Classification of Hyperspectral Imagery Based on Dictionary Learning and Extended Multi-attribute Profiles

Qishuo Gao[1(✉)] [iD], Samsung Lim[1] [iD], and Xiuping Jia[2]

[1] School of Civil and Environmental Engineering,
The University of New South Wales, Sydney, NSW, Australia
`qishuo.gao@student.unsw.edu.au`
[2] School of Engineering and Information Technology,
The University of New South Wales, Canberra, ACT, Australia

Abstract. In recent years, sparse representation has shown its competitiveness in the field of image processing, and attribute profiles have also demonstrated their reliable performance in utilizing spatial information in hyperspectral image classification. In order to fully integrate spatial information, we propose a novel framework which integrates the above-mentioned methods for hyperspectral image classification. Specifically, sparse representation is used to learn a posteriori probability with extended attribute profiles as input features. A classification error term is added to the sparse representation-based classifier model and is solved by the k-singular value decomposition algorithm. The spatial correlation of neighboring pixels is incorporated by a maximum a posteriori scheme to obtain the final classification results. Experimental results on two benchmark hyperspectral images suggest that the proposed approach outperforms the related sparsity-based methods and support vector machine-based classifiers.

Keywords: Extended multi-attribute profiles · Markov Random Field
Hyperspectral imagery · Sparse representation classification

1 Introduction

Hyperspectral imagery (HSI) has been widely used in the field of remote sensing for the past decade. Its capability to acquire hundreds of images with a wide range of wavelengths makes HSI a powerful tool in many areas, such as military surveillance, natural resources detection, land cover classification, etc. [1]. However, the unique properties of HSI have posed difficult image processing problems; for instance, it has been identified as a challenging task to analyze the spectral and spatial information simultaneously for HSI classification [2].

Many attempts have been made to solve this problem. In [3], the authors proposed a new sampling strategy to extract both spectral and spatial information. In addition, Markov Random Field (MRF) is considered to be a powerful method for integrating both spectral and spatial information [4]. However, its efficiency and effectiveness are questionable due to the high computational complexity and the uncertainty of the smoothing parameter to be chosen. Another category of approaches to deal with

contextual information is based on attribute profiles (APs) which are constructed by a set of attribute filters (AFs). AFs operate only on the connected components based on a criterion that evaluates the attribute against a threshold. When multiple layers of HSI are considered, the stack of individually computed APs can be referred to as an extended attribute profile (EAP) [5]. Moreover, if different attributes are considered and multiple EAPs are stacked together, an extended multi-attribute profile (EMAP) can be constructed [6]. EMAP can effectively deal with both spectral and contextual information [7].

Sparse representation-based classifiers (SRCs) have been found to be efficient tools in many image processing areas in the last few years [8–10]. SRC assumes that each signal can be expressed as a linear combination of prototypes selected from a dictionary. The advantages of applying SRC as a classification method have been investigated [11–13]. SRC can achieve good performance on HSI classification because the pixels with highly correlated bands can be sparsely represented. Usually a SRC dictionary is directly constructed by training atoms, which limits the HSI classification accuracy due to the large number of training atoms. Hence it is sensible to construct a reliable dictionary for the classification problem.

Based on the aforementioned knowledge, a novel framework using both EMAP and SRC is developed and presented in this paper particularly for HSI classification, which we name it as extended SRC (ESRC) method. Apart from spectral information, EMAPs have been constructed to initialize the dictionary for SRC. Thus, both spectral and spatial context can be considered to maximize among class separability. Subsequently, we optimize the dictionary using an effective method known as *k*-singular value decomposition (K-SVD) [14]. Similar to Jiang [15], we add a classification error term to the SRC model, then the reconstruction error and the classification error can be modelled simultaneously. Finally, the class label can be derived via the MRF-based maximum a posteriori (MAP) method, where the spatial energy term is improved by a Gaussian framework in this paper. It should be noted that the spatial information is utilized via EMAPs and then regularized by the MRF-MAP method, therefore our ESRC can further improve the classification results.

The remainder of this paper is organized as follows. The proposed framework is described in Sect. 2. The effectiveness of the framework is demonstrated by the experiments in Sect. 3. Finally, we conclude and provide some remarks in Sect. 4.

