# EFFICIENT SUPERPIXEL-ORIENTED MULTI-TASK JOINT SPARSE REPRESENTATION CLASSIFICATION FOR HYPERSPECTRAL IMAGERY

*Jiayi Li, Hongyan Zhang, and Liangpei Zhang**

The State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, P. R. China
Corresponding author: zlp62@whu.edu.cn

**ABSTRACT**

With regard to the specific role of each pixel within a spatial parcel of a hyperspectral image (HSI), we propose a novel superpixel-oriented sparse representation classification method with a multi-task learning approach. The proposed algorithm exploits the class-level sparsity prior for multiple-feature fusion, and also the correlation and distinctiveness of pixels in a spatial local region. Compared with the state-of-the-art hyperspectral classifiers, the superiority of the spatial prior utilization, the multiple-feature fusion, and the computational efficiency are maintained at the same time in the proposed method. The proposed classification framework was tested on two HSIs. The experimental results suggest that the proposed algorithm performs better than the other representation-based classification algorithms and some popular hyperspectral multiple-feature classifiers.

***Index Terms***—Joint sparsity, multi-task learning, superpixel representation, hyperspectral image classification

## 1. INTRODUCTION

Hyperspectral images (HSIs), spanning the visible to infrared spectrum with hundreds of contiguous and narrow spectral bands, are favored by subtle discriminative spectral characteristics, as well as spatial information. With such fine discriminative information, supervised classification, which labels each pixel in the image with a given training set from each class, is an important task for the subsequent processing and analysis.

The framework of supervised classification consists of two procedures: discriminative feature extraction and classifier design. For the first issue, techniques in both the spectral and spatial domains have been studied in recent years, and it is believed that there is no optimal single feature for the various classification tasks [1]. Therefore, some multiple-feature HSI classifiers have been proposed to combine the complementary features and enhance the discriminability with state-of-the-art performance. When constructing an HSI classifier, the lack of training samples, the Hughes phenomenon, and the high computational burden caused by the "curse of high dimensionality" are inevitable obstacles [2].

Considering the above problems, we propose a superpixel-oriented multi-task joint sparse representation classification (SMTJSRC) algorithm for HSIs, which exploits the joint sparsity prior information for multiple-feature fusion, and the correlation and distinctiveness of pixels in a spatial local region. The algorithm is implemented in the following steps, as shown in Fig. 1. Firstly, several complementary features of the HSI are constructed. Secondly, superpixels, which can be regarded as small local regions with an adaptive shape and size, are acquired by partitioning the whole HSI scene. Thirdly, the multiple-feature joint sparse linear regression model is extended in a set-to-set collaborative representation (CR) manner to obtain the coding coefficient vector for the subsequent classification decision. The proposed method aims at enhancing the discrimination of superpixels by combining the complementary information of different features and highlighting the major patterns of pixels within a spatial similar region in a multi-task learning (MTL) fusion to achieve an improved classification performance.

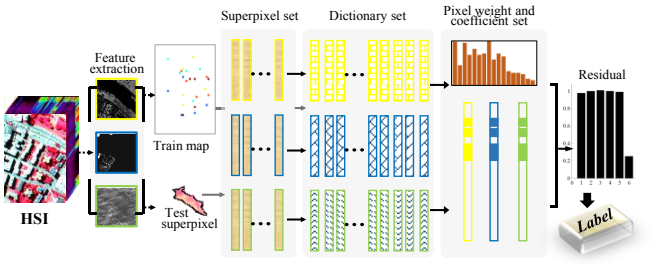


Fig. 1. Schematic illustration of the proposed work. Multiple modalities of the features are first extracted for the whole hyperspectral scene. A dictionary set and a superpixel set both containing multiple features are constructed. Each coefficient vector is represented as a linear combination of the corresponding training feature dictionary. To preserve the invariant similarities between various features, a multi-task joint sparsity norm, which enforces the joint selection of a few common classes of training samples to represent a test superpixel over each feature and each instance, is introduced. Meanwhile, the practical instance of the convex hull can be simultaneously learned with the bounded coefficient (known as the pixel weight in this framework). Finally, the classification decision is made according to the reconstruction error of each individual class.

