) were considered as spectral part of stacked features. More than 50

different combinations were evaluated on three hyperspectral data using SVM classifier and the best combination of each data set was reported. In Fang et al. (2017) the spectral, texture and shape features are combined by adaptive sparse representation (MFASR) method. The sparse coefficients of these multiple-feature matrix are used to determine the label of each pixel.

Multi-scale superpixel-based spectral-spatial classification (Li et al. 2016) is another novel technique for spectral-spatial classification. Superpixel-based classification and segmentation are performed in each scale on which the classification accuracy is improved using the segmentation results. Finally, the multi-scale classification result is obtained by the majority voting.

Some other widely used methods are based on the combination of the pixel-wise classification with segmentation or clustering techniques. Tarabalka et al. (2010b) used an SVM classifier for pixel-wise classification and a watershed transformation to generate segmentation map. The improved classification results were obtained by combining classification and segmentation maps through majority voting approach within the watershed regions. Seifi Majdar and Ghassemian (2017a) investigated the efficiency of this method in functional data analysis framework. A three steps spectral-spatial classification method based on joint bilateral filtering and graph cut segmentation is proposed by Wang et al. 2016a, b in which the regions obtained by the spectral-spatial segmentation are properly labelled.

The prominent potential of the Neural network (NN) for hyperspectral image classification has recently proved. In conventional NNs the typical pooling layers are fixed and cannot be changed adaptively for feature downsampling. On the other hand, the sampling locations of traditional convolutional kernels cannot be changed based on the complex spatial structures in hyperspectral images. Zhu et al. (2018) proposed a deformable CNN-based hyperspectral image classification method in which a deformable convolutional sampling locations with adaptive size and shape can be adjusted according to the spatial information. In Fang et al. (2018c), a squeeze multibias network (SMBN) is proposed for hyperspectral image classification. In this method, a multibias module is placed behind the convolutional layer to decouple feature maps to multiple maps according to the magnitudes of responses. Then, the combination of the response maps by the subsequent layer is used to reach a better classification accuracy.

Integrating of the spectral data and the spatial information in the probabilistic framework (Liu and Lu 2016) is one of the newest spectral-spatial classification approaches. The per-pixel probability is separately estimated for spectral data and spatial information and then the joint probability is obtained using these two probabilities. We can find outstanding results in this method, but still remained some issues to be addressed. This method only applied in fixed-size window and is not

considered the border effects between regions. In another work, a spectral-spatial classification method was proposed in probabilistic framework. At first, spectral and texture features were separately classified by a probabilistic SVM classifier to estimate the per-pixel probability. Then, the total probability was calculated by a linear combination of the previous probabilities for each pixel. Finally, one pixel was assigned to a class with maximum total probability (Seifi Majdar and Ghassemian 2017b). Texture features is the only spatial information which applied in this method to enhance the classification results and it had not been used of the shape features and contextual information.

To address the above issues and to enhance the classification results further, a new combination of spectral, texture and shape features, as well as, contextual information in the probabilistic framework is proposed to improve the classification of the hyperspectral images, especially with limited training samples. Texture features are extracted by Gabor filters and the shape features are represented by MPs. The spectral and spatial features are separately fed into a probabilistic support vector machine (SVM) classifier to estimate the per-pixel probability. These probabilities are combined together to calculate the total probability, on which three weights determine the efficacy of each one. Finally, the classification result obtained in the previous step is refined by majority voting within the shape adaptive (SA) neighbourhood of each pixel. Instead of the simple majority vote, we applied the majority vote in the probabilistic framework on which the reliability of the labels in the region is also considered. The main contributions of this article contain three-folds: 1) weighted combination of the spectral, texture, shape features and the contextual information in probabilistic framework is the main contribution of this article that didn't appear in the previous works in which all extracted spatial information are combined with the spectral data by using probability distribution functions, 2) Unlike some spectral-spatial classification methods which use of the fixed-window neighbourhood, in this article, the contextual information in a SA neighbourhood is used to enhance the classification results, 3) in the previous works, the number of the similar and dissimilar labels in the neighbourhood is used to refine the label of the central pixel, but the reliability of the neighbourhood labels is not considered. In the proposed method, the number of labels and the reliability of that labels are considered to determine the label of the central in a SA neighbourhood. The remaining parts of this article are organized as follows: In the next section, the mentioned methods to extract the texture and the shape features are introduced briefly. In the Proposed classification method section, we explain the proposed method. The hyperspectral data sets and the experimental results are represented in the Experimental results and discussions section, and the conclusion is drawn in Conclusion section.

Main components of the proposed method

In this study we exploit spatial information using Gabor filters as texture features and MP as shape features. In the following subsections, we introduce them briefly.

Texture feature extraction using Gabor filters

Gabor filters have widely been conducted in image processing, pattern recognition and computer vision (Mirmehdi 2008). They can provide accurate time-frequency location and robust against contrast and brightness of the images. Two-dimensional Gabor filters are used to extract texture features. Daugman (1985) extended the Gabor filters in two-dimensions for the first time. A two-dimensional Gabor wavelet consists of a complex plane wave modulated by an elliptical Gaussian envelope (Zhang et al. 2012). 2-D Gabor function can be represented as follows:

$$G_{s,d}(x,y) = G_{\vec{k}}(\vec{x}) = \frac{\|\vec{k}\|}{\delta^2} \cdot e^{-\frac{\|\vec{k}\|^2 - \|\vec{x}\|^2}{2\delta^2}} \cdot \left[e^{i \cdot \vec{k} \cdot \vec{x}} - e^{-\frac{\delta^2}{2}} \right],$$

$$\vec{k} = (\pi/2f^s) \cdot e^{i(\frac{\pi d}{8})}, f = 2. \tag{1}$$

Where (x,y) implies the spatial position of a pixel in the image, \vec{k} is the frequency vector, d and s are direction and scale respectively. The number of oscillation under Gaussian envelope is determined by $\delta = 2\pi$. The convolution of the image I with the Gabor function in a specific scale and direction is performed to generate the Gabor texture features:

$$F_{s,d}(x,y) = G_{s,d}(x,y) * I(x,y) \tag{2}$$

The Gabor texture features of one pixel are given as follows:

$$v_{texture}(x,y) = [F_{1,1}(x,y), \dots, F_{s,d}(x,y)] \in R^{sd} \tag{3}$$

Shape features representation by morphological profiles

Mathematical morphology, developed by Matheron (1975); Serra (1983), is a powerful methodology which shows some operators to extract image components used to represent the region shape. Erosion and dilation, two fundamental morphological operations, act on an image by using structuring element (SE) of arbitrary size and shape to evaluate the geometrical structures of the image. More complex morphological operations, opening and closing, are obtained by the combination of erosion and dilation. The morphological opening is defined as an erosion of an image followed by a dilation. Moreover, the morphological closing has an opposite

definition of the opening operation which is defined as a dilation, followed by an erosion operation. Reconstruction is a morphological transformation which tends to restore the shape of the objects that remain after morphological opening or closing. When opening by reconstruction or closing by reconstruction operators act on an image, preserve the objects that can contain SE, while the other objects are completely removed (Soille 2013). Applying these operators on a grayscale image \mathbf{I} with m disk-shaped SE of radius $\lambda \in \{1, 2, \dots, m\}$, lead to a stack of $(2n + 1)$ features,

$$\begin{aligned} \text{MP}(\mathbf{I}) &= \Pi_i \\ &: \left\langle \begin{array}{l} \Pi_i = \Pi_{\phi_\lambda} \text{ with } \lambda = (m-1 + i), \forall \lambda \in [1, m] \\ \Pi_i = \Pi_{\gamma_\lambda} \text{ with } \lambda = (i-m-1), \forall \lambda \in [m+1, 2m+1] \end{array} \right\rangle \end{aligned} \quad (4)$$

Where Π_{ϕ_λ} and Π_{γ_λ} are closing and opening by reconstruction operators with a disk-shaped SE of radius λ , respectively.