2 Design of Framework

2.1 EMAP Feature Extraction

EAPs are built by concatenating many attribute profiles (APs), and each AP is generated for each feature in a scalar hyperspectral image. That is:

$$EAP = \{AP(f_1), AP(f_2), \ldots, AP(f_n)\} \tag{1}$$

APs are a generalized form of morphological profiles, which can be obtained from an image by applying a criterion T. By using n morphological thickening (φ^T) and n thinning ($\emptyset^T$) operators, an AP can be constructed as:

$$AP(f) = \{\varphi_n^T(f), \varphi_{n-1}^T(f), \ldots, \varphi_1^T(f), f, \phi_1^T(f), \ldots, \phi_{n-1}^T(f), \phi_n^T(f)\} \tag{2}$$

Generally, there are some common criteria associated with the operators, such as area, volume, diagonal box, and standard deviation. According to the operators (thickening or thinning) used in the image processing, the image can be transformed to an extensive or anti-extensive one. In this paper, area, standard deviation, the moment of inertia, and length of the diagonal are used as the attributes to compute EMAP features for classification tasks. The stack of different EAPs leads to EMAPs, and the detailed information of EMAPs can be found in the report [7].

2.2   Dictionary Learning for SRC

Suppose $\{x_i\}_{i=1}^N$ represents N training samples from a L-dimensional hyperspectral dataset, and x belongs to c labelled classes while $y_i \in \{1 \ldots c\}$ is the label of each observed pixel x_i. For each class c, there exists a matrix $D_C \in R^{m \times n^c} (N = \sum_{i=1}^c n_c)$ containing n_c prototype atoms for columns. Each pixel with a c-th label can be represented approximately as follows:

$$x_i \approx D_C r_i^c, \quad \left\| r_i^c \right\| \leq K, \qquad \forall y_i = c \tag{3}$$

where $r^c \in R^m$ is the representation coefficient of signal x, $\| \cdot \|_0$ is a l_0 norm which counts the number of nonzero atoms in a coefficient vector, and K is a predefined sparsity constraint. Assuming that the global dictionary $D = [D_1, D_2 \ldots D_c]$ is known, the corresponding representation coefficients $r = [r^1, r^2 \ldots r^c]$ can be computed by solving the following equation:

$$r_i^c = \arg\min_r \left\| x_i - Dr_i^c \right\|_2^2 \quad s.t. \quad \forall i, \|r_i\|_0 \leq K \tag{4}$$

There exist many algorithms to optimize this problem, for example, orthogonal matching pursuit (OMP) [16] is one of the most efficient methods. OMP is a greedy method which simply selects the dictionary prototypes in sequence. The pixels belong to the class which has the minimum class-wise reconstruction error e_i^c, where $e_i^c = x_i - D_C r_i^c$:

$$\hat{y}_i = \arg\min_c \left\| e_i^c \right\|_2^2 \tag{5}$$

The main goal of this paper is to find a dictionary D which can help maximize classification accuracy. In order to minimize the reconstruction error and the classification error simultaneously, a classification error term $\left\| H - W^T r \right\|_2^2$ which has the sparse code directly as a feature for classification, is included in the objective function.

Following the solution by Jiang et al. [15], the objective function for learning optimal dictionary and sparse code can be redefined as follows:

$$< D, r, W > \ = \arg \min_{D,r,W} \|X - Dr\|_2^2 + \alpha \|H - W^T r\|_2^2$$

$$s.t. \forall i \quad \|r_i\|_0 \leq K \tag{6}$$

where $H = [h_1, h_2 \ldots h_N] \in R^{c \times N}$ represents the class labels of the training samples; $h_i = [0, 0 \ldots 1 \ldots 0, 0] \in R^c$, where the nonzero c-th position represents the c-th class that contains x_i; Wr is a linear classification function that supports learning an optimal dictionary; and α is a scalar controlling corresponding terms.

In order to use $K - SVD$ as the efficient solution, Eq. (6) can be rewritten as the following:

$$< D, r, W > \ = \arg \min_{D,r,W} \left\| \begin{matrix} X \\ \sqrt{\alpha}H \end{matrix} \ \begin{pmatrix} D \\ \sqrt{\alpha}W^T \end{pmatrix} r \right\|_2^2$$

$$s.t. \forall i \quad \|r_i\|_0 \leq K \tag{7}$$

The initial dictionary is obtained by EMAPs; given the initialized D, the original K−SVD is employed to obtain r, and r can be used to compute the initial W with linear support vector machine (SVM); both D and W are updated by the K−SVD algorithm.

