

# Spatial-Spectral Metric Learning for Hyperspectral Remote Sensing Image Classification

Jiangtao Peng<sup>a,b</sup>, Yicong Zhou<sup>b,\*</sup>, C. L. Philip Chen<sup>b</sup>

<sup>a</sup>Faculty of Mathematics and Statistics, Hubei University, Wuhan, China

<sup>b</sup>Department of computer and Information Science, University of Macau, Macau, China

## ABSTRACT

A spatial-spectral metric learning (SSML) framework for hyperspectral image (HSI) classification is proposed. SSML learns a metric by considering both the spectral characteristics and spatial features represented as the mean of neighboring pixels. It first performs the local pixel neighborhood preserving embedding (LPNPE) to reduce the dimensionality of HSI and meanwhile to preserve the spatial local similarity structure. Then, it learns a spectral and spatial distance metric, separately. Finally, the combination of the spectral and spatial metrics yields a joint spatial-spectral metric. It is followed by a nearest neighbor (NN) classifier for HSI classification. SSML shows good performance over the spectral and spatial NN and SVM on the benchmark hyperspectral data set of Indian Pines.

**Keywords:** Hyperspectral image, metric learning, classification, dimension reduction

## 1. INTRODUCTION

Hyperspectral image (HSI) classification receives increasing interests recently due to the discriminant ability of hyperspectral data with a high spatial and spectral resolution.<sup>1,2</sup> Traditional HSI classification methods generally perform on the spectral features based on the fact that the samples in the same class have similar spectral characteristics. By measuring the similarity distance among different spectral curves, they can discriminate and classify different samples. For example, k-nearest neighbor (k-NN) classifier and support vector machine (SVM) perform the HSI classification based on the Euclidean distance and kernel distance, respectively. The success of these methods depends on designing an approximate distance similarity metric (e.g., Euclidean distance or kernel function).

Metric learning directly learns a distance metric from the given samples such that the samples in the same class have small distances while the samples from different classes are as far as possible.<sup>3</sup> Many methods have been proposed to learn distance metrics. E. Xing et al. put metric learning into a convex programming framework with the objective function of minimizing the pairwise distance between all similar points under the constraint that the dissimilar points are well separated.<sup>3</sup> Weinberger et al. learned a distance metric in a maximum margin framework by semidefinite programming.<sup>4</sup> Other metric learning methods can be referred to Ref. 5. Most of these methods focus on learning a Mahalanobis distance metric for k-NN classification.

In this paper, we perform the distance metric learning for k-NN classification of hyperspectral images. Because HSIs have both the spectral and spatial features, we propose a spatial-spectral metric learning (SSML) framework for HSI classification. Rather than using only spectral characteristics as in traditional metric learning methods, SSML learns the metric by considering both the spectral features and spatial neighboring pixel relations.

Considering the high HSI dimensionality, we first perform the local pixel neighborhood preserving embedding (LPNPE) to reduce the dimensionality and meanwhile to preserve the local similarity structure of HSIs. Then, we represent each HSI pixel as a spectral vector and a spatial vector simultaneously. The spatial vector is extracted from the spatial local pixel neighborhood, i.e., the mean or standard deviation of neighboring pixels. Based on the constructed spectral and spatial features, we can learn the spectral and spatial distance metrics, separately. Finally, integrating the spectral and spatial metrics yields a spatial-spectral distance metric. It is followed by a nearest neighbor (NN) classifier for HSI classification. In learning the distance metric, we generate

---

E-mail: yicongzhou@umac.mo. Telephone: +853-83978458

positive and negative sample pairs from the  $k$  nearest neighbors in the training set, and learn a Mahalanobis distance metric by using a support vector machine (SVM) solver on the generated sample pairs.<sup>6</sup>

The rest of this paper is organized as follows. In Section 2, a simple metric learning model is introduced. In Section 3, the proposed spatial-spectral metric learning method is described in detail. The experimental results and analysis are provided in Section 4. Finally, Section 5 gives a summary of our work.