Probabilistic SVM

SVM is a versatile method for the classification of hyperspectral images with linear or nonlinear models. The nonlinear SVM model, using kernel functions, provides better classification accuracies rather than the linear model (Camps-Valls and Bruzzone 2005). In this article, all of the spectral and spatial features are fed into the multiclass one-versus-one SVM classifier with a polynomial kernel, $K(x, y) = (\gamma x^T y + r_0)^d$. For simplification and fair comparisons, the default values are considered as kernel parameters, $d = 3$, $\gamma = 1/\text{number of features}$, and $r_0 = 0$. The SVM classifier cannot directly generate the estimation of probability. Therefore, different techniques are used to estimate per-pixel probability by combining all pair-wise comparisons (Wu et al. 2004). The popular LIBSVM library is employed to implement the SVM classifier and to estimate per-pixel probability (Chang and Lin 2011).

Proposed classification method

The flowchart of the proposed classification method is shown in Fig. 1. We have d -band hyperspectral image $\mathbf{X} = \{x_i \in R^d, i = 1, 2, \dots, n\}$; n is the number of pixels. In this method, at first, Gabor and MP features are extracted from the first principal components (PC) of the original data (Zhang et al. 2012; Mirzapour and Ghassemian 2015). In the second step, the spatial features, Gabor and shape features, as well as spectral data are separately classified using the probabilistic SVM classifier. Therefore, three per-pixel probabilities are obtained for each pixel of the image.

Suppose there are K different classes $\{w_1, w_2, \dots, w_K\}$ in the image. The estimation of the probability based on the spectral, texture and shape features are computed as follows:

$$p^{\text{spc}}(x_i) = \{p_k^{\text{spc}}(x_i) = p^{\text{spc}}(y = k|x_i), k = 1, 2, \dots, K; i = 1, 2, \dots, n\} \quad (5)$$

$$p^{\text{txt}}(x_i) = \{p_k^{\text{txt}}(x_i) = p^{\text{txt}}(y = k|x_i), k = 1, 2, \dots, K; i = 1, 2, \dots, n\} \quad (6)$$

$$p^{\text{shp}}(x_i) = \{p_k^{\text{shp}}(x_i) = p^{\text{shp}}(y = k|x_i), k = 1, 2, \dots, K; i = 1, 2, \dots, n\} \quad (7)$$

Where ‘spc’, ‘txt’, and ‘shp’ are the abbreviations of spectral, texture, and shape, respectively.

Now, three probability distributions are obtained, i.e., the spectral distribution calculated from the spectral data, the spatial distributions computed from the texture and shape features. In the third step, the spectral, texture and shape probability distributions are combined together on which three positive weights, $\omega = \{\omega_1, \omega_2, \omega_3\}$; $\omega_j > 0$, determine the efficacy of each one. It means that for each feature, the larger ω_j represents the more important role of that feature in the combination. Therefore, total probability distributions defined as follows:

$$\begin{aligned} p^{\text{total}}(x_i) &= \omega_1 p^{\text{spc}}(x_i) + \omega_2 p^{\text{txt}}(x_i) + \omega_3 p^{\text{shp}}(x_i) \\ &= \sum_{j=1}^3 \omega_j p^{\text{feature}_j}(x_i) \end{aligned} \quad (8)$$

Where $\text{feature} = \{\text{spc}, \text{txt}, \text{shp}\}$ and also we have the following constraint

$$\sum_{j=1}^3 \omega_j = 1, \quad \omega_j > 0 \quad (9)$$

In the final step, the spatial contextual information is applied to enhance the classification performance. For this purpose, Instead of defining a fixed-size window, a shape adaptive (SA) window is assumed for each pixel x_i (Fu et al. 2016). We know that the label of central pixel x_i can be modified by two factors of the neighbouring pixels. The first one is the inter-pixel class dependence assumption on which one pixel with a specific label has tendency to have neighbouring pixels with the same label (Tarabalka et al. 2010b). The second one is the classification reliability of the neighbouring pixels in which the neighbouring pixels with higher reliability have more influence on the label of pixel x_i (Negri et al. 2014). Negri et al. (2014) determined the classification reliability of one-pixel x as follows:

$$f_{\text{SVM}}(\mathbf{x}) = \langle w, \mathbf{x} \rangle + b \quad (10)$$

Where, w is the orthogonal vector to the separating hyperplane and $\frac{b}{\|w\|}$ determines the offset of the hyperplane from the origin along w . The notations \langle, \rangle and $\| \cdot \|$ represents the inner product and the vector normal, respectively. In a given neighbourhood of a pixel x_i , the influence of each neighbouring pixel is proportional to the reliability of that pixel.

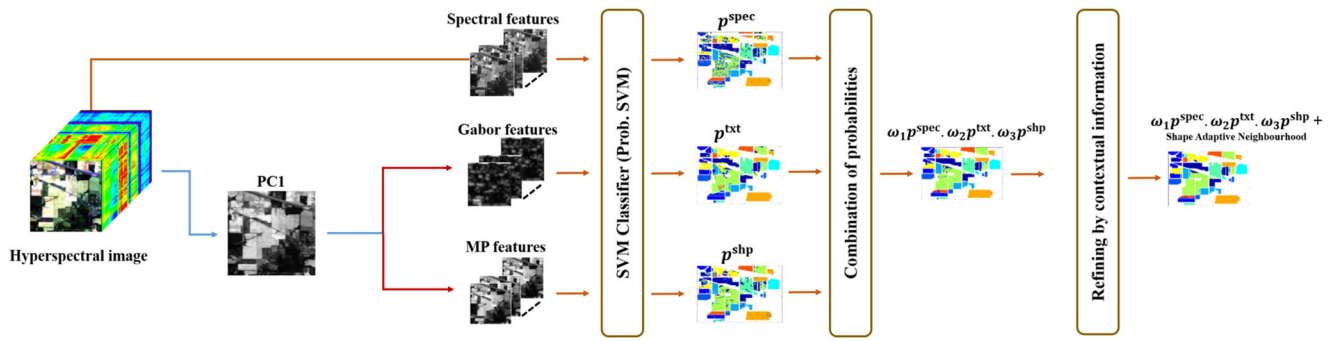


Fig. 1 Flowchart of the proposed classification method

In our proposed method, the classification reliability of the pixels is determined by per-pixel probability which obtained previously. Accordingly, in a specific class, the classification reliability of a pixel with probability near one is high and the classification reliability of a pixel with probability near zero is low. For instance, Fig. 2 shows the per-pixel probability of the pixels x_1 and x_2 , we supposed that there are only two different classes in the sense. As shown in this figure the probability of belonging pixel x_1 to class1 is high but there is a lower probability of belonging pixel x_2 to class1. As a result, it can be said that the classification reliability of pixel x_1 is higher than the classification reliability of pixel x_2 in class1. Accordingly, we can say with high reliability that the red pixel belongs to the class1 while for blue pixel the degree of reliability is low.

The general idea of applying spatial contextual information to regularize the initial classification maps are given in Fig. 3. In this procedure, at first, a neighbourhood with SA is considered for each pixel x_i . Then the total probability of pixel x_i and its neighbourhood, which obtained in the previous step, are summed together:

$$p^{SA}(x_i) = p^{total}(x_i) \tag{11}$$

Where N_i is the set of neighbouring pixels of a given pixel x_i . Finally, the pixel x_i is assigned to the class with the highest

probability.

$$label(x_i) = \operatorname{argmax}_{k=1}^K p^{SA}(x_{ik}) \tag{12}$$

It should be noted that this procedure occurs when there is at least one pixel with the different label in the neighbouring pixels.

Our proposed method to use of the contextual information of the neighbouring pixels is the same as the majority voting model but it has a main advantage. The simple majority voting model is just based on the inter-pixel class dependence assumption in which all pixels within a region are assigned to the most frequent label of that region and the reliability of the neighbourhood pixels are not considered. On the other hand, in our proposed majority voting in the probabilistic framework, the regularization of the classification map is performed based on the number of labels and the reliability of the neighbourhood patterns. The former indicates that one pixel with a specific label has tendency to have neighbouring pixels with the same label, while the latter means that the neighbouring pixels with higher classification reliability have more influence on refining the classification maps. When the probabilities of one pixel and its neighbouring pixels collected together, these two assumptions considered simultaneously.