Let $X_{new} = (X^T, \sqrt{\alpha}H^T)^T$, $D_{new} = (D^T, \sqrt{\alpha}W)^T$, then Eq. (7) can be rewritten as:

$$< D_{new}, r > \ = \arg \min_{D_{new},r} \{\|X_{new} - D_{new}r\|_2^2\}$$

$$s.t. \forall i, \quad \|r_i\|_0 \leq K \tag{8}$$

Then K−SVD algorithm is employed to optimize this problem. Let d_k and r_k represent the kth row in D and its corresponding coefficients, respectively. The overall processing steps of K−SVD is summarized in the report by Ahron et al. [14]. D_{new} and r are computed by K−SVD, and then $D = \{d_1, d_2 \ldots d_k\}$ and $r = \{r_1, r_2 \ldots r_k\}$ can be obtained from D_{new}. The representation error vector can be computed from $e_{ic} = x_i - D_c r_i^c$. Additionally, the posterior probability $\rho(y_{ic}/x_i)$ is inversely proportional to e_{ic}[17]:

$$\rho(y_{ic}/x_i) = \frac{1}{\|e_{ic}\|_2^2 \mu} \tag{9}$$

where y_{ic} refers to labelled class c for the pixel x_i and μ is a normalized constant.

2.3   Spatial Information Regularization

We have described the mechanism that is used to obtain the class probability in the previous section. In this section, we will show how to implement the spatial information characterization.

By utilizing the MAP theory, a pixel is likely to have the same label with its neighboring pixels. According to the Hammersley-Clifford theory [18], the MAP estimation of y is represented as follows:

$$\hat{y} = \arg\min_{y}\left(-\sum_{i \in N(i)} \log\rho(y_i/x_i) - \log\rho(y)\right)$$

$$\rho(y) = e^{\sigma \sum_{i \in N(i)} \delta(y_i - y_j)}$$

(10)

where the term $\rho(y_i/x_i)$ is the spectral energy function and can be estimated from previous Eq. (8). $\rho(y)$ is regularized by a unit impulse function $\delta(y_i - y_j)$, where $\delta(0) = 1$ and $\delta(y) = 0$ for $y \neq 0$; additionally, σ is a smoothing parameter. This term attains the probability of one to equal neighboring labels and zero to the other way around. In this way, non-probability might result in a misclassification for some mixing regions. In this paper, we modify this term by a Gaussian radial basis network to make it a more efficient way, and the entire image is decomposed into local patches with a neighborhood size $N \times N$. For $j \in N(i)$, the function applied on the pixels constrained by the neighborhood size can be represented as:

$$\delta(y_i - y_j) = \begin{cases} 1, & if \quad y_i = y_j \\ \exp\left(-\left(\frac{u_{ij}}{s_{u_{ij}}}\right)^2 - \left(\frac{v_{ij}}{s_{v_{ij}}}\right)^2\right) & if \quad y_i \neq y_j \end{cases}$$

(11)

We improve the unit impulse function by optimizing the weight of different class probability using a smoothing function. $\delta(y_i - y_j)$ is a function of standardized Euclidian distance between pixels x_i and x_j, where u_{ij} and v_{ij} are the horizontal and vertical distances from x_i and x_j, respectively. $s_{u_{ij}}$ and $s_{v_{ij}}$ represent the stand deviation in each direction. The range of Eq. (11) can meet the definition of probability constrained by (0,1]. Given the prior class information and spatial locations of pixels, this improvement can be trained quickly and efficiently.