## 2. METRIC LEARNING MODEL

Given a set of training samples  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  with labels  $\{y_i\}_{i=1}^n$ , metric learning aims to learn a positive semi-definite matrix  $M$  associated with a Mahalanobis-like metric distance:

$$d_M(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j) \quad (1)$$

such that the within-class metric distance is minimized while the between-class metric distance is maximized. To achieve this goal, an SVM-like metric learning model is proposed in Ref. 6:

$$\begin{aligned} \min \quad & \|M\|_F^2 + \gamma(\sum_{i,j} \xi_{ij}^N + \sum_i \xi_i^P) \\ \text{s.t.} \quad & d_M(\mathbf{x}_i, \mathbf{x}_j) + b \geq 1 - \xi_{ij}^N, \quad y_j \neq y_i; \\ & d_M(\mathbf{x}_i, \mathbf{x}_k) + b \leq -1 + \xi_i^P, \quad y_k = y_i; \\ & M \succeq 0, \quad \xi_{ij}^N \geq 0, \quad \xi_i^P \geq 0 \end{aligned} \quad (2)$$

where  $\xi_{ij}^N$  and  $\xi_i^P$  are slack variables for the between-class sample pairs (negative pairs) and within-class sample pairs (positive pairs), respectively. This optimization model (2) can be solved by the SVM solver.<sup>6,8</sup>

## 3. SPATIAL-SPECTRAL METRIC LEARNING

Metric learning measures and classifies different samples based on the similarity metric distance directly. For HSI, the high dimensionality largely affects the computational accuracy and processing speed of metric learning, and finally will lead to poor classification performance. To overcome this effect, we propose a local pixel neighborhood preserving embedding (LPNPE) method to reduce the dimensionality of the HSI data.

LPNPE aims to preserve the spatial local pixel neighborhood structures. Given a sample  $\mathbf{x}_i$  with pixel coordinate  $(p_i, q_i)$ , the local pixel neighborhood centered at  $\mathbf{x}_i$  is

$$N(\mathbf{x}_i) = \{\mathbf{x} \triangleq (p, q) | p \in [p_i - a, p_i + a], q \in [q_i - a, q_i + a]\}$$

where  $a = (w - 1)/2$ , the odd number  $w$  is the width of neighborhood window (called the scale). The pixels in the spatial pixel neighborhood  $N(\mathbf{x}_i)$  are denoted as  $\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{is}$ , where  $s = w^2 - 1$  is the number of neighbors of  $\mathbf{x}_i$ . We can define a distance scatter in the local pixel neighborhood as:

$$h_i = \sum_{k=1}^s (\mathbf{x}_i - \mathbf{x}_{ik})(\mathbf{x}_i - \mathbf{x}_{ik})^T$$

Consider all the training samples in the HSI, the local pixel neighborhood preserving scatter matrix is:

$$H = \sum_{i=1}^n h_i = \sum_{i=1}^n \sum_{k=1}^s (\mathbf{x}_i - \mathbf{x}_{ik})(\mathbf{x}_i - \mathbf{x}_{ik})^T$$

Denote the total scatter matrix as

$$S = \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

where  $\mathbf{m}$  is the mean of training samples.

LPNPE seeks a linear projection matrix such that the local pixel neighborhood preserving scatter is minimized while the total scatter is maximized in the projected space. The optimal projection  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_\ell]$  can be obtained by solving the generalized eigenvalue problem:

$$S\mathbf{v} = \lambda H\mathbf{v} \quad (3)$$

Due to the fact that the neighboring pixels in a spatial local homogeneous region consist of the same materials and belong to the same class, the proposed LPNPE can preserve the intrinsic structure of HSI well.

On the dimension-reduced HSI data, we can perform the joint spatial and spectral metric learning. For each sample  $\mathbf{x}_i$ , we denote the corresponding spectral feature vector as  $\mathbf{x}_i^\omega$  and the spatial feature vector as  $\mathbf{x}_i^s$ , where the spatial feature vector is the mean of neighboring pixels in  $N(\mathbf{x}_i)$ . Based on the defined features, we can compute a spectral metric matrix  $M^\omega$  and a spatial metric matrix  $M^s$  according to the model (2). The associated metric distance can be represented as:

$$d_{M^\omega}(\mathbf{x}^\omega, \mathbf{z}^\omega) = (\mathbf{x}^\omega - \mathbf{z}^\omega)^T M^\omega (\mathbf{x}^\omega - \mathbf{z}^\omega) \quad (4)$$

$$d_{M^s}(\mathbf{x}^s, \mathbf{z}^s) = (\mathbf{x}^s - \mathbf{z}^s)^T M^s (\mathbf{x}^s - \mathbf{z}^s) \quad (5)$$