The mentioned details of the proposed spectral-spatial classification method can be found in Fig. 1. The proposed

Fig. 2 Example of using probability estimates of one pixel to determine the classification reliability



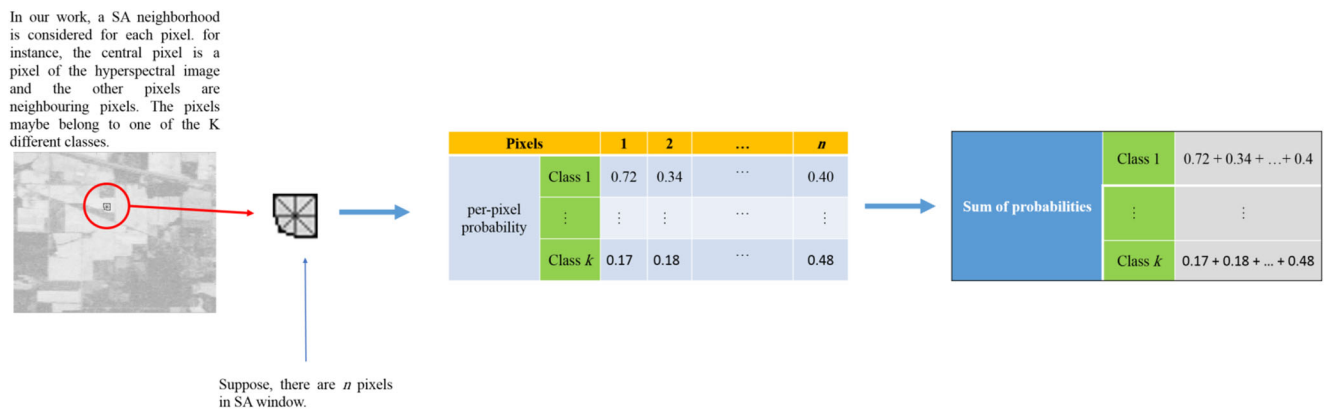


Fig. 3 The procedure of applying spatial contextual information to refine the classification maps (the numbers in the tables were arbitrarily selected)

method for each data set with different scheme of training samples will be introduced in the next section. The only drawback of the proposed method is the best selection of weights ω_j correspond to each feature. In the proposed method, the best set of these weights are derived experimentally. The effects of the weights on the classification results are given in the next section.

Experimental results and discussions

In this article, three hyperspectral data sets are used to evaluate the proposed method and to compare it with some recent spectral–spatial classification methods: (1) Indian Pines data set has a low spatial resolution image and it contains agricultural and forest land covers which gathered by AVIRIS sensor over the Indian Pines test site. (2) Pavia University data set was captured by the ROSIS sensor over Pavia with a high spatial resolution and it contains urban structures. (3) Salinas data set contains agricultural land cover with a high spatial resolution which were acquired by the AVIRIS sensor over Salinas Valley.

Hyperspectral data sets

(1) *Indian Pines data set*: Indian Pines data set is a high spectral resolution hyperspectral image but it has a low spatial resolution (20 m). It relates to non-urban, agricultural/forest land covers which captured by the AVIRIS sensor over North-western Indiana in June 1992. It has 145×145 pixels and 224 spectral bands on which 24 water absorption and noisy bands (104–108, 150–163, 220) were excluded and the remaining 200 bands are used in the experiment. This image contains 16 different classes in which six classes with the small number of samples are removed and ten classes will be applied for the experiments. The false-colour and the ground truth image are shown in Fig. 4.

- (2) *Pavia University data set*: Pavia University data set which has a very high spatial resolution (1.3 m) acquired by the ROSIS-3 sensor over the urban area of Pavia University, northern Italy. It has made of 610×340 pixels, and 115 spectral bands in which 12 most noisy bands were omitted and the remaining 103 spectral bands were used in our experiment. The false-colour and the ground truth image are shown in Fig. 5.
- (3) *Salinas data set*: Salinas data set is a high spectral resolution and also it has a high spatial resolution (3.7 m) which captured by AVIRIS sensor over Salinas Valley, California. This data set comprises 512×217 pixels, 16 classes, and 224 bands, while 20 water absorption bands (108–112, 154–167 and 224) were discarded. The false-colour and the ground truth image are shown in Fig. 6.

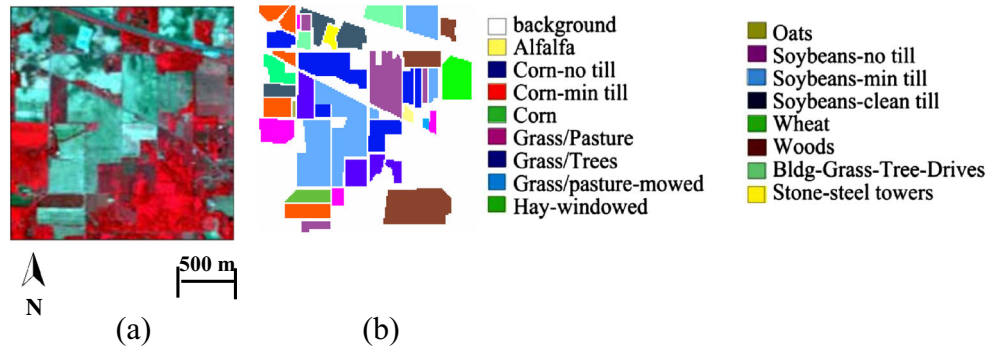
General description

In order to evaluate the performance of the proposed method, it is applied to three data sets described in advance. In all data sets, the texture and shape features are extracted by Gabor filters and MP, respectively.

In Gabor function, the direction and scale parameters are $d=12$ and $s=5$ (Zhang et al. 2012). Therefore, $5 \times 12 = 60$ Gabor filters are applied on the first PC. Accordingly, the texture features for one pixel of the hyperspectral image are represented by 60 features, $\mathbf{v}_{\text{texture}} \in R^{60}$. To extract MP features, the first PC is used. The parameter m is arbitrarily chosen 25, therefore, the disk-shaped SE of radius $\lambda \in \{1, 2, \dots, 25\}$, leads to a stack of 51 features (Mirzapour and Ghassemian 2015). It should be noted that, we used of first PC because it contains as much of the variability in the image as possible and has the most spatial information.

The probabilistic SVM with a polynomial kernel, $K(x, y) = (\gamma x^T y + r_0)^d$, is used in all classifications. The default values are considered as kernel parameters,

Fig. 4 Indian Pines image: **a** false-colour image **b** 16-class ground truth



$d = 3$, $\gamma = \frac{1}{\text{number of features}}$, and $r_0 = 0$. In all experiments the training samples of each class are randomly selected from the entire scene to train the classifier, however, the remaining samples are used as testing samples. Each experiment is run ten times with ten different sets of training samples and the average results are reported. Three schemes for the number of training samples is considered in all experiments which called set A, set B and set C respectively: (1) ill-posed classification problem i.e. $n_i = 10 < N < d$; (2) poorly-posed classification problem i.e. $n_i = 50 < d < N$; (3) a proportional scheme i.e. we select 1% of samples from each class as training set; where n_i is the number of training samples of class i , N is the number of all training samples and d is the number of spectral bands of hyperspectral image. All schemes relate to the limited training samples problems. It should be noted that we consider type 1 and 2 for Indian Pines and Pavia University data sets and type 3 only for Salinas data set in order to make fair comparison between the proposed method and the other spectral–spatial methods. The number of training samples in each scheme for three data sets are shown in Table 1.

Three measures of accuracy were used: overall accuracy (OA), average accuracy (AA) (Kianisarkaleh and Ghassemian 2016)) and kappa coefficient (κ) (Cohen 1960). The experiments are conducted on a computer with an Intel Pentium (R) Dual-Core 2.2 GHz processor and 4 GB RAM running Windows 7 64Bit operating system. In addition, MATLAB version 2018a is used to implement the algorithms.

