Applied Mathematics & Information Sciences
*An International Journal*

http://dx.doi.org/10.12785/amis/091L10

# A Semi-Supervised Feature Extraction based on Supervised and Fuzzy-based Linear Discriminant Analysis for Hyperspectral Image Classification

*Cheng-Hsuan Li*[2,*], *Hsin-Hua Ho*[1], *Bor-Chen Kuo*[2], *Jin-Shiuh Taur*[1], *Hui-Shan Chu*[3] *and Min-Shian Wang*[2]

[1] Department of Electrical Engineering, National Chung Hsing University, 402 Taichung, Taiwan
[2] Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, 403 Taichung, Taiwan
[3] Department of Computer Science and Information Engineering, National Cheng Kung University, 701 Tainan, Taiwan

Received: 11 Nov. 2013, Revised: 12 Mar. 2014, Accepted: 13 Mar. 2014
Published online: 1 Feb. 2015

**Abstract:** Linear discriminant analysis (LDA) is a commonly used feature extraction method to resolve the Hughes phenomenon for classification. Moreover, many studies show that the spatial information can greatly improve the classification performance. Hence, for hyperspectral image classification, it is not only necessary to use the available spectral information but also to exploit the spatial information. Recently, we proposed a fuzzy-based LDA (FLDA), an unsupervised feature extraction, and used it for clustering problem. However, it is hard to apply in the image segmentation because the optimization problem is nonlinear and non-convex and the number of membership values, the product of the number of clusters and the number of pixels in the image, is too large. In this paper, a semi-supervised feature extraction method which is based on the scatter matrices of LDA and FLDA (FLDA) is proposed. The unknown samples and their membership values which are determined by the posteriors after applying the classifier are used to form the within- and between-cluster scatter matrices of FLDA. The experimental results on two hyperspectral images, the Washington DC Mall and the Indian Pine Site, show that the proposed method can yield a better classification performance than LDA in the small sampling size problem.

**Keywords:** Linear Discriminant Analysis, LDA, Unsupervised LDA, Fuzzy-Based LDA, Semi-Supervised LDA

## 1 Introduction

Linear discriminate analysis (LDA) [1] is a commonly used feature extract (FE) method to resolve the Hughes phenomenon, a kind of problem of the curse of dimensionality, is often encountered in classification when the dimensionality of the space grows and the size of the training set is fixed, especially in the small sampling size problem. Recently, a fuzzy-based linear discriminant analysis (FLDA) based on the membership values is proposed [2,3]. The membership values of the scatter matrices in FLDA are obtained from the cluster result of fuzzy linear discriminant clustering (FLDC). However, it is hard to apply FLDA in the image segmentation because the optimization problem is nonlinear and non-convex and the number of membership values, the product of the number of clusters and the number of pixels in the image, is too large.

Moreover, the spatial information by considering semi-labels can greatly improve the accuracy of the hyperspectral image classification [4,5]. Using the semi-labels of the unknown samples can not only increase the training samples but also decrease the imprecise estimation of samples, which have very similar spectral properties, and have difficulty distinguishing the unlabeled patterns.

Therefore, in this study, we want to integrate the scatter matrices of LDA and FLDA, and develop semi-supervised scatter matrices. The unknown samples and their membership values which are determined by the posteriors after applying the classifier are used to form the within- and between-cluster scatter matrices of FLDA. After obtaining the within- and between-cluster scatter matrices based on the semi-information, combine with the scatter matrices of LDA to form semi-scatter matrices. Therefore, in this paper, a semi-supervised LDA (SLDA)

* Corresponding author e-mail: ChengHsuanLi@gmail.com

based on the semi-information obtained by the applying classifier is proposed.

The study is organized as follows. The LDA and its fuzzy-based version will be introduced in section 2. Section 3 describes our proposed method, SLDA, and the experimental results are shown in Section 4 Finally, some conclusions are addressed in Section 5.

## 2 Linear Discriminant Analysis and Its Fuzzy-Based Version

### 2.1 Linear Discriminant Analysis

LDA [1] is one of the most well-known feature extraction methods, which aims to estimate a linear transformation to map data from a high-dimensional space into a low-dimensional subspace that maintains an unchanged separability. LDA uses the mean vector and covariance matrix of each class, to formulate within-class, between-class, and mixture-class scatter matrices. Based on the Fisher criterion, the LDA method finds features such that the ratio of the between-class scatter to the average within-class scatter is maximized, in a lower dimensional space. By applying the concept of class scattering to class separation, the Fisher criterion, takes the large values from samples when they are well clustered around their mean within each class, and the clusters of the different classes are well separated. The formula is stated as the following:

Suppose $H_i = \{x_1^{(i)}, \ldots, x_{N_i}^{(i)}\} \subset R^d$ were the set of samples in class $i$, $N_i$ were the number of samples in class $i$, $i$ = 1,..., $L$, and $N = N_1 + \cdots + N_L$ were the number of all training samples. The between-class scatter matrix $S_b^{LDA}$ and the within-class scatter matrix $S_w^{LDA}$ would be defined as

$$S_b^{LDA} = \sum_{i=1}^{L} \frac{N_i}{N} (m_i - m)(m_i - m)^T$$

and

$$S_w^{LDA} = \sum_{i=1}^{L} \sum_{j=1}^{N_i} \frac{1}{N} \left(x_j^{(i)} - m_i\right) \left(x_j^{(i)} - m_i\right)^T$$

where $m_i$ were the class mean defined by

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j^{(i)} \text{ and } m = \frac{1}{N} \sum_{i=1}^{L} \sum_{j=1}^{N_i} x_j^{(i)}$$

representing the total mean.

The optimal features are determined by optimizing the Fisher criterion given by $J_{LDA} = tr[(S_w^{LDA})^{-1} S_b^{LDA}]$. This is equivalent to solving the generalized eigenvalue problem,

$$S_b^{LDA} v_z = \lambda_z S_w^{LDA} v_s \ , \ z = 1, \ldots, d$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ where the extracted eigenvectors are used to form the transformation matrix of LDA, i.e., the transformation matrix from the original space to the reduced subspace is defined by: $A^{LDA} = [v_1, v_2, \ldots, v_p]$. The Fisher criterion $J_{LDA}$ is able to detect the capacity of the separability for the transformed training samples.

### 2.2 Fuzzy-Based Linear Discriminant Analysis

In the subsection, a fuzzy-based LDA (FLDA) using the concept of membership values is described [2,3]. Let $X$ be a hyperspectral $d$-dimensional image of size $I \times J$ pixels, and $\{x_1, \ldots, x_{I \times J}\} \subset R^d$ be all samples in the image. The between scatter matrix $S_b^{FLDA}$ and the within scatter matrix $S_w^{FLDA}$ as follows:

$$S_b^{FLDA} = \sum_{i=1}^{L} \frac{\sum_{j=1}^{I \times J} u_{ij}}{I \times J} (m_i - m)(m_i - m)^T$$

and

$$S_w^{FLDA} = \sum_{i=1}^{L} \sum_{j=1}^{I \times J} \frac{u_{ij}}{I \times J} (x_j - m_i)(x_j - m_i)^T$$

where $u_{ij}$ means the membership value of the sample $x_j$ in class $i$,

$$m_i = \sum_{j=1}^{I \times J} \frac{u_{ij}}{\sum_{k=1}^{N} u_{ik}} x_j$$

is the class mean, and $m = \frac{1}{I \times J} \sum_{j=1}^{I \times J} x_j$ represents the total mean.

The weight $\sum_{j=1}^{I \times J} u_{ij}/(I \times J)$ indicates the influence of the samples in class $i$. If clustering results show that the number of samples in cluster $i$ is larger than in cluster $k$, then

$$\sum_{j=1}^{I \times J} u_{ij}/(I \times J) > \sum_{j=1}^{I \times J} u_{kj}/(I \times J),$$

i.e., the influence of $m_i$ is larger than $m_k$ in measuring the between separability. The weight $u_{ij}/(I \times J)$ indicates the influence of the $j$-th sample in class $i$. If $u_{ij}/(I \times J) > u_{kj}/(I \times J)$, then the $j$-th sample favors in class $i$. Hence, the corresponding weights in the within measure is larger.

Furthermore, it is easy to observe that the scatter matrices of LDA are special cases of the scatter matrices of FLDA, i.e., $u_{ij} = 1$ and $u_{kj} = 0$, $k \neq i$ when $x_j$ is the training sample in class $i$. In [2], the clustering performance after applying UFLDA outperforms the clustering performance after applying principal component analysis (PCA) and independent component analysis (ICA).

However, it is hard to apply FLDA in the image segmentation because the optimization problem is nonlinear and non-convex and the number of membership values, the product of the number of clusters and the number of pixels in the image, is too large. For overcoming this problem, the posteriors of a sample after applying the classifier are used to be the membership values of this sample.

## 3 A Semi-Supervised Linear Discriminant Analysis

In this section, we propose a semi-information based LDA (SLDA) algorithm. The purpose of incorporating the semi-information in feature extraction is to increase the useable samples. In this study, the posteriori probabilities by applying the Gaussian classifier and the posteriori probabilities defined by $k$NN classifier are used to form the membership values of the scatter matrices of FLDA. Then the semi-information then can be both used to adjust the extracted direction, and the classification performance will be improved.

### *3.1 Scatter Matrices of SLDA*

The between scatter matrix $S_b^{SLDA}$ and the within scatter matrix $S_w^{SLDA}$ are defined as follows:

$$S_b^{SLDA} = \alpha S_b^{LDA} + \beta S_b^{FLDA},$$

and

$$S_w^{SLDA} = \alpha S_w^{LDA} + \beta S_w^{FLDA}$$

where $\alpha$ and $\beta$ are two parameters to control the influences of the scatter matrices of LDA and semi-scatter matrices of FLDA. In this paper, the $k$-fold cross validation method is used to determine the proper parameters, $\alpha$ and $\beta$.

The features of SLDA can be obtained by solving the following generalized eigenvalue problem:

$$S_b^{SLDA} v_z = \lambda_z S_w^{SLDA} v.$$

The extracted eigenvectors form the transformation matrix of SLDA, i.e.,

$$A^{SLDA} = [v_1, v_2, ..., v_p].$$

One can find that if $\beta = 0$, then $A^{SLDA} = A^{LDA}$.

### *3.2 Posteriori Probabilities by the Classifier*

*(1) Gaussian Classifier:* After applying the Gaussian classifier [1], the class-conditional probability density function of the unknown sample $x_j$ is

$$p(x_j|i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \cdot \exp\left(-\frac{1}{2}(x_j - m_i)^T \Sigma_i^{-1}(x_j - m_i)\right)$$

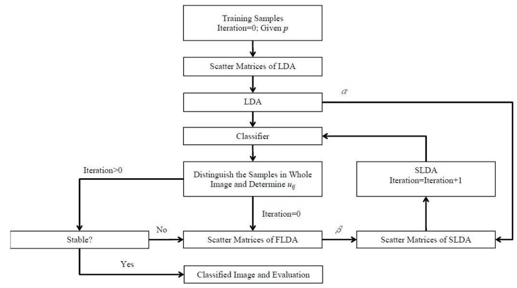



**Fig. 1:** The framework of the proposed method, SLDA.

where $m_i = \frac{1}{N_i}\sum_{\ell=1}^{N_i} x_\ell^{(i)}$ is the mean of class $i$,

$$\Sigma_i = \frac{1}{N_i}\sum_{\ell=1}^{N_i}(x_\ell^{(i)} - m_i)(x_\ell^{(i)} - m_i)^T$$

is the $d \times d$ covariance matrix, and $|\Sigma_i|$ denotes the determinant of $\Sigma_i$. Then $x_j$ is distinguished to class $i$, if

$$p(x_j|i)p(i) \geq p(x_j|k)p(k)$$

for all $k \neq i$, where $p(i)$ is the priori probability of class $i$. In SLDA, the membership value $u_{ij}$ is defined by

$$u_{ij} = \frac{p(x_j|i)p(i)}{\sum_{k=1}^{L} p(x_j|k)p(k)}.$$

*(2) k-Nearest-Neighbor (kNN) Classifier:* Let $X_j = [x_{j1}, \cdots, x_{jk}]$ be the $k$ nearest neighbors of $x_j$. The $k$NN classifier [1] finds the $k$ nearest neighbors of semi-points and assigns the semi-information to the most frequently occurring class of its $k$ neighbors.

In SLDA, the membership values of the semi-scatter matrices are based on the distance measure. According to the Euclidean distance rule, the membership value $u_{ij}$ of the unknown sample $x_j$, is defined as

$$u_{ij} = \frac{\sum_{r=1}^{a_i} \left\|x_{jr} - x_j\right\|^{-1}}{\sum_{i=1}^{L}\sum_{r=1}^{a_i} \left\|x_{jr} - x_j\right\|^{-1}}.$$

where $a_i$ is the number of samples in $x_j$ which belong class $i$, $i = 1, \ldots, L$, and $j = 1, \ldots, I \times J$.

### *3.3 The Framework of SLDA*

Fig. 1 is the framework of the proposed method, SLDA, which is based on semi-information. Given the training samples and determine the reduced dimensionality $p$. In the first part, the scatter matrices of LDA are computed and LDA is used for feature extraction. After applying the classifier, the semi-labels and membership values can be obtained, and the scatter matrices of FLDA in the second part can be computed. By multiplying the parameters $\alpha$

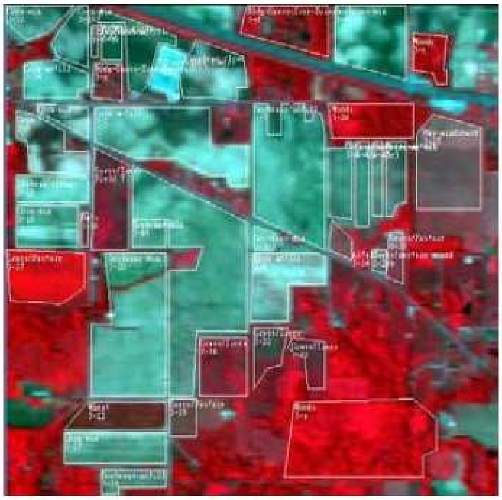


**Fig. 2:** A portion of the Indian pine site image with a size of 145x145 pixels.