The value 1 is assigned to $\rho(y_i = y_j)$ which indicates that the pixels tend to appear around those of the same class. Since homogenous areas are dominant, the improved function will yield a good approximation for the regions, especially for the edge area.
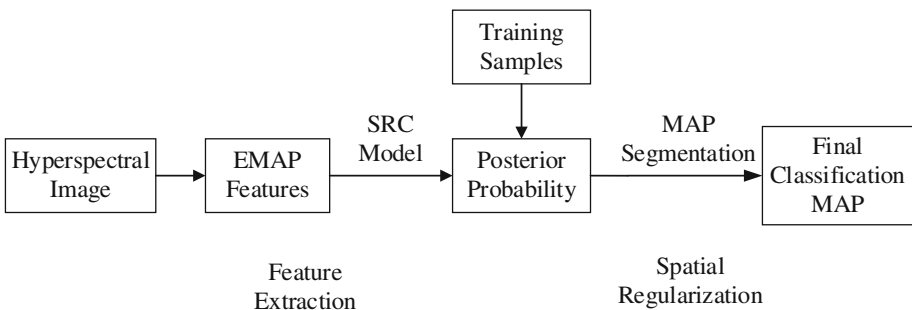

Fig. 1. Flowchart of the proposed framework

As shown in Eq. (11), the spatial relationship is modelled directly by the spatial locations.

In this paper, we utilize the $\propto$ −expansion algorithm to optimize this combinatorial problem. To better understand the main procedures of presented framework, the flowchart is shown as in Fig. 1.

3   Experiment Analysis

3.1   Experimental Setup

Two benchmark hyperspectral images are used to evaluate the proposed method. The attribute values used for EMAPs transformations are described as follows. Area of regions: 5000, length of the diagonal: 100, moment of inertia: 0.5, and standard deviation: 50. Smoothing parameter σ is set as 0.2 in this paper.

The best parameters are chosen via cross-validation for classifiers in this paper, and the results are compared with those acquired by kernel SVM (KSVM), sparsity representation model using OMP (SRC), Kernel SVM with EMAP features (EKSVM), sparsity representation model using OMP with EMAP features (ERAP), SVM probability with EMAP features and original MRF (SVM_AP), and original MRF with probability obtained with ERAP (referred as ERAP_MRF).

All experiments are implemented with Matlab 2015b. Average accuracy (AA), Overall accuracy (OA), kappa coefficient (k) are calculated as the accuracy assessment, which are commonly used for classification tasks.

3.2   Experiments on Indian Pines Data Set

The Indian Pines data set was acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in 1992. It covers 145×145 pixels with a spatial resolution of 20 m. After removing 20 water absorption bands, 200 spectral bands from 0.2 to 2.4 µm as the original features. There are 16 labelled classes, which are shown in Table 1, as well as the numbers of training and testing datasets.

The average results of ten experiments repeatedly run on different randomly chosen training and testing datasets are shown in Table 2. The classification maps as well as the ground truth are shown in Fig. 2.

The listed results in Table 2 and Fig. 2 show that our framework outperforms most techniques and is especially better than KSVM which is known as a state-of-the-art method. The visual results also show that the spatial information involved techniques lead to a much smoother classification map than algorithms with only spectral information involved.

One can observe that EMAP with normal classifiers have already provided high classification accuracy, however, the transform of a probabilistic work and including a MAP segmentation work improve the classification accuracies as it can be particularly observed in the ERAP_MRF, SVM_AP and ESRC. This confirms the ability of the work in a probabilistic sense that MAP segmentation can indeed correct the results by

Table 1. Class information for Indian Pines image

Class No.	Class name	Train	Test
1	Alfalfa	25	29
2	Corn-no till	50	1384
3	Corn-min till	50	784
4	Corn	50	184
5	Grass/trees	50	447
6	Grass/pasture	50	697
7	Grass/pasture-mowed	13	13
8	Hay-windrowed	50	439
9	Oats	10	10
10	Soybeans-no till	50	918
11	Soybeans-min till	50	2418
12	Soybeans-clean till	50	564
13	Wheat	50	162
14	Woods	50	1244
15	Buildings-grass	50	330
16	Stone-steel towers	45	50
Total		693	9673