Results

At first, effectiveness of the different values of ω_1 , ω_2 and ω_3 on the proposed weighted combination of multiple-features classification in probabilistic framework has been evaluated (Fig. 7). This experiment demonstrates that by selecting different values for these weights, the classification results are changed considerably. With set A, in Indian Pines image, the effectiveness of the shape probability distribution is higher than the texture and the spectral probability distribution. We can find almost the same effects in Salinas image, but in Pavia

Fig. 5 Pavia University image: **a** false-colour image **b** ground truth

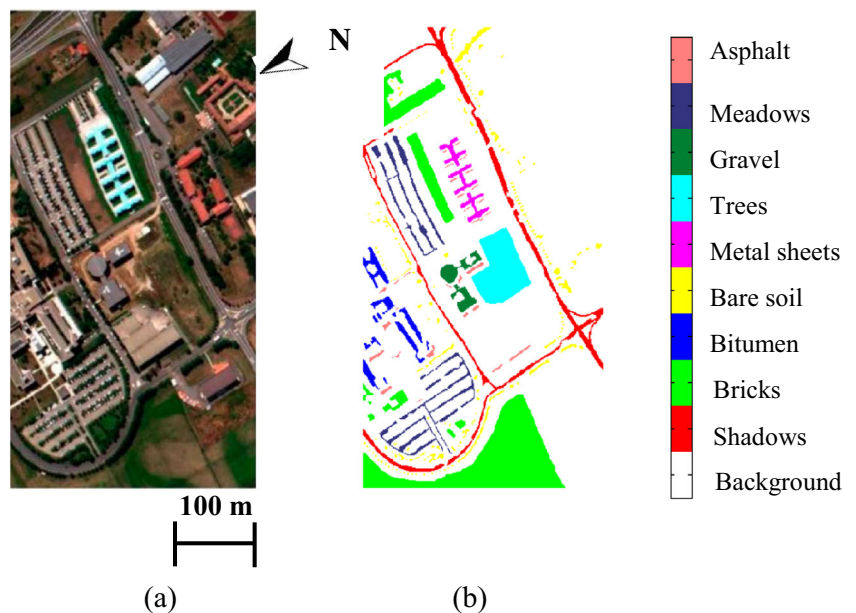
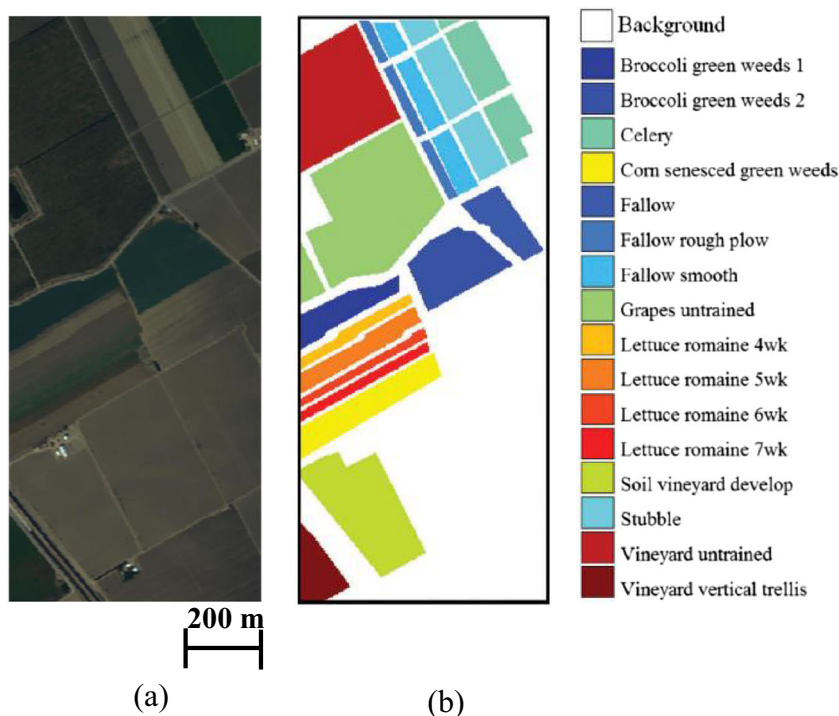


Fig. 6 Salinas image: **a** false-colour image **b** ground truth



University image, the texture probability distribution has more effectiveness. With set B, the classification results of the three data set are almost the same. In these condition the effectiveness of the shape and texture probability distribution are higher than the spectral probability distribution. It can be see that the effectiveness of the weights increases by decreasing the number of training samples.

Our experiment with different set of training samples demonstrate that various reasons, such as hyperspectral image, spatial resolution and the number of training samples, have effect to select the optimum weights ω_1 , ω_2 and ω_3 to reach the best classification results. The best choices of these parameters were derived experimentally and in each case are given in Fig. 7. Automatically selecting the weights in different situation, can be investigated in future works.

Now, the classification results of the spectral, texture and shape features are obtained on three hyperspectral images with two schemes of the training samples, set A and B, and compared with the classification results of the proposed method to show its superiorities.

1) *Indian Pines image*: the pixel-wise classification results of spectral, texture, shape features and the proposed method with limited training samples are given in Table 2. The low spatial resolution and the problem of highly mixed pixels of the Indian Pines image leads to the low classification accuracies with spectral and texture features, especially with set A. It can be seen that among these features, the shape one reveals better results than the spectral and texture features with set A. Moreover, the texture features reveal better results with set B. The OA, AA and κ of the shape features with set A are 74.75%, 77.74%, and 71.16%, respectively. On the other hand, the OA, AA and κ of the texture features with set B are 88.20%, 91.68%, and 86.36%, respectively. On the other hand, the OA, AA and κ of the proposed method with set A are 84.86%, 89.80%, and 82.63%, respectively and with set B are 96.54%, 97.70%, and 95.97%, respectively. If we compare these results, we can see the considerable improvement in classification accuracies. The proposed method improves the OA accuracy about 25%, 27% and

Table 1 The number of spectral bands and the total number of training samples for three data sets

Data set	Number of bands	10 training samples	50 training samples	1% training samples
Indian Pines (10-classes)	200	100	500	97
Pavia University	103	90	450	426
Salinas	204	160	800	543