## 2. PROPOSED FRAMEWORK

### 2.1. Superpixel Segmentation and Distance Definition

A superpixel in a scene can be defined as a pure perceptual uniform parcel. The superpixel segmentation method utilized

here is based on graph partitioning and the entropy rate [3], and has an efficient computational complexity approximated as $O(|V|\log|V|)$, where $V$ refers to the number of pixels in the scene. The only free parameter $T$ (the number of superpixels) controls the segmentation scale of the scene. The first principal component (PC) of the HSI, which maintains the most important information of the whole scene, is utilized as the base image for the superpixel segmentation.

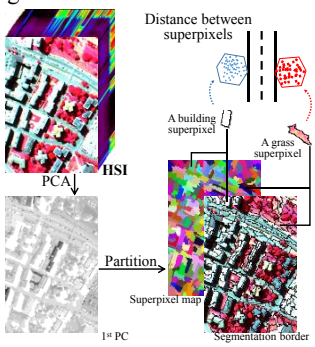


Fig. 2. The procedure of the superpixel segmentation and the hull-based superpixel-to-superpixel distance.

As shown in Fig. 2, a parcel belonging to the grass class can be clustered as a superpixel, and denoted as $\mathbf{Y}=\{\boldsymbol{y}_1,\ldots\boldsymbol{y}_i,\ldots,\boldsymbol{y}_g\}$, where $\boldsymbol{y}_i\in\mathbb{R}^d$, $d$ is the dimension of the feature, and $g$ is the number of pixels in the parcel. The hull of set $\mathbf{Y}$ is defined as $H(\mathbf{Y})=\{\sum a_i\boldsymbol{y}_i\}$. Usually, $\sum a_i=1$ is required to be bounded:

$$H(\mathbf{Y})=\{\sum a_i\boldsymbol{y}_i\,|\sum a_i=1,0\le a_i\le\tau\} \tag{1}$$

In the same scene, another superpixel belonging to the building class can also be denoted as a sample set $\mathbf{Z}=\{\boldsymbol{z}_1,\ldots\boldsymbol{z}_i,\ldots,\boldsymbol{z}_l\}$, where $\boldsymbol{z}_i\in\mathbb{R}^d$, and $l$ is the number of pixels in the building superpixel. The distance between two superpixels is defined by modeling each set as a convex set:

$$\begin{aligned}&\min_{a,b}\left\|\sum a_i\boldsymbol{y}_i-\sum b_i\boldsymbol{z}_i\right\|_2^2\\&\text{s.t. }\sum a_i=1,0\le a_i\le\tau;\ \sum b_i=1,0\le b_i\le\tau\end{aligned} \tag{2}$$

For HSI classification, it is believed that the over-segmentation procedure usually ensures that there is no intersection between superpixels of different classes. Therefore, (2) denotes the distance between two support hyperplanes, which is equivalent to the distance measured by support vector machine (SVM) [1]. That is to say, minimizing the distance can be transformed into maximizing the margin between the two convex hulls.

### 2.2. Superpixel Collaborative Representation

The superiority of the recent popular CR-based hyperspectral classifiers [4, 5] is due to the utilization of the similar training samples from different classes to represent the test pixel. In view of this, it is natural to inherit such a mechanism in the following superpixel-oriented framework.