Table 2. Classification accuracies (%) for Indian Pines image

Method	SRC	KSVM	ERAP	EKSVM	SVM_AP	ERAP_MRF	ESRC
1	54.17	48.57	64.63	61.45	70.67	71.23	85.71
2	77.68	74.10	85.12	84.60	90.46	90.36	96.22
3	78.91	73.84	84.49	83.70	92.17	91.00	95.96
4	50.48	45.43	60.33	57.18	68.87	74.74	83.09
5	86.43	83.94	90.51	88.76	92.26	94.53	97.41
6	94.91	93.82	96.34	96.25	97.92	97.66	98.92
7	95.65	91.67	95.83	100.00	96.00	100.00	100.00
8	98.34	97.93	98.56	97.96	99.59	98.77	99.79
9	73.08	72	82.61	79.17	90.48	90.00	90.91
10	69.55	65.52	78.23	75.70	84.35	87.33	91.93
11	83.82	80.83	89.85	88.81	93.63	93.51	96.85
12	72.60	68.90	81.56	78.31	87.31	86.63	93.79
13	97.17	97.17	98.12	98.10	98.59	99.05	99.53
14	97.89	97.56	98.48	98.80	99.18	99.27	99.45
15	60.58	54.92	69.81	68.54	78.35	80.00	89.05
16	83.04	80.87	86.24	86.11	90.29	94.00	95.92
OA	80.90	77.63	86.90	85.71	91.38	91.98	95.76
AA	79.64	76.69	85.05	83.97	89.38	90.50	94.56
k	78.44	74.78	85.17	83.84	90.23	90.90	95.17
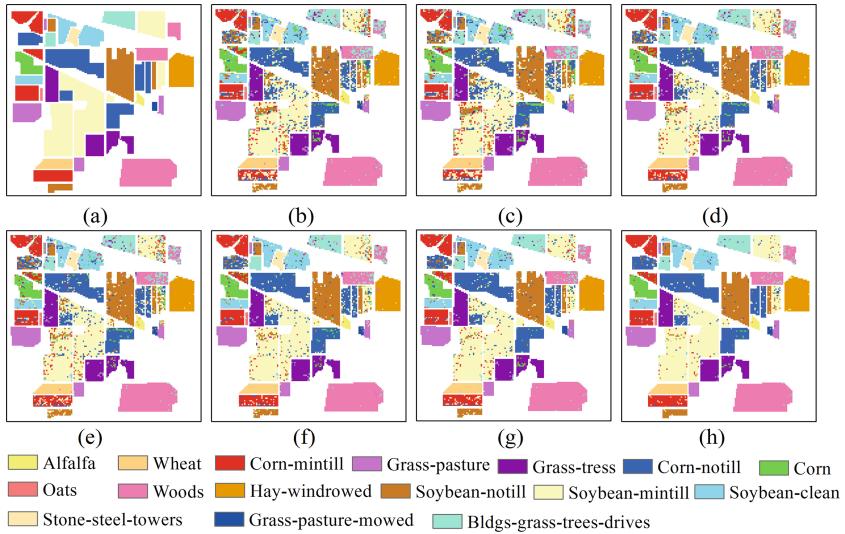
Fig. 2. Classification maps of Indian Pines. (a) Ground truth. (b) SRC. (c) KSVM. (d) ERAP. (e) EKSVM. (f) SVM_AP (g) ERAP_MRF. (h) ESRC.

regularizing the spatial information. It also should be noted that our proposed ESRC obtains the best result, which classifies most of regions accurately.

As for a SRC based method, the proposed ESRC exhibits the best performance especially in the edge areas, which can be observed from the classification maps. In addition, ESRC has also shown its potential effectiveness in dealing with the small training data sets, which is meaningful for practice applications. ESRC achieves the best result when compared to EKSVM and SVM_AP, which further confirms that our method can learn a discriminative dictionary and implement a more accurate spatial regularization. The results show that the proposed ESRC method is more accurate than original MRF-based spatial regularization methods. Particularly, ESRC performs well on the minority classes (e.g. Class 1 and Class 7).

The visual comparison of ERAP_MRF and SVM_AP also confirms the competitive efficiency of optimization model for SRC used in this paper. However, both of them fail to identify Class 9. This is due to the insufficient training samples for this class. The experiments also indicate that the improved method has a potential to obtain a more accurate result with a smaller training set.

For ERAP_MRF and ESRC, a window size 8×8 is applied. The former method over-smooths the oat-covered region and misclassifies Oats as Grass/Trees or Corn-min. This is because each oat pixel is dominated by Class 3 (Corn-min) and Class 6 (Grass/Trees). The improvement of spatial regularization gives a weight of pixels far away from the central pixel via spatial locations, which is helpful for the dominated regions.