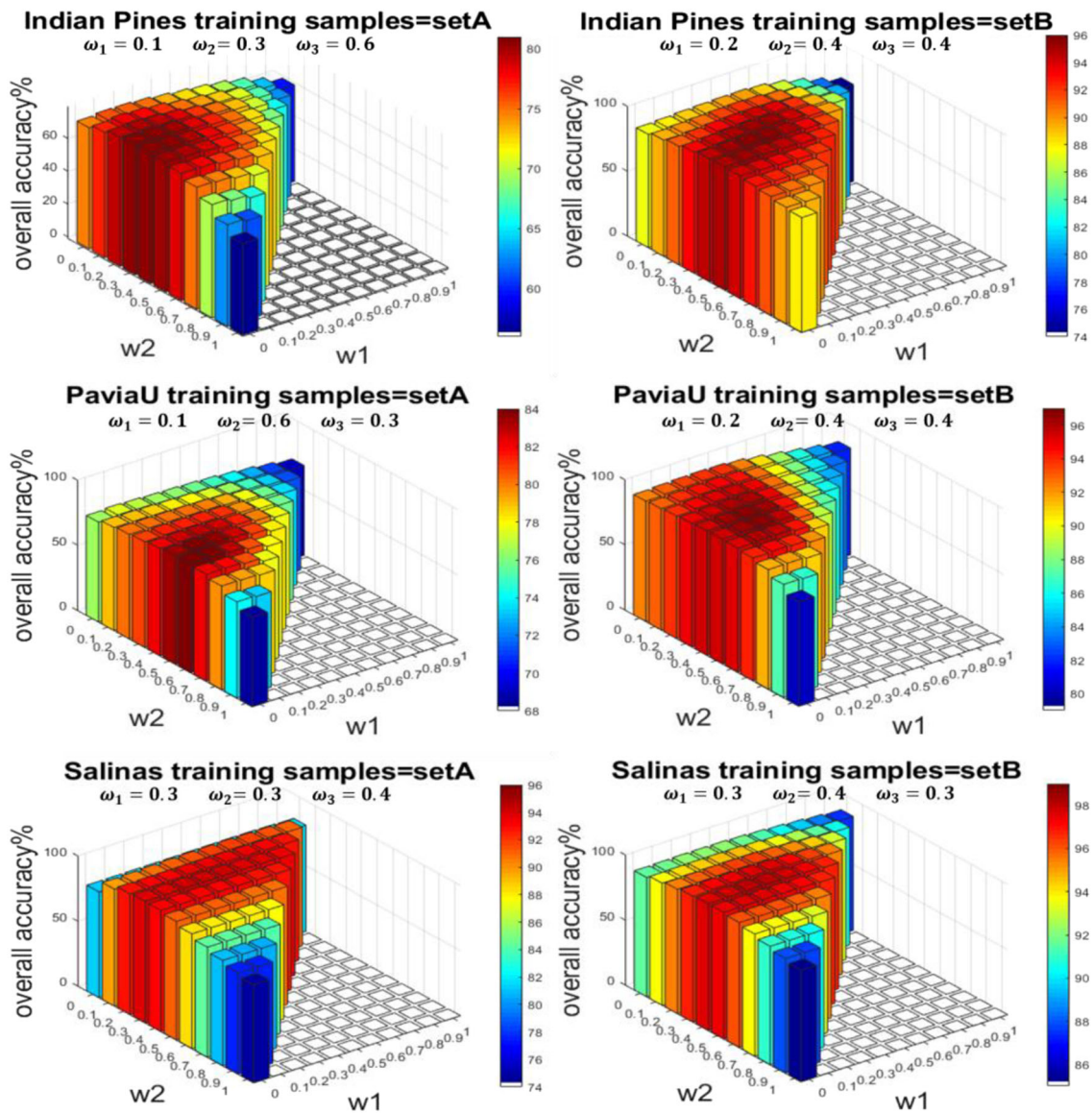


Fig. 7 The evaluation of the weighted combination with some different values of ω_1 , ω_2 and ω_3 on three data sets and with training samples set A and set B, in each case the best weights are given

10% in comparison with pixel-wise spectral, texture and shape features classification with set A, respectively. Similar improvements, but of lesser degree, can also be found with set B. The classification maps of set A and set B are shown in Fig. 8.

2) *Pavia University image*: as Indian Pines image, the classification accuracies of the spectral, texture and shape features; the proposed method are given in Table 3. It can be seen that among all features, shape features have the best performance with all schemes of training samples. On the other hand, the proposed method improves the classification accuracies about 10% with set A and about 5% with set B in comparison with the shape features. The classification maps of set A and set B are shown in Fig. 9.

3) *Salinas image*: as previous data sets, OA, AA and κ of the classification results with set A and set B are listed in Table 4. We can also see the benefits of the proposed weighted combination method in probabilistic framework with this data set. The classification maps are given in Fig. 10.

The simplest conclusion which can be drawn from the previous experiment is that the classification results of the proposed method depend on the data set and the number of training samples. On the other hand, we could not find a unique set of weights for all data sets to give the best results with different schemes of training samples. But the proper weights can improve the classification results

Table 2 Classification results of Indian Pines image

Classes		SVM (Spectral Features)		SVM (Texture Features)		SVM (Shape Features)		Proposed method	
Class Name	No. of Samples	Set A	Set B	Set A	Set B	Set A	Set B	Set A	Set B
Corn- no till	1428	26.72	72.46	08.39	78.11	56.69	77.10	61.14	89.47
Corn-min till	830	23.17	62.94	72.07	96.92	86.09	92.05	96.46	99.48
Grass-pasture	83	63.84	91.28	72.30	95.64	75.26	87.38	87.94	97.92
Grass-trees	730	91.38	94.72	42.50	83.45	88.19	91.50	97.63	99.70
Hay-windrowed	478	99.35	99.76	99.14	100	99.14	99.06	100	100
Soybean-no till	972	72.76	76.81	59.77	89.92	77.75	76.59	87.62	95.88
Soybean-min till	2455	61.63	52.22	51.61	81.45	69.89	87.44	77.99	95.59
Soybean-clean till	593	54.03	85.81	80.44	96.31	60.89	91.34	98.62	98.89
Woods	1265	74.74	94.738	71.39	95.06	82.39	95.06	90.27	99.50
Building-Grass-Trees-Drives	386	57.44	72.32	92.81	100	81.11	86.60	61.14	89.47
AA		62.51	80.31	65.04	91.68	77.74	88.41	89.80	97.70
OA		59.56	74.36	56.44	88.20	74.75	87.24	84.86	96.54
κ		53.60	70.64	50.37	86.36	71.16	85.19	82.63	95.97

considerably. We select the optimum weights experimentally in this article. The benefits of the proposed combination method are more considerable with set A, ill-posed training samples, in comparison with set B, poorly-posed training samples.

Overall time complexity analysis

Since the run time of the competing methods are not accessible to compare with our method, the overall time complexity of the proposed method is studied. The overall time complexity of the proposed method is studied here. The most time-

consuming parts are texture features extraction by Gabor filters, shape features extraction by MP and spatial contextual information extraction by shape-adaptive algorithm. Because of the use of SVM classifier with a polynomial kernel, without tuning the kernel parameters, the classification of all pixels based on the spectral, Gabor and MP features is not time-consuming. In addition, the process of combining probability distributions and applying spatial contextual information of the neighbouring pixels are not complex and they are not considered in total time complexity. Therefore, the time complexity of Gabor features extraction, MP extraction, shape-adaptive algorithm and SVM classification are denoted by

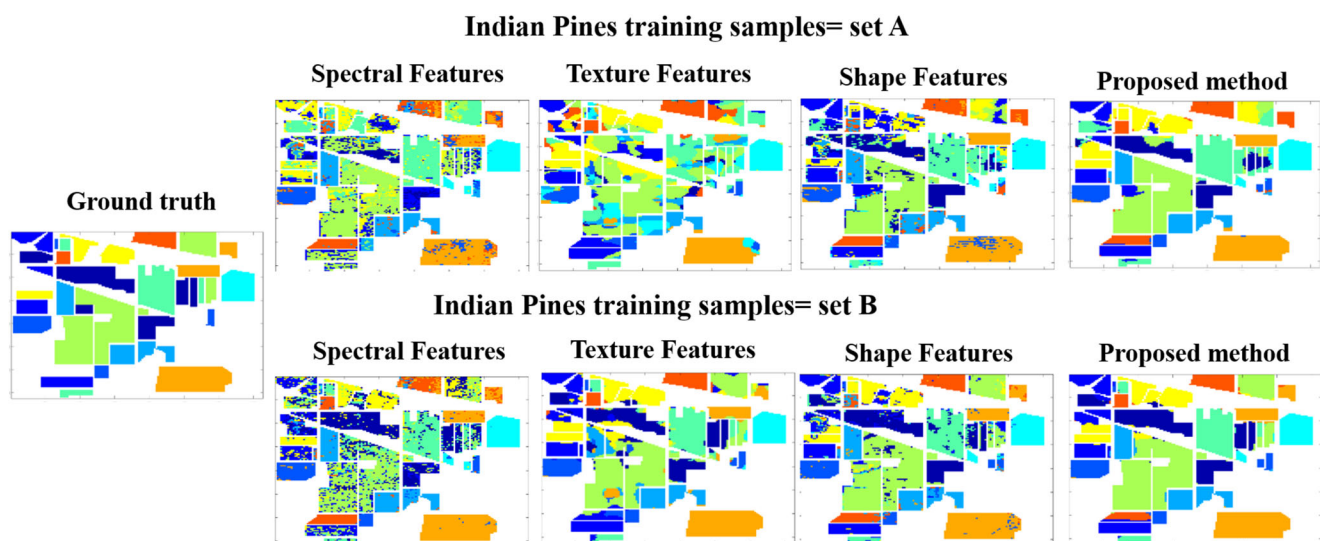


Fig. 8 Indian Pines image: the ground truth image and comparison of the classification maps of spectral, texture and shape features and proposed method with training samples set A and set B

Table 3 Classification results of Pavia University image

Classes		SVM (Spectral Features)		SVM (Texture Features)		SVM (Shape Features)		Proposed method	
Class Name	No. of Samples	Set A	Set B	Set A	Set B	Set A	Set B	Set A	Set B
Asphalt	6631	71.39	86.47	49.93	66.97	64.44	91.80	88.25	99.20
Meadows	18,649	71.39	78.06	90.80	92.99	81.75	95.69	88.45	99.10
Gravel	2099	50.40	78.83	47.86	83.60	85.83	93.17	79.88	97.49
Trees	3064	94.07	89.83	10.83	19.20	73.73	96.15	75.36	95.67
Metal sheets	1345	42.39	92.17	99.02	97.65	52.13	97.41	99.60	99.60
Bare Soil	5029	42.35	81.91	47.34	74.36	68.71	94.26	65.11	98.22
Bitumen	1330	81.21	83.13	45.22	85.51	94.69	90.26	91.36	100
Bricks	3682	76.22	80.98	76.68	88.73	76.44	67.69	94.65	94.56
Shadows	947	100	100	67.55	76.55	100	100	95.66	98.55
AA		69.66	85.71	68.65	76.17	77.52	91.83	85.46	98.04
OA		69.02	82.04	58.47	80.21	76.58	92.41	86.44	98.33
κ		60.48	77.03	58.23	73.55	69.73	90.02	80.90	97.79

O(Gabor), O(MP), O(SA) and O(SVM). Accordingly, the overall time complexity of the proposed method is calculated by:

$$(n-N) \times [O(\text{Gabor}) + O(\text{MP}) + O(\text{SA}) + 3 \times O(\text{SVM})] \quad (13)$$

Where, n is the number of pixels in the image and N .