Suppose we have $M$ distinct classes, then we set $\mathbf{D}_i\in\mathbb{R}^{d\times N_i}$ $i=1,\ldots,M$ as the $i$th sub-dictionary, whose columns are the $N_i$ training samples from the $i$th class, where $N=\sum N_i$, and each $\mathbf{D}_i$ can model a convex set for a specific class. The collaborative dictionary $\mathbf{D}$, which is made up of all the sub-dictionaries $\mathbf{D}_i$, $i=1,\ldots,M$, and is concatenated as a uniform convex set, maps each hyperspectral pixel into a high-dimensional space corresponding to the dictionary. To classify the unlabeled superpixel $\mathbf{Y}$, we model each of $\mathbf{Y}$ and $\mathbf{D}$ as a hull, i.e., $\mathbf{Y}\boldsymbol{a}\in\mathbb{R}^d$ and $\mathbf{D}\boldsymbol{b}\in\mathbb{R}^d$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are coefficient vectors. The CR model of the convex hull of the superpixel $\mathbf{Y}$ can then be expressed as:

$$\mathbf{Y}\boldsymbol{a}=\left[\mathbf{D}_1\boldsymbol{b}_1\ \ldots\mathbf{D}_i\boldsymbol{b}_i\ \ldots\mathbf{D}_M\boldsymbol{b}_M\right]=\mathbf{D}\boldsymbol{b}+\boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{b}_i$ represents the coefficient sub-vector over the $i$th sub-dictionary $\mathbf{D}_i$, and $\boldsymbol{\varepsilon}$ is the random noise.

### 2.3. The Multi-task Learning Algorithm for Superpixel-Oriented Collaborative Representation Classification

Since none of the feature descriptors has the optimal discriminative power for all classes, the MTL approach can further fuse the complementary discriminative abilities of different features by the simultaneous use of the specific learned convex hull of each feature.

Considering each feature as a modality, (3) can be extended as:

$$\begin{aligned}\mathbf{Y}^1\boldsymbol{a}^1&=\left[\mathbf{D}_1^1\boldsymbol{b}_1^1\ \ldots\mathbf{D}_i^1\boldsymbol{b}_i^1\ \ldots\mathbf{D}_M^1\boldsymbol{b}_M^1\right]=\mathbf{D}^1\boldsymbol{b}^1+\boldsymbol{\varepsilon}^1\\&\vdots\qquad\qquad\qquad\qquad\qquad\qquad\vdots\\\mathbf{Y}^K\boldsymbol{a}^K&=\left[\mathbf{D}_1^K\boldsymbol{b}_1^K\ \ldots\mathbf{D}_i^K\boldsymbol{b}_i^K\ \ldots\mathbf{D}_M^K\boldsymbol{b}_M^K\right]=\mathbf{D}^K\boldsymbol{b}^K+\boldsymbol{\varepsilon}^K\end{aligned} \tag{4}$$

where the convex sets of $\mathbf{Y}^k$ are extracted from the different perspectives of the unlabeled hyperspectral superpixel, their corresponding sub-dictionaries $\mathbf{D}^k$ are constructed with the features of the same training samples, and $K$ is the number of modalities. $\boldsymbol{a}^k$, $\boldsymbol{b}^k$, and $\boldsymbol{\varepsilon}^k$ ($k=1,\ldots,K$) are the pixel weight set, collaborative coefficient set, and random noise set, respectively.

To make the representation step stable, the MTL-based framework also utilizes regularizations as:

$$\begin{aligned}&\min_{\{\boldsymbol{a}^k,\boldsymbol{b}^k\}}\sum_k^K\left\|\mathbf{Y}^k\boldsymbol{a}^k-\mathbf{D}^k\boldsymbol{b}^k\right\|_2^2\\&\text{s.t. }\sum a_i^k=1,\ \left\|\boldsymbol{a}^k\right\|_{\ell_p}<\delta_1,\ \left\|\boldsymbol{b}^k\right\|_{\ell_q}<\delta_2,\ k=1,\ldots,K\end{aligned} \tag{5}$$