The accuracy varies with different sparsity constraint factor K and neighborhood size N. The effect of K and N on the classification accuracy is shown in Fig. 3. The sparsity constraint factor plays an important role in the experiment, which produces less

sparse codes when set too small, and makes the dictionary no longer sparse and more time-consuming when set too large. The neighborhood size is also an important parameter in the spatial regularization network. As shown in Fig. 3, a too large N may cause over-smoothing and produce noisy information, while a too small N cannot preserve enough spatial features. N is set as 8×8 throughout the experiment to achieve the best accuracy.
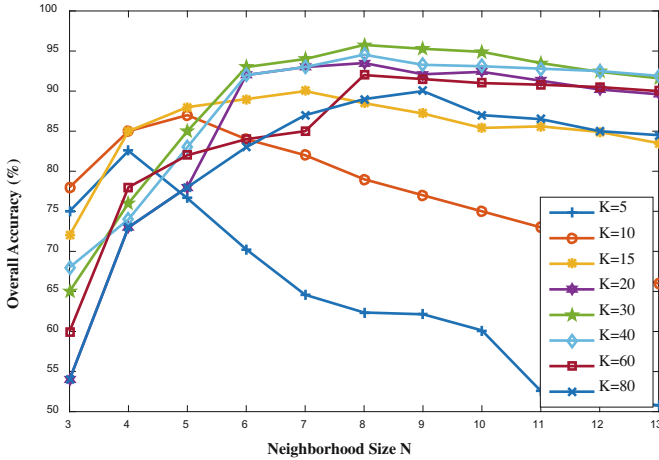

Fig. 3. The effect of neighborhood size N and sparsity constraint factor K

We apply SRC, ERAP, ERAP_MRF and ESRC with different sparsity level factor K, ranging from K = 5 to K = 80, and the best accuracy is chosen for this experiment.

The parameter α which controls the contribution of the classification error is determined by cross validation experiments on training images. For Indian Pines data set, it is set to 0.001. This small value for α is due to the high dimension sparse vector of the training samples. The normalized large scale sparse vector results in small

Table 3. Class information for Pavia University image

Class No.	Class Name	Train	Test
1	Asphalt	50	6631
2	Meadows	50	18649
3	Gravel	50	2099
4	Trees	50	3064
5	Meta sheets	50	1345
6	Bare soil	50	5029
7	Bitumen	50	1330
8	Bricks	50	3682
9	Shadows	50	947
Total		450	42776

Table 4. Classification accuracies (%) for Pavia University image

Method	SRC	KSVM	ERAP	EKSVM	SVM_AP	ERAP_MRF	ESRC
1	92.76	92.24	94.42	95.95	96.60	97.70	97.86
2	97.06	96.79	98.09	98.74	98.89	99.21	99.53
3	71.93	70.74	79.57	85.31	86.89	90.56	93.00
4	51.47	50.05	59.82	66.28	69.76	75.67	81.29
5	91.64	91.52	94.31	95.79	94.84	97.18	98.61
6	56.74	55.24	64.90	71.92	74.32	80.09	85.59
7	75.62	74.54	81.42	86.15	87.71	89.82	93.22
8	83.58	82.59	88.32	91.29	92.89	94.25	96.32
9	99.89	99.89	99.89	99.89	99.89	100.00	100.00
OA	79.32	78.14	84.99	88.87	90.27	92.84	94.98
AA	80.08	79.29	84.53	87.93	89.09	91.61	93.94
k	74.20	72.83	80.97	85.74	87.47	90.71	93.44

component values for the extracted features, therefore the weight of this controlling term is lower, compared to the sparse reconstruction errors.