Comparison with some recent spectral–spatial classification methods

The proposed method with set B for Indian Pines and Pavia University images and set C for Salinas image are compared with some recent spectral–spatial classification methods in order to evaluate the capability of the proposed method in hyperspectral image classification (Tables 5, 6 and 7). For

Indian Pines and Pavia University images, the proposed method is quantitatively compared with the extended morphological profiles (EMP) (Benediktsson et al. 2005), edge-preserving filter (EPF) (Kang et al. 2014), SVM-composite kernel (SVM-CK) (Camps-Valls et al. 2006), generalized composite kernel-based multivariate logistic regression (GCK-MLR) (Li et al. 2013), superpixel-based classification via multiple kernels (SC-MK) (Fang et al. 2015), multiple nonlinear feature learning with multivariate logistic regression (MNFL) (Li et al. 2015), EPs with a stacking manner (EPs-stacking) (Ghamisi et al. 2016), EPs-fusion (EPs-F) (Fang et al. 2018a, b, c). The spatial information in EMP and EPF are extracted by the morphological profiles and edge-filtering, respectively. In the SVM-CK method, the spectral and spatial features are combined together by the composite kernel on which the weights for each features are selected manually. Similarly, in the GCK-MLR method, the generalized

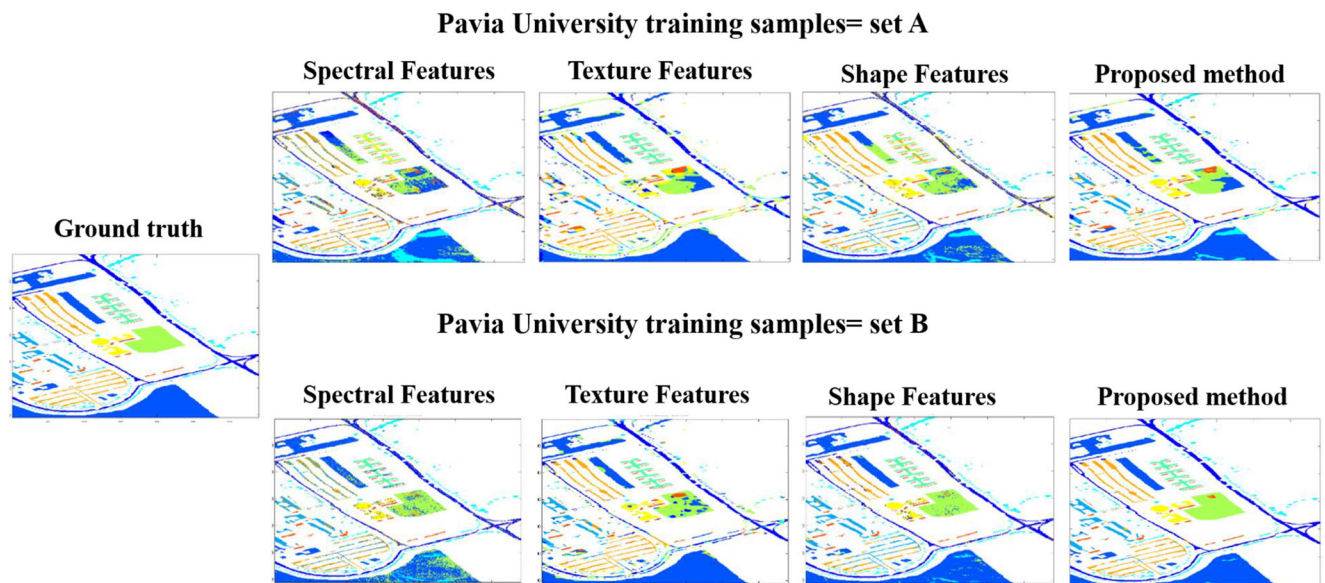


Fig. 9 Pavia University image: the ground truth image and comparison of the classification maps of spectral, texture and shape features and proposed method with training samples set A and set B

Table 4 Classification results of Salinas image

Classes		SVM (Spectral Features)		SVM (Texture Features)		SVM (Shape Features)		Proposed method	
Class Name	No. of Samples	Set A	Set B	Set A	Set B	Set A	Set B	Set A	Set B
Brocoli-green-weeds-1	2009	98.58	99.37	74.66	88.89	95.79	100	100	100
Brocoli-green-weeds-2	3726	94.60	95.11	76.07	85.25	66.54	98.16	100	100
Fallow	1979	68.03	99.57	70.166	86.09	98.03	99.52	98.97	99.89
Fallow-rough-plow	1394	98.03	99.32	65.72	75.90	91.19	99.24	99.63	99.78
Fallow-smooth	2678	92.13	97.64	73.75	77.55	95.42	98.50	95.66	98.18
Stubble	3959	94.28	96.83	74.50	85.90	95.38	98.48	99.84	99.89
Celery	3579	87.14	99.00	81.04	89.52	69.73	99.29	99.97	99.94
Grapes-untrained	11,271	67.43	78.09	73.79	87.57	99.15	86.86	99.67	100
Soil-vinyard-develop	6203	97.38	98.30	73.89	84.45	87.98	98.72	100	100
Corn-senesced-green-weeds	3278	81.85	86.12	69.26	85.26	98.62	93.67	98.58	99.35
Lattice-romaine-4wk	1068	97.63	95.86	72.91	88.47	99.45	96.55	100	100
Lattice-romaine-5wk	1927	99.39	99.89	66.89	86.45	96.89	97.21	99.52	99.84
Lattice-romaine-6wk	916	98.39	99.88	63.48	73.44	89.96	100	98.12	100
Lattice-romaine-7wk	1070	90.25	93.99	75.49	69.48	40.21	94.29	99.52	98.01
Vinyard-vertical-trellis	7268	56.29	56.17	75.77	87.64	81.30	66.66	98.69	99.49
Vinyard-vertical-trellis	1807	73.84	97.55	74.66	84.33	95.79	98.83	99.10	100
AA		87.20	87.43	72.73	83.51	87.82	95.37	99.20	99.65
OA		81.49	93.29	74.95	85.41	81.27	91.55	99.31	99.74
κ		79.43	85.99	71.99	83.74	79.12	90.58	99.23	99.71

composite kernel is constructed by the spectral–spatial information. A superpixel approach in the SC-MK method is used to extract the spatial and spectral information. Various extracted features are combined together in the MNFL method for better classification of hyperspectral images. A fusion framework in the EPs-F method is used to draw out the spatial information of the EPs. On the other hand, for Salinas image,

the proposed method is compared with the extended morphological profiles (EMP) (Benediktsson et al. 2005), the Logistic regression via variable splitting and augmented Lagrangian-multilevel logistic (LORSAL-MLL) (Li et al. 2011), the sparse representation classification-pixelwisely (SRC-pixelwisely), the joint sparse representation model (JSRM) (Chen et al. 2011), the nonlocal weighting sparse