To combine the discriminative abilities of the multiple features for classification, a joint learning procedure is utilized for (5) by imposing a class-level sparsity-inducing term on $\boldsymbol{b}^k$, $k=1,\ldots,K$. Similar to the motivation behind CR-based classifiers, it is useful to jointly select a few common classes of training samples to represent a test convex hull over each feature. That is to say, the desired representation vectors for the multiple features should share certain class-level sparsity patterns. Given the optimal superpixel convex hull $\mathbf{Y}^k\boldsymbol{a}^k$, the CR coefficient vector $\boldsymbol{b}^k$ can be rewritten as $\boldsymbol{b}^k=\left[\boldsymbol{b}_1^{k\prime},\cdots,\boldsymbol{b}_M^{k\prime}\right]$, in which $\boldsymbol{b}_i^k$ consists of the components of $\boldsymbol{b}^k$ restricted on class $i$. We stack all the CR vectors together, and $\mathbf{B}_i=\left[\boldsymbol{b}_i^1,\cdots,\boldsymbol{b}_i^K\right]$ denotes the representation coefficients associated with class $i$ across the different features. Inspired by the sparsity constraint utilized in sparse representation classification (SRC) [5] and the role of the multiple features, the class-level joint sparsity-inducing term, which applies the $\ell_0$-norm across the $\ell_2$-norm of $\mathbf{B}_i$, can be shown as $\left\|\left[\|\mathbf{B}_1\|_F,\ldots,\|\mathbf{B}_M\|_F\right]\right\|_0$, and relaxed as $\left\|\left[\|\mathbf{B}_1\|_F,\ldots,\|\mathbf{B}_M\|_F\right]\right\|_1$.

Considering the homogeneities and the similarities of the pixels within a parcel, the $\ell_2$-norm regularization is available to make the problem stable, and with a light computational complexity. We can rewrite the regularized MTL model in (5) as its Lagrangian formulation:

$$L\left(\mathbf{A},\mathbf{B},\left\{\gamma^k\right\}_{k=1,\ldots,K}\right)=\sum_k^K\left\|\mathbf{Y}^k\boldsymbol{a}^k-\mathbf{D}^k\boldsymbol{b}^k\right\|_2^2+\lambda_1\left\|\mathbf{A}\right\|_F+\eta\left\|\mathbf{B}\right\|_{1,2}+\sum_k^K\left\langle\gamma^k,\boldsymbol{e}\boldsymbol{a}^k-1\right\rangle \quad (6)$$

where $\mathbf{A}=\left[\boldsymbol{a}^1,\ldots,\boldsymbol{a}^K\right]$, $\mathbf{B}=\left[\boldsymbol{b}^1,\ldots,\boldsymbol{b}^K\right]$, $\lambda$ and $\eta$ are positive constants to balance the data fidelity term and the regularizations, $\gamma^k$, $k=1,\ldots,K$ is the Lagrange multiplier set, $\langle\bullet,\bullet\rangle$ is the inner product, and $\boldsymbol{e}$ is a row vector whose elements are 1. As a convex optimization problem with two variables, (6) can be solved by alternating the optimization with the two corresponding sub-problems, until the solutions converge to a minimum.

For the first sub-problem, we optimize $\mathbf{A}$ by fixing $\mathbf{B}$, and the optimization of (6) becomes:

$$L\left(\boldsymbol{a}^k,\gamma^k\right)=\left\|\mathbf{Y}^k\boldsymbol{a}^k-\mathbf{D}^k\boldsymbol{b}^k\right\|_2^2+\lambda\left\|\boldsymbol{a}^k\right\|_2^2+\left\langle\gamma^k,\boldsymbol{e}\boldsymbol{a}^k-1\right\rangle \quad (7)$$

There is a closed-form solution to (7):

$$\gamma^k=2\left(\boldsymbol{e}\boldsymbol{p}_k-1\right)/\boldsymbol{e}\mathbf{Q}_k^{-1}\boldsymbol{e}',\ \boldsymbol{a}^k=\boldsymbol{p}_k-0.5\gamma^k\mathbf{Q}_k^{-1}\boldsymbol{e}' \quad (8)$$

where $\mathbf{Q}_k=\left(\mathbf{Y}^{k\prime}\mathbf{Y}^k+\lambda\right)^{-1}$ and $\boldsymbol{p}_k=\mathbf{Q}_k\mathbf{Y}^k\boldsymbol{x}^k$.