The composite spatial-spectral metric  $d_{M^{s,\omega}}$  is defined as:

$$d_{M^{s,\omega}}(\mathbf{x}, \mathbf{z}) = (1 - \mu)d_{M^\omega}(\mathbf{x}^\omega, \mathbf{z}^\omega) + \mu d_{M^s}(\mathbf{x}^s, \mathbf{z}^s) \quad (6)$$

where  $\mu$  is a weight controlling the relative proportion of the spatial and spectral information in the final metric distance. Note that, a linear combination of two metric matrices also yields a metric matrix. The spatial-spectral metric learning (SSML) framework is shown in Fig. 1. After learning a spatial-spectral metric, the metric distance-based nearest neighbor classifier is used to perform HSI classification.

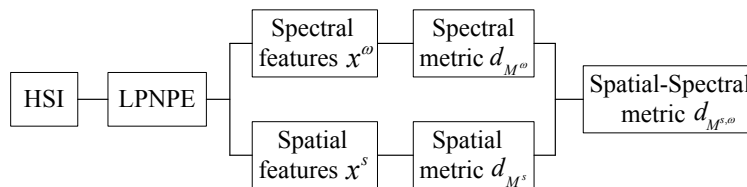


Figure 1. Flowchart of the proposed SSML framework.

## 4. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed SSML framework on the Indian Pines data set. SSML is compared with the spatial and spectral SVMs<sup>7</sup> on the original and LPNPE dimension-reduced data. In the experiments, speML and spaML refer to the metric learning using only the spectral features and spatial features, respectively. Similarly, speSVM and spaSVM represent the spectral and spatial SVM respectively. SSSVM refers to the weighted summation kernel based spatial-spectral SVM. NN refers to the 1-nearest neighbor classifier. In all the spatial method,  $11 \times 11$  neighborhood window is used.

### 4.1 Hyperspectral data set

The Indian Pines data was acquired by the AVIRIS sensor in 1992.<sup>†</sup> The image scene contains with  $145 \times 145$  pixels and 220 spectral bands, where 20 channels were discarded because of atmospheric affection. There are 16 classes in the data. The total number of samples is 10249 ranging from 20 to 2455 in each class. The false color composition of bands 50, 27, and 17 and the ground-truth map are shown in Fig. 2.

<sup>†</sup> [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

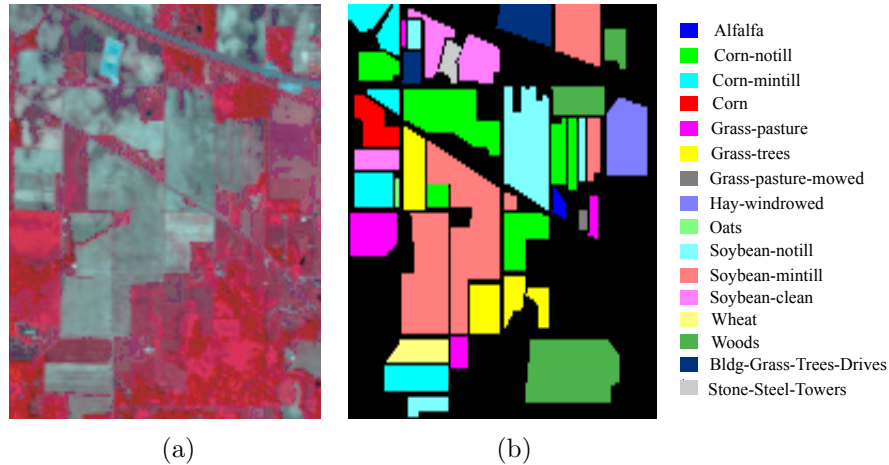


Figure 2. The Indian Pines data set. (a) RGB composite image of three bands 50, 27 and 17. (b) Ground-truth map.

#### 4.2 Investigation on the number of labeled samples per class

We investigate the performance of the proposed SSML under different numbers of labeled samples per class. We randomly choose  $N = 5, 10, 15, 20, 25, 30$  samples from each class to form the training set, respectively (For the class less than  $N$  samples, half of total samples are chosen). The remaining samples compose the testing set. In each case, the experiment is repeated ten times with randomly chosen training samples. The results of ten runs are averaged. The classification overall accuracies (OAs) and kappa coefficients ( $\kappa$ ) under different numbers of training samples are shown in Table 1.