3.3   Experiments on ROSIS Pavia University Data Set, Italy

The image was collected by the Reflective Optics Systems Imaging spectrometer (ROSIS), and the sensor generates 115 spectral bands covering from 0.43 to 0.86 µm with a spatial resolution of 1.3 m. In our experiments, 103 bands are used with 12 noisy
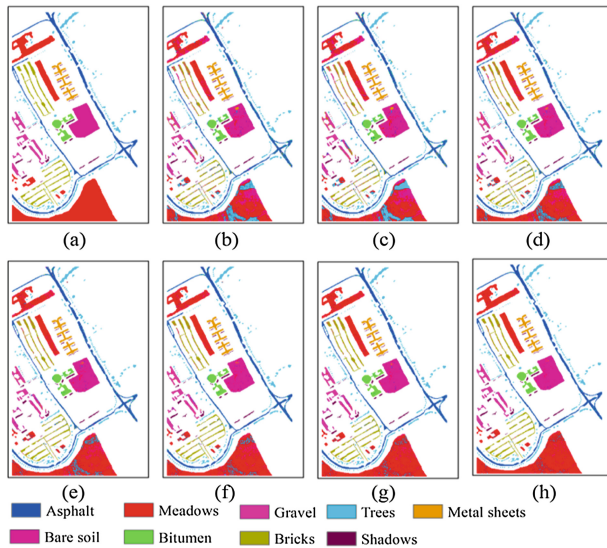
Fig. 4. Classification maps of Pavia University. (a) Ground truth. (b) SRC. (c) KSVM. (d) ERAP. (e) EKSVM. (f) SVM_AP (g) ERAP_MRF. (h) ESRC

bands removed from both data sets. It consists of 610×340 pixels with 9 labelled classes. As discussed above, the sparsity constraint factor T is set to range from 5 to 80 for SRC, ERAP, ERAP_MRF and ESRC. N is set as a 9×9 and α is set as 0.005 for this dataset. Table 3 shows the class information of this data set.

The classification accuracies and classification maps are summarized in Table 4 and Fig. 4. The classification results for this imagery are consistent with Indian Pines imagery. Our proposed method (i.e. ESRC) performs better than the other methods in most cases.

4   Conclusion

In this paper, a novel framework is proposed for HSI classification. This framework is based on EMAP and SRC. HSI pixels are considered as sparse representation by the atoms in a selected dictionary. In the proposed algorithm, a classification error term is added to the SRC model and is solved by the K-SVD algorithm. To improve the classification accuracy, we also have taken into account the influence of neighboring pixels of the pixel of interest in the MAP spatial regularization model. Experiments conducted on two different hyperspectral images show that the proposed method yields high classification accuracy, which is especially better than the state-of-the-art SVM classifiers.

References

1. Chang, C.-I.: Hyperspectral Data Exploitation: Theory and Applications. Wiley, New York (2007)
2. Camps-Valls, G., et al.: Advances in hyperspectral image classification: earth monitoring with statistical learning methods. IEEE Signal Process. Mag. **31**(1), 45–54 (2014)
3. Liang, J., et al.: On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **55**(2), 862–880 (2017)
4. Aghighi, H., et al.: Dynamic block-based parameter estimation for MRF classification of high-resolution images. IEEE Geosci. Remote Sens. Lett. **11**(10), 1687–1691 (2014)
5. Dalla Mura, M., et al.: Extended profiles with morphological attribute filters for the analysis of hyperspectral data. Int. J. Remote Sens. **31**(22), 5975–5991 (2010)
6. Song, B., et al.: Remotely sensed image classification using sparse representations of morphological attribute profiles. IEEE Trans. Geosci. Remote Sens. **52**(8), 5122–5136 (2014)
7. Dalla Mura, M., et al.: Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci. Remote Sens. Lett. **8**(3), 542–546 (2011)
8. Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
9. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Trans. Image Process. **17**(1), 53–69 (2008)
10. Wu, Z., et al.: Exemplar-based sparse representation with residual compensation for voice conversion. IEEE/ACM Trans. Audio Speech Lang. Process. **22**(10), 1506–1521 (2014)

11. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press (2008)
12. Wright, J., et al.: Sparse representation for computer vision and pattern recognition. Proc. IEEE **98**(6), 1031–1044 (2010)
13. Ye, M., et al.: Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **55**(3), 1544–1562 (2017)
14. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
15. Jiang, Z., Lin, Z., Davis, L.S.: Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1697–1704 (2011)
16. Chen, S., Billings, S.A., Luo, W.: Orthogonal least squares methods and their application to non-linear system identification. Int. J. Control **50**(5), 1873–1896 (1989)
17. Li, J., Zhang, H., Zhang, L.: Supervised segmentation of very high resolution images by the use of extended morphological attribute profiles and a sparse transform. IEEE Geosci. Remote Sens. Lett. **11**(8), 1409–1413 (2014)
18. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)