For the second sub-problem, we optimize $\mathbf{B}$ by fixing $\mathbf{A}$, and the optimization of (6) becomes:

$$\min_{\mathbf{B}}\sum_k^K\left\|\boldsymbol{s}^k-\mathbf{D}^k\boldsymbol{b}^k\right\|_2^2+\eta\left\|\mathbf{B}\right\|_{1,2}\ \text{s. t. }\boldsymbol{s}^k=\mathbf{Y}^k\boldsymbol{a}^k \quad (9)$$

The problem of (9) is known as the multi-task joint covariate selection model in sparse learning, and can be efficiently solved within several iterations by the accelerated proximal gradient (APG) method [6]. As discussed in [7], the alternating minimization approach for such a general convex problem will converge to the correct solution, as both sub-problems are convex. As in the MTL approaches [4], the label of the unlabeled superpixel is then determined by the minimal total residual:

$$\text{class}\left(\mathbf{Y}\right)=\arg\min_{i=1,\ldots,M}\sum_{k=1}^K\left\|\mathbf{Y}^k\boldsymbol{a}^k-\mathbf{A}_i^k\boldsymbol{b}_i^k\right\|_2^2 \quad (10)$$

The computational burden for the proposed SMTJSRC algorithm is as follows. The running time for the first sub-problem is $O\left(Kd_{\max}^2\right)$, where $d_{\max}$ is the maximum of the dimension of the multiple features, as $\mathbf{Q}$ and its inverse can be pre-computed. The second sub-problem takes $O\left(\left(2L_1+1\right)Kd_{\max}\right)$, where $L_1$ is the average iteration times utilized to solve (9). It is also notable that the computational load of (10) is negligible in the superpixel MTL procedure. Finally, $L_2$ is the average iteration times for the alternating optimization of the second sub-problem, and the computational complexity for labeling the whole hyperspectral scene is $O\left(TL_2\left(\left(2L_1+1\right)Kd_{\max}+Kd_{\max}^2\right)+|V|\log|V|\right)$.

## 3. EXPERIMENTS AND ANALYSIS

### 3.1. Data Sets

The first HSI scene was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in June 12, 1992, over the Indian Pines test site in North-Western Indiana, and consists of $145\times145$ pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.0 $\mu m$. We used the 200 available bands in following procedure. The spatial resolution for this image is about 20 m. This image contains 10 ground-truth classes, and the numbers of the reference data and the corresponding visual map are shown in Table II and Fig. 3(b), respectively.

The second scene was acquired by the Reflective Optics Systems Imaging Spectrometer (ROSIS) sensor over Pavia University, northern Italy. We selected 103 of the bands and cut a patch sized $610\times340$. The geometric resolution of this image is 1.3 m. This image contains nine reference classes, and the numbers of the reference data and the corresponding visual map are shown in Table III and Fig. 4(b), respectively.

### 3.2. Experimental Setting

For the multiple-feature extraction, we utilized three meaningful features: the spectral value feature, the Gabor texture feature, and the differential morphological profile (DMP) feature. The discriminative and complementary nature of these three features for HSI classification has been detailed in Li *et al.* [4].

TABLE I
CLASSIFICATION APPROACHES IN THE COMPARISON

| | | | |
|---|---|---|---|
| SVM-VS | 1) SVM with RBF kernel<br>2) Nonlinear<br>3) Vector stacking (VS) | SVM-CK [8] | 1) SVM with RBF kernel<br>2) Nonlinear<br>3) Composite kernel (CK) |
| GCK-MLR [8] | 1) MLR<br>2) Nonlinear<br>3) Generalized CK (GCK) | MNFL [9] | 1)Regularization-free MLR<br>2) Nonlinear<br>3) Vector stacking (VS) |
| SRC-MTL [4] | 1) CR<br>2) Linear<br>3) Multiple task learning:<br>Class-level sparsity prior | JCRC-MTL [4] | 1) CR<br>2) Linear<br>3) Multiple task learning:<br>a. Similarity for labeling<br>b. Dissimilarities across features<br>c. Spatial homogeneity |

Based on these multiple features, a number of the state-of-the-art multiple-feature classification algorithms were taken as benchmarks, as illustrated in Table I in detail. In this table, the first line of each item refers to the basic classification principle, the second line categories the classifier as linear or not, and the last line represents the manner with which the classifier combines the multiple features.