Table 1. Classification accuracies (%) using different numbers of training samples per class (In each case of labeled samples, the first and second rows correspond to the OAs and  $\kappa$  coefficients, respectively).

$N$	Original data				DR data by LPNPE						
	NN	SpeSVM	SpaSVM	SSSVM	NN	SpeSVM	SpaSVM	SSSVM	SpeML	SpaML	SSML
5	46.7±4.34	47.0±2.93	59.8±4.16	60.5±3.77	58.3±3.99	59.8±3.32	67.8±4.21	70.1±2.44	59.2±3.16	70.7±2.59	<b>72.2±2.40</b>
	40.7±4.44	40.8±3.03	55.1±4.45	56.0±4.00	53.4±4.01	55.1±3.30	64.0±4.49	66.5±2.69	54.4±3.21	67.1±2.86	<b>68.8±2.64</b>
10	50.9±2.05	53.8±2.53	70.8±3.33	71.8±2.83	63.7±3.30	65.3±1.95	78.3±5.25	80.4±2.08	67.2±3.15	80.9±1.59	<b>82.2±1.76</b>
	45.3±2.18	48.4±2.59	67.3±3.63	68.4±3.08	59.4±3.51	61.1±2.07	75.6±5.76	77.9±2.28	63.3±3.40	78.4±1.76	<b>79.9±1.97</b>
15	54.9±1.68	60.7±1.80	77.1±1.66	77.7±1.56	68.0±1.44	67.7±1.85	83.1±3.01	84.0±1.35	70.8±2.14	83.1±1.41	<b>84.6±1.90</b>
	49.6±1.76	56.0±1.97	74.2±1.77	75.0±1.67	64.1±1.56	63.8±1.92	80.9±3.45	81.9±1.50	67.2±2.36	80.9±1.55	<b>82.6±2.13</b>
20	55.6±2.34	62.2±2.14	81.2±2.18	81.6±1.89	69.8±1.13	71.3±1.32	87.6±2.07	88.1±1.67	73.2±0.75	86.8±1.05	<b>88.6±1.10</b>
	50.4±2.43	57.5±2.33	78.7±2.44	79.2±2.10	66.0±1.27	67.6±1.35	86.0±2.33	86.5±1.87	69.7±0.80	85.0±1.17	<b>87.0±1.23</b>
25	57.5±1.19	65.0±2.44	82.9±2.52	83.5±1.91	71.1±1.62	71.5±1.16	87.9±2.22	88.8±1.63	75.2±1.21	88.0±1.40	<b>89.3±1.59</b>
	52.4±1.31	60.6±2.92	80.6±2.80	81.4±2.13	67.4±1.71	67.9±1.24	86.3±2.49	87.3±1.81	72.0±1.27	86.4±1.57	<b>87.9±1.77</b>
30	58.6±1.21	68.0±1.40	84.3±1.46	85.1±2.08	72.3±1.41	73.1±1.40	89.8±1.46	90.1±1.44	75.9±1.20	90.1±0.70	<b>90.5±0.81</b>
	53.6±1.31	64.0±1.57	82.2±1.63	83.2±2.33	68.7±1.56	69.7±1.54	88.5±1.64	88.8±1.63	72.8±1.32	88.7±0.79	<b>89.2±0.92</b>

As shown in the table, by reducing the HSI dimensionality, LPNPE coupled with NN and SVM dramatically improve the classification performance on the original data. The metric learning on the spatial data provides higher accuracies than that on the spectral data. By combining the spatial and spectral data, the proposed SSML shows a significant improvement over the single spectral metric learning and the single spatial metric learning,

especially in the cases of extremely limited training samples. In all cases, SSML provides better classification results than the NN and SVM methods on the original and LPNPE dimension-reduced data. The classification overall accuracy of the different methods on different numbers of training samples are shown in Fig. 3.

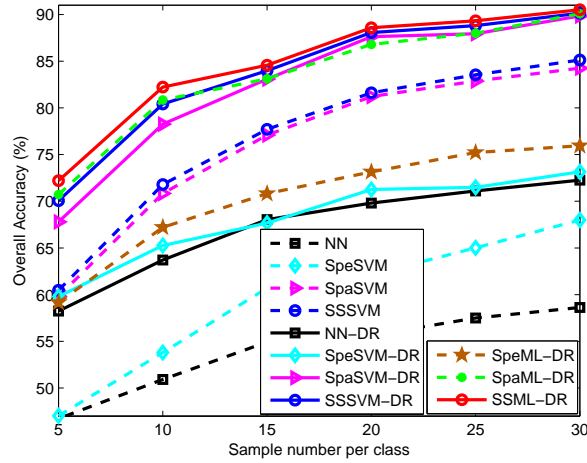


Figure 3. Overall accuracy versus sample number per class.

Table 2. Classification accuracies (%) for the balanced training set (5% labeled samples per class used for training, totally 518 training samples, all the remaining samples used for testing).

C	# Samples		Classification algorithms										
	Train	Test	Original data				DR data by LPNPE						
			NN	SpeSVM	SpaSVM	SSSVM	NN	SpeSVM	SpaSVM	SSSVM	SpeML	SpaML	SSML
1	3	43	34.9±17.9	49.3±15.8	83.3±17.6	83.0±18.9	65.8±12.6	63.0±18.9	<b>94.7±5.03</b>	<b>94.7±5.03</b>	72.3±11.5	71.4±17.5	93.0±8.63
2	71	1357	49.0±4.91	69.2±5.14	89.4±1.91	92.3±2.10	75.5±2.91	78.4±2.71	92.1±2.95	92.0±2.36	79.2±1.99	<b>95.2±1.94</b>	94.8±1.44
3	42	788	46.8±2.42	62.0±5.17	91.3±3.87	94.4±2.77	66.3±3.45	62.8±3.66	94.4±2.47	<b>95.0±2.30</b>	73.2±2.99	92.5±2.57	94.2±1.88
4	12	225	31.4±7.04	42.9±5.91	87.8±7.97	89.6±6.21	54.3±8.94	48.0±11.1	90.4±5.00	<b>90.5±5.62</b>	59.4±8.87	86.0±5.73	89.2±5.75
5	24	459	78.9±4.07	87.0±3.10	92.8±5.82	94.0±5.78	88.5±3.11	90.3±2.35	94.5±4.13	95.5±4.11	88.4±3.26	95.3±3.71	<b>96.6±2.32</b>
6	37	693	88.9±3.55	93.1±2.49	94.6±1.88	97.4±1.19	96.3±1.06	96.8±2.31	95.9±2.02	96.6±1.76	95.8±2.57	99.0±0.55	<b>99.3±0.81</b>
7	3	25	86.0±9.28	82.4±8.47	94.0±9.09	98.0±2.83	94.0±4.32	95.6±4.40	94.8±7.79	97.6±5.06	95.6±3.98	<b>99.6±1.26</b>	<b>99.6±1.26</b>
8	24	454	93.7±2.08	95.7±1.77	98.4±2.46	98.5±2.48	98.9±1.31	99.7±0.39	99.9±0.24	99.9±0.24	99.6±0.44	99.8±0.69	<b>99.9±0.11</b>
9	3	17	55.3±15.2	64.1±17.6	85.3±20.0	85.9±19.4	90.0±14.1	89.4±14.1	95.3±7.23	95.9±6.23	88.8±8.96	<b>99.4±1.86</b>	98.2±3.97
10	49	923	58.0±3.76	63.4±6.33	92.0±2.88	93.3±2.37	69.8±3.53	53.1±6.09	91.9±3.33	91.3±3.07	77.2±4.07	94.1±2.03	<b>95.6±2.07</b>
11	123	2332	68.7±3.16	76.7±2.31	94.9±1.57	95.0±1.64	80.1±2.80	82.0±3.81	94.8±2.34	95.3±2.14	84.3±2.46	95.4±1.21	<b>96.6±1.40</b>
12	30	563	39.6±4.10	55.9±7.24	80.2±7.34	82.3±8.20	73.7±3.94	71.6±5.70	88.8±4.84	90.7±4.97	78.2±4.99	<b>93.6±3.64</b>	92.8±3.30
13	10	195	86.1±10.1	93.3±4.38	95.7±5.26	98.2±2.78	98.1±0.91	98.3±0.94	97.4±5.32	98.0±4.59	97.8±1.83	99.1±0.96	<b>99.5±0.51</b>
14	63	1202	89.3±1.97	94.0±1.88	97.9±1.47	<b>99.1±0.87</b>	95.1±1.55	97.1±0.96	98.8±0.66	99.0±0.62	95.9±1.29	98.2±1.55	99.0±0.98
15	19	367	29.4±3.25	42.5±5.85	89.9±7.41	91.5±7.31	55.9±5.70	58.9±4.90	93.2±6.89	<b>93.5±7.03</b>	57.8±5.77	90.6±4.97	90.0±7.21
16	5	88	87.4±3.10	88.3±4.39	87.6±10.6	93.6±8.57	86.6±3.29	87.1±3.13	86.1±7.14	91.3±8.50	88.1±4.49	<b>97.0±2.40</b>	94.7±5.08
Overall accuracy			65.2±0.85	74.9±1.10	92.6±1.05	94.2±0.99	80.2±0.65	79.4±0.83	94.4±0.54	94.8±0.75	83.5±0.82	95.3±0.45	<b>96.1±0.42</b>
Average accuracy			64.0±1.04	72.5±1.86	90.9±2.52	92.9±2.50	80.6±1.28	79.5±2.12	93.9±1.04	94.8±0.99	83.2±1.10	94.1±1.31	<b>95.8±0.92</b>
Coefficient $\kappa$			60.2±0.91	71.3±1.31	91.6±1.20	93.4±1.13	77.3±0.72	76.3±1.01	93.6±0.62	94.1±0.85	81.2±0.94	94.6±0.52	<b>95.5±0.47</b>

### 4.3 Investigation on the balanced training set

The Indian Pines data have an unbalanced data distribution, where the total number of samples ranges from 20 to 2455 in each class. In order to investigate the performance of our proposed approach in this challenging

case, we present the classification results in the case of the balanced training set. We randomly choose 5% of the labeled samples per class for training (for the class with extremely limited training samples, a minimum of three samples is chosen as training samples, resulting 518 training samples totally). The remaining labeled samples are used for testing. In this case, the training samples per class are different, but the ratio of the labeled samples to total samples in each class is balanced.

The class accuracies, overall accuracies, average accuracies, and  $\kappa$  coefficients on the test set are recorded in Table 2, where the number of training and testing samples for each class is included. The metric learning methods provide higher class accuracies in most of the classes. The overall accuracies, average accuracies, and  $\kappa$  coefficients of the proposed SSML are consistently better than those of other methods. The classification maps of different methods are shown in Fig. 4, where SSML provides slightly better results with little noise in the map.

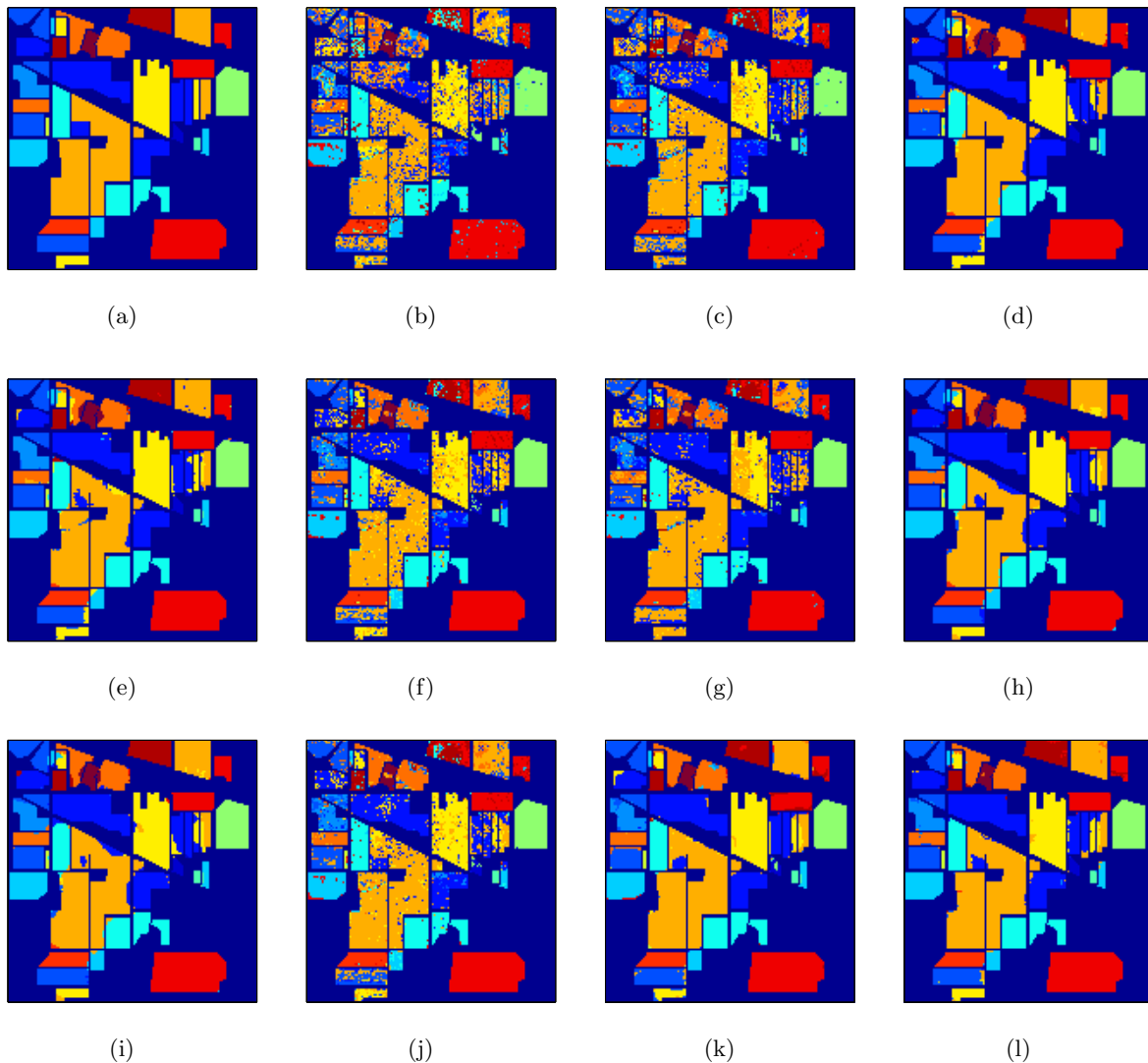


Figure 4. Classification maps. (a) Ground-truth. (b) NN (65.2%). (c) SpeSVM (74.9%). (d) SpaSVM (92.6%). (e) SSSVM (94.2%). (f) NN-DR (80.2%). (g) SpeSVM-DR (79.4%). (h) SpaSVM-DR (94.4%). (i) SSSVM-DR (94.8%). (j) SpeML (83.5%). (k) SpaML (95.3%). (l) SSML (96.1%).