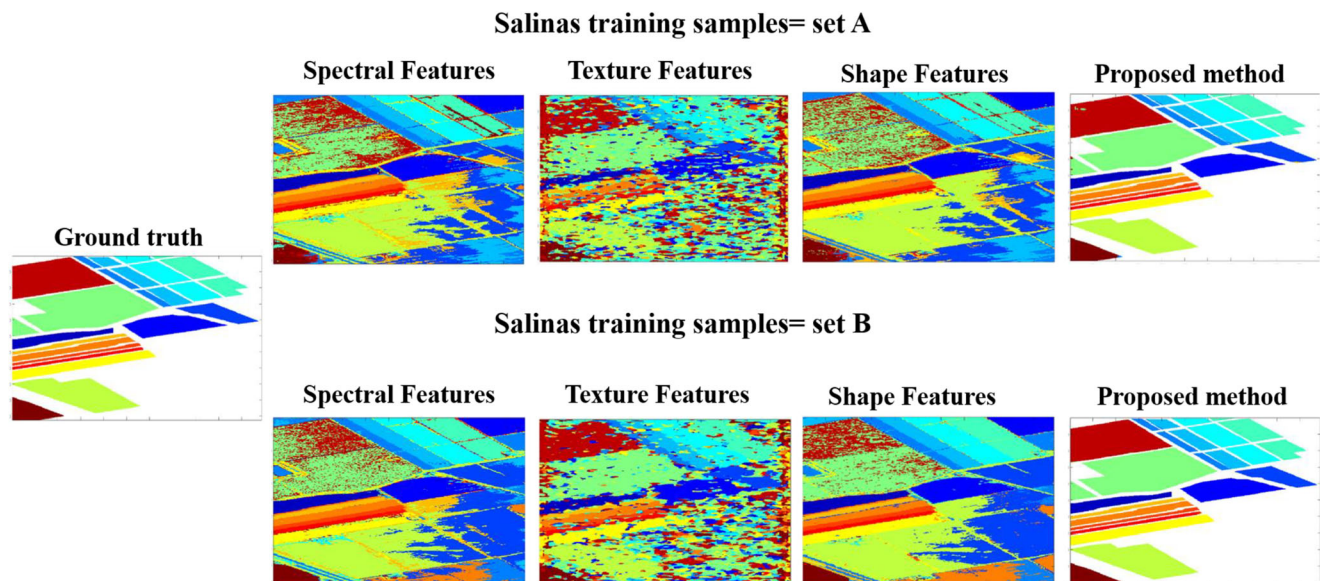


Fig. 10 Salinas image: the ground truth image and comparison of the classification maps of spectral, texture and shape features and proposed method with training samples set A and set B

Table 5 Indian Pines: classification accuracies of the proposed method compared with some recent spectral–spatial classification methods for 50 training samples (set B)

Methods	EMP	EPF	SVM-CK	GCK-MLR	SC-MK	MNFL	EPs- stacking	EPs-F	Proposed method
OA	89.43	88.31	83.68	89.36	95.06	90.86	87.66	95.85	96.54
AA	87.76	86.51	87.39	91.71	96.50	91.85	90.43	97.09	97.70
κ	90.01	88.22	81.14	87.73	94.32	89.44	85.89	95.22	95.97

representation (NLW-SR) (Zhang et al. 2014), the singlescale adaptive sparse representation(SASR), the multiscale separate sparse representation (MSSR), the multiscale joint sparse representation (MJSR), the multiscale adaptive sparse representation (MASR) (Fang et al. 2014). For the LORSAL-MLL method, the spatial information of the hyperspectral image is exploited by the multilevel logistic prior-based segmentation technique. The SRC-Pixel-wise is a sparse representation-based classifier on which the spectral information is used for classification. Similarly, the JSRM is a sparse representation-based classifier on which the spatial context is utilized within one fixed single scale. In NLW-SR method, a dynamic weight based on the spectral similarity and the local neighbourhood region is assigned for each pixel. The SASR method is a single-scale sparse representation classifiers based on modifications of the JSRM. In SASR, the adaptive atom selection strategy is applied on the JSRM. The MSSR and MJSR are two multiscale sparse representation-based classifiers. The MASR simultaneously represents pixels of multiple scales via an adaptive sparse strategy to exploit correlations among multiple scales and to represent the pixels of each scale to by an appropriate representation. Since we did not have the codes of these methods, the classification accuracies of the competing methods were obtained from the corresponding articles. The details and the parameter settings of different methods can be found in the original articles. It can be seen that the proposed method in Indian Pines and Salinas data sets outperforms the competing methods. The proposed method in Pavia University is better than the other methods although it is mostly similar to the EPs-F method. The OA of the proposed method is 98.33% and the OA of the EPs-F method is 98.67%. It can be seen that the proposed method in Indian Pines and Salinas data sets outperforms the competing methods on which the level of enhancement is noticeable.

The proposed method in Pavia University is better than the other methods although it is mostly similar to the EPs-F method. Eps-F is an effective spatial-spectral feature extraction method for hyperspectral images (HSIs), which has better performance in Pavia University, high resolution hyperspectral image. The given results in these tables demonstrate the performance of the proposed spectral–spatial classification method. These results show that the weighted combination of the probability distributions contain useful information about texture and shape features of the image. In addition, the spatial contextual information of the neighbouring pixels can significantly improve the classification accuracy. On the other hand, according to the results presented in Tables 5, 6 and 7, the proposed method with set B for Indian Pines and Pavia University images and set C for Salinas image, poorly-posed classification problem, outperforms most of the competing methods.

Conclusion

In this article a novel weighted combination of spectral and spatial information in probabilistic framework for classification of hyperspectral images, especially with limited training samples was proposed. Three probabilities of spectral, texture and shape features were combined together for each pixel on which three weights determine the efficacy of each one. Then the spatial contextual information of the neighbouring pixels in a SA region was applied to improve the classification results further. The main contributions of this article contain three-folds: 1) weighted combination of the spectral, texture, shape features and the contextual information in probabilistic framework is the main contribution of this article that didn't appear in the previous works in which all extracted spatial

Table 6 Pavia University: classification accuracies of the proposed method compared with some recent spectral–spatial classification methods for 50 training samples (set B)

Methods	EMP	EPF	SVM-CK	GCK-MLR	SC-MK	MNFL	EPs- stacking	EPs-F	Proposed method
OA	94.57	93.57	91.14	96.40	96.28	94.95	96.76	98.67	98.33
AA	92.85	91.60	91.95	96.34	97.51	94.91	97.53	99.13	98.04
κ	93.21	92.82	88.34	95.24	95.11	93.43	95.68	98.23	97.99

Table 7 Salinas: classification accuracies of the proposed method compared with some recent spectral–spatial classification methods for 1% training samples (set C)

Methods	EMP	LORSAL- MLL	SRC- pixelwised	JSRM	NLW- SR	SASR	MSSR	MJSR	MASR	Proposed method
OA	96.23	93.75	85.89	93.96	95.25	95.19	95.93	93.46	97.33	99.44
AA	97.29	96.39	91.95	95.36	97.34	94.03	97.14	96.95	98.53	99.18
κ	95.00	93.00	84.00	93.00	95.00	95.00	95.00	93.00	97.00	99.38

information are combined with the spectral data by using probability distribution functions, 2) Unlike some spectral–spatial classification methods which use of the fixed-window neighbourhood, in this article, the contextual information in a SA neighbourhood is used to enhance the classification results, 3) in the previous works, the number of the similar and dissimilar labels in the neighbourhood is used to refine the label of the central pixel, but the reliability of the neighbourhood labels is not considered. The proposed method was performed on three data sets with different limited training samples. Studying the results showed that the use of probabilistic framework to combine the spectral and spatial information is a simple and robust technique for enhancing the classification accuracies in hyperspectral images. Finally, Comparison with some recent spectral–spatial classification methods demonstrated better performance of the proposed method.