For each dataset, we randomly selected 30 pixels for each class as the training samples, and the rest as the test samples from the reference data to validate the performances. Each parameter was selected by cross-validation from a reasonable range. The classification accuracies were averaged over 10 runs for each classifier to reduce the possible bias induced by the random sampling. All the experiments, except for the SVM-related work (accelerated by C++), were carried out using MATLAB on a PC with one 3.10 GHz processer and 8.0 Gb of RAM.

### 3.3. Experimental Results

The thematic maps of the various classifiers are visually shown in Figs. 3–4. In the quantitative evaluation tables, the average accuracy for each class, the average overall accuracy (OA), and the average kappa coefficient ($\kappa$) with their standard deviations for the different classifiers are shown in sequence. In addition, the average running times for the labeling of the whole scene are shown in the bottom line of the quantitative evaluation tables.

For the homogenous Indian Pines image, the inferior accuracy of SRC-MTL, which is even a bit weaker than the direct feature stacking approach (VS), suggests the limitations of the pixel-oriented methods without contextual information. For such an image with subtle spectral differences between classes, the spatial prior can effectively stabilize the signal and alleviate the "salt-and-pepper" phenomenon, as both the quantitative evaluation and the thematic map indicate. For the two classifiers with spatial information, it can be observed from Fig. 3 that JCRC-MTL suffers from the "over-smoothing" problem at the optimal neighboring size, while the proposed SMTJSRC method can effectively preserve the edge information in most cases.

For the second hyperspectral dataset with a high spatial resolution, the shape and texture features are more meaningful, and can effectively improve the classification result. Based on the complementary multiple features, different learning methods for utilizing these features significantly affect the discriminability. Firstly, the VS-based classifiers are inferior to the others, and the proposed MTL mechanism shows the best performance, which is consistent with the classification results of the Indian Pines image. Secondly, utilizing the contextual information with a regular pattern (i.e., stacking neighboring pixels to the test pixel together with an equal weight) is still useful, as each parcel in the scene covers tens of pixels at least. The tuned optimal neighboring size of JCRC-MTL for classifying the Pavia University image is small, and the "over-smoothing" problem is not significant. Therefore, JCRC-MTL is significantly superior to SRC-MTL in this image. All in all, it can be seen that the proposed spatial prior related SMTJSRC algorithm shows a superior performance.

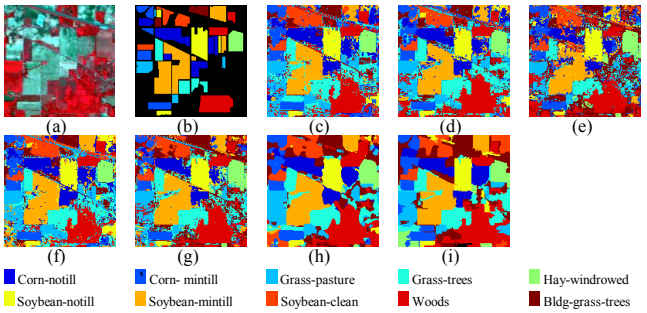


Fig. 3. Classification results for the Indian Pines image: (a) false-color image (R: 57, G: 27, B: 17), (b) reference set, (c) SVM-VS, (d) SVM-CK, (e) GCK-MLR, (f) MNFL, (g) SRC-MTL, (h) JCRC-MTL, and (i) SMTJSRC.