## 5. CONCLUSION

In this paper, we have proposed a spatial and spectral information based metric learning framework for hyperspectral image classification. In the proposed method, a local pixel neighborhood preserving embedding has been first proposed to reduce the HSI dimensionality and meanwhile preserves the intrinsic data structure. Then, a composite spatial-spectral distance metric has been learned based on both the spectral and spatial information in the dimension reduced data. Experimental results show the effectiveness of our proposed SSML.

## Acknowledgments

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDC-T/017/2012/A1 and by the Research Committee at University of Macau under Grants MYRG2014-00003-FST, MRG017/ZYC/2014/FST, MYRG113(Y1-L3)-FST12-ZYC and MRG001/ZYC/2013/FST.

## REFERENCES

- [1] Fauvel, M., Tarabalka, Y., Benediktsson, J. A., Chanussot, J., and Tilton, J. C., "Advances in Spectral-Spatial Classification of Hyperspectral Images," *Proceedings of the IEEE* **101**(3), 652–675 (2013).
- [2] Camps-Valls, G., Tuia, D., Bruzzone, L., and Benediktsson, J. A., "Advances in hyperspectral image classification: Earth Monitoring with Statistical Learning Methods," *IEEE signal processing magazine* **31**(1), 45–54 (2014).
- [3] Xing, E., Ng, A., Jordan, M., and Russell, S., "Distance metric learning with application to clustering with side-information," In *NIPS 2003*.
- [4] Weinberger, K., Blitzer, J., and Saul, L., "Distance metric learning for large margin nearest neighbor classification," In *NIPS 2006*.
- [5] Yang, L., Jin, R., "Distance metric learning: a comprehensive survey", Technical Report, Michigan State University, 2006.
- [6] Zhu, P., Zhang, L., Zuo, W., and Zhang, D., "From Point to Set: Extend the Learning of Distance Metrics," In *ICCV 2013*.
- [7] Camps-Valls, G., Gomez-Chova, L., Muñoz-Maré, J., Vila-Francés, J., Calpe-Maravilla, J., "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letter* **3**(1), 93–97 (2006).
- [8] Chang, C., and Lin, C., "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27 (2011). Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.