References

- Benediktsson JA, Ghamisi P (2015) Spectral–spatial classification of hyperspectral remote sensing images. Artech House
- Benediktsson JA, Palmason JA, Sveinsson JR (2005) Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Trans Geosci Remote Sens* 43(3):480–491
- Camps-Valls G, Bruzzone L (2005) Kernel-based methods for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 43(6):1351–1362
- Camps-Valls G, Gomez-Chova L, Munoz-Mari J, Vila-Frances J, Calpe-Maravilla J (2006) Composite kernels for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 3(1):93–97
- Camps-Valls G, Tuia D, Bruzzone L, Benediktsson JA (2014) Advances in hyperspectral image classification: earth monitoring with statistical learning methods. *IEEE Signal Process Mag* 31(1):45–54
- Canty MJ (2014) Image analysis, classification, and change detection in remote sensing with algorithms for ENVI/IDL, 3rd edn. CRC Press, Boca Raton
- Chang CC, Lin CJ (2011) LIBS SVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–27 (software available at: www.csie.ntu.edu.tw/~cjlin/libsvm)
- Chen Y, Nasrabadi NM, Tran TD (2011) Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans Geosci Remote Sens* 49:3973–3985
- Cohen J (1960) A coefficient of agreement from nominal scales. *Educ Psychol Meas* 20: 37–46. <https://doi.org/10.1177/001316446002000104>
- Dalla Mura M, Benediktsson JA, Waske B, Bruzzone L (2010) Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *Int J Remote Sens* 31(22):5975–5991
- Daugman J (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A* 2(7):1160–1169
- Fang L, Li S, Kang X, Benediktsson JA (2014) Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation. *IEEE Trans Geosci Remote Sens* 52(12):7738–7749
- Fang L, Li S, Duan W, Ren J, Benediktsson JA (2015) Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE Trans Geosci Remote Sens* 53:6663–6674
- Fang L, Wang C, Li S, Benediktsson JA (2017) Hyperspectral image classification via multiple-feature-based adaptive sparse representation. *IEEE Trans Instrum Meas* 66:1646–1657
- Fang L, He N, Li S, Ghamisi P, Benediktsson JA (2018a) Extinction profiles fusion for hyperspectral images classification. *IEEE Trans Geosci Remote Sens* 56:1803–1815
- Fang L, He N, Li S, Plaza AJ, Plaza J (2018b) A new spatial–spectral feature extraction method for hyperspectral images using local covariance matrix representation. *IEEE Trans Geosci Remote Sens* 56:3534–3546
- Fang L, Liu G, Li S, Ghamisi P, Benediktsson JA (2018c) Hyperspectral image classification with squeeze multibias network *IEEE Transactions on Geoscience and Remote Sensing* 57:1291–1301
- Fauvel M, Tarabalka Y, Benediktsson JA, Chanussot J, Tilton JC (2013) Advances in spectral–spatial classification of hyperspectral images. *Proc IEEE* 101(3):652–675
- Fu W, Li S, Fang L, Kang X, Benediktsson JA (2016) Hyperspectral image classification via shape-adaptive joint sparse representation. *IEEE J Sel Top Appl Earth Obs Remote Sens* 9:556–567
- Ghamisi P, Souza R, Benediktsson JA, Rittner L, Lotufo R, Zhu XX (2016) Hyperspectral data classification using extended extinction profiles. *IEEE Geosci Remote Sens Lett* 13:1641–1645
- Golipour M, Ghassemian H, Mirzapour F (2016) Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework. *IEEE Trans Geosci Remote Sens* 54:805–816
- Guo X, Huang X, Zhang L (2014) Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery. *IEEE Geosci Remote Sens Lett* 11:2183–2187
- Huo L-Z, Tang P (2011) Spectral and spatial classification of hyperspectral data using SVMs and Gabor textures. In: 2011 IEEE International Geoscience and Remote Sensing Symposium pp 1708–1711
- Jia X, Kuo B-C, Crawford MM (2013) Feature mining for hyperspectral image classification. *Proc IEEE* 101:676–697
- Kang X, Li S, Benediktsson JA (2014) Spectral–spatial hyperspectral image classification with edge-preserving filtering. *IEEE Trans Geosci Remote Sens* 52:2666–2677
- Khodadadzadeh M, Li J, Plaza A, Ghassemian H, Bioucas-Dias JM, Li X (2014) Spectral–spatial classification of hyperspectral data using

- local and global probabilities for mixed pixel characterization. *IEEE Trans Geosci Remote Sens* 52:6298–6314
- Kianisarkaleh A, Ghassemian H (2016) Nonparametric feature extraction for classification of hyperspectral images with limited training samples. *ISPRS J Photogramm Remote Sens* 119:64–78
- Li G, Wan Y (2015) A new combination classification of pixel-and object-based methods. *Int J Remote Sens* 36:5842–5868
- Li J, Bioucas-Dias JM, Plaza A (2011) Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Trans Geosci Remote Sens* 49(10):3947–3960
- Li J, Marpu PR, Plaza A, Bioucas-Dias JM, Benediktsson JA (2013) Generalized composite kernel framework for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 51:4816–4829
- Li J, Huang X, Gamba P, Bioucas-Dias JM, Zhang L, Benediktsson JA, Plaza A (2015) Multiple feature learning for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 53:1592–1606
- Li S, Ni L, Jia X, Gao L, Zhang B, Peng M (2016) Multi-scale superpixel spectral–spatial classification of hyperspectral images. *Int J Remote Sens* 37:4905–4922
- Liu J, Lu W (2016) A probabilistic framework for spectral–spatial classification of hyperspectral images. *IEEE Trans Geosci Remote Sens* 54:5375–5384
- Matheron G (1975) *Random sets and integral geometry*. Wiley, New York
- Mirmehdi M (2008) *Handbook of texture analysis*. Imperial College Press
- Mirzapour F, Ghassemian H (2015) Improving hyperspectral image classification by combining spectral, texture, and shape features. *Int J Remote Sens* 36:1070–1096
- Negri RG, Dutra LV, Sant’Anna SJS (2014) An innovative support vector machine based method for contextual image classification. *ISPRS J Photogramm Remote Sens* 87:241–248
- Seifi Majdar R, Ghassemian H (2017a) Spectral-Spatial classification of hyperspectral images using functional data analysis. *Remote Sens Lett* 8:488–497
- Seifi Majdar R, Ghassemian H (2017b) A probabilistic SVM approach for hyperspectral image classification using spectral and texture features. *Int J Remote Sens* 38:4265–4284
- Serra J (1983) *Image analysis and mathematical morphology*. Academic Press, Inc.
- Soille P (2013) *Morphological image analysis: principles and applications*. Springer Science & Business Media
- Solberg AHS, Taxt T, Jain AK (1996) A Markov random field model for classification of multisource satellite imagery. *IEEE Trans Geosci Remote Sens* 34:100–113
- Tarabalka Y, Fauvel M, Chanussot J, Benediktsson JA (2010a) SVM-and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosci Remote Sens Lett* 7:736–740
- Tarabalka Y, Chanussot J, Benediktsson JA (2010b) Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recogn* 43:2367–2379
- Vapnik VN (2000) *The nature of statistical learning theory*, 2nd edn. Springer-Verlag, New York
- Wang Y, Duan H (2018) Classification of hyperspectral images by SVM using a composite kernel by employing spectral, spatial and hierarchical structure information. *Remote Sens* 10:441
- Wang Y, Duan H (2019) Spectral–spatial classification of hyperspectral images by algebraic multigrid based multiscale information fusion. *Int J Remote Sens* 40:1301–1330
- Wang L, Shi C, Diao C, Ji W, Yin D (2016a) A survey of methods incorporating spatial information in image classification and spectral unmixing. *Int J Remote Sens* 37:3870–3910
- Wang Y, Song H, Zhang Y (2016b) Spectral-spatial classification of hyperspectral images using joint bilateral filter and graph cut based model. *Remote Sens* 8:748
- Wu T-F, Lin C-J, Weng RC (2004) Probability estimates for multi-class classification by pairwise coupling. *J Mach Learn Res* 5:975–1005
- Zehtabian A, Ghassemian H (2015) An adaptive pixon extraction technique for multispectral/hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 12:831–835
- Zhang B, Li S, Jia X, Gao L, Peng M (2011) Adaptive Markov random field approach for classification of hyperspectral imagery. *IEEE Geosci Remote Sens Lett* 8:973–977
- Zhang L, Zhang L, Tao D, Huang X (2012) On combining multiple features for hyperspectral remote sensing image classification. *IEEE Trans Geosci Remote Sens* 50:879–893
- Zhang H, Li J, Huang Y, Zhang L (2014) A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7:2056–2206
- Zhu J, Fang L, Ghamisi P (2018) Deformable convolutional neural networks for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 15:1254–1258

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.