TABLE II
REFERENCE INFORMATION, CLASSIFICATION ACCURACY (%), AND RUNNING TIME (SECONDS) FOR THE INDIAN PINES IMAGE WITH THE TEST SET

| CLASS Name | Num. | SVM-VS | SVM-CK | GCK-MLR | MNFL | SRC-MTL | JCRC-MTL | SMTJSRC |
|---|---|---|---|---|---|---|---|---|
| Corn-notill | 1428 | 84.08 | 85.33 | 84.82 | 82.55 | 78.21 | 85.77 | **88.44** |
| Corn-mintill | 830 | 94.09 | 95.14 | 93.60 | 91.86 | 92.23 | 94.13 | **99.05** |
| Grass-pasture | 483 | 88.48 | 93.36 | **91.81** | 91.30 | 87.79 | 91.79 | 87.97 |
| Grass-trees | 730 | 93.57 | 99.53 | 99.14 | 98.47 | 99.63 | 99.69 | **100.0** |
| Hay-windrowed | 478 | 99.53 | 99.96 | 99.80 | 99.46 | 99.98 | **100** | 99.84 |
| Soybean-notill | 972 | 83.48 | 86.94 | **87.39** | 83.97 | 82.98 | 85.89 | 87.20 |
| Soybean-mintill | 2455 | 84.16 | 85.14 | 87.81 | 85.93 | 85.42 | 89.56 | **92.45** |
| Soybean-clean | 593 | 91.37 | 91.60 | 91.85 | 91.67 | 93.85 | 98.53 | **99.13** |
| Woods | 1265 | 96.33 | 98.60 | 98.97 | 97.72 | 96.54 | **99.75** | 98.11 |
| Bldg-grass-trees | 386 | 96.43 | 96.83 | 98.76 | 96.26 | 98.15 | **99.27** | 98.15 |
| OA | | 89.10 ±1.84 | 91.02 ±1.65 | 91.58 ±1.60 | 89.89 ±1.55 | 89.03 ±1.77 | 92.65 ±0.77 | **93.96 ±0.60** |
| $\kappa$ | | 87.39 ±2.08 | 89.60 ±1.89 | 90.25 ±1.82 | 88.29 ±1.78 | 87.32 ±2.01 | 91.48 ±0.90 | **92.99 ±0.70** |
| Time | | 2.1 | 14 | 1.8 | 0.7 | 50.5 | 407.8 | 3.33 |

For the running time comparison, it can be seen that MNFL is the fastest, and SVM-VS, SVM-CK, GCK-MLR, and the proposed SMTJSRC are comparable but a bit slower than the former, and the other two classifiers are the slowest. Here, it can be concluded that the proposed SMTJSRC is much more efficient than the other two MTL-based classifiers, with a superior classification accuracy at the same time.

## 4. CONCLUSION

In this paper, we have proposed an efficient superpixel-oriented multi-task joint sparse representation classification algorithm for hyperspectral imagery. Therein, an HSI superpixel is represented by an adaptive combination of the pixels in a parcel, and class-level sparsity is utilized to simultaneously integrate the multiple features into a uniform classification framework. The main advantage of the proposed SMTJSRC is that the superiorities of the multiple-feature fusion strategy, the spatial prior utilization, and the light computational complexity can be maintained at the same time. The extensive experimental results clearly indicate that the proposed method is computationally efficient and achieves a competitive classification performance.

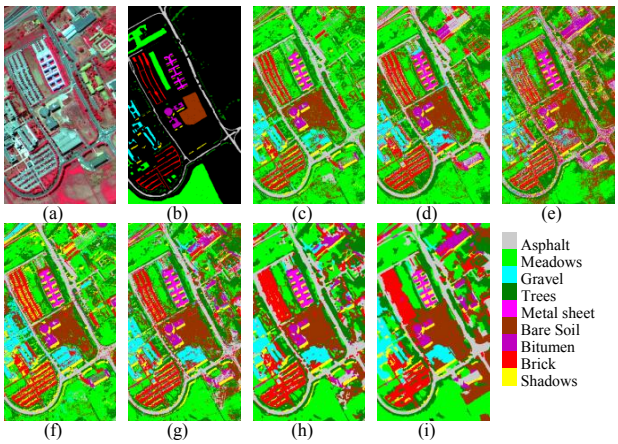


Fig. 4. Classification results for the Pavia University image: (a) false-color image (R: 102, G: 56, B: 31), (b) reference set, (c) SVM-VS, (d) SVM-CK, (e) GCK-MLR, (f) MNFL, (g) SRC-MTL, (h) JCRC-MTL, and (i) SMTJSRC.

TABLE III
REFERENCE INFORMATION, CLASSIFICATION ACCURACY (%), AND RUNNING TIME (SECONDS) FOR THE PAVIA UNIVERSITY IMAGE WITH THE TEST SET

| CLASS Name | Num. | SVM-VS | SVM-CK | GCK-MLR | MNFL | SRC-MTL | JCRC-MTL | SMTJSRC |
|---|---|---|---|---|---|---|---|---|
| Asphalt | 6631 | 93.74 | 93.66 | 95.05 | 93.96 | 91.22 | **97.65** | 96.69 |
| Meadows | 18649 | 79.71 | 86.51 | 87.88 | 84.81 | 90.32 | 93.41 | **95.79** |
| Gravel | 2099 | 95.77 | 95.67 | 97.55 | 96.17 | 96.49 | 97.96 | **98.66** |
| Trees | 3064 | 96.43 | 96.29 | 96.72 | 97.12 | **98.67** | 96.80 | 92.44 |
| Metal sheet | 1345 | 99.76 | 99.55 | 99.56 | 99.81 | 99.98 | **100** | 99.24 |
| Bare soil | 5029 | 92.28 | 94.58 | 94.65 | 94.95 | 94.79 | 98.69 | **99.59** |
| Bitumen | 1330 | 95.35 | 97.45 | 97.14 | 97.78 | 97.18 | **99.69** | 98.66 |
| Brick | 3682 | 94.72 | 96.65 | 97.87 | 97.28 | 97.23 | **98.93** | **98.93** |
| Asphalt | 6631 | 98.76 | 98.85 | 98.69 | 97.04 | **99.30** | 98.57 | 93.54 |
| OA | | 88.14 ±2.05 | 91.63 ±1.70 | 92.67 ±2.04 | 91.09 ±1.94 | 93.18 ±1.65 | 96.13 ±1.33 | **96.49 ±0.90** |
| $\kappa$ | | 84.78 ±2.57 | 89.18 ±2.12 | 90.49 ±2.57 | 88.52 ±2.42 | 91.12 ±2.07 | 94.94 ±1.70 | **95.39 ±1.16** |
| Time | | 14.8 | 13 | 35.20 | 4.1 | 368.3 | 2944.2 | 31.63 |

## 5. REFERENCES


[1] 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, 2004.

[2] A. Plaza, J. A. Benediktsson, J. W. Boardman *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.,* vol. 113, pp. S110-S122, 2009.

[3] M. Liu, O. Tuzel, S. Ramalingam *et al.*, "Entropy rate superpixel segmentation," *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2097-2104, 2011.

[4] J. Li, H. Zhang, L. Zhang *et al.*, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 9, pp. 5923 - 5936, 2014.

[5] J. Wright, A. Y. Yang, A. Ganesh *et al.*, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 31, no. 2, pp. 210-227, 2009.

[6] X. Yuan, X. Liu, and S. Yan, "Visual classification with multi-task joint sparse representation," *IEEE Trans. Image Process.,* vol. 21, no. 10, pp. 4349-4360, 2012.

[7] A. Gunawardana, and W. Byrne, "Convergence theorems for generalized alternating minimization procedures," *The Journal of Machine Learning Research,* vol. 6, pp. 2049-2073, 2005.

[8] J. Li, P. R. Marpu, A. Plaza *et al.*, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 9, pp. 4816- 4829 2013.

[9] J. Li, X. Huang, P. Gamba *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 53, no. 3, pp. 1592- 1606, 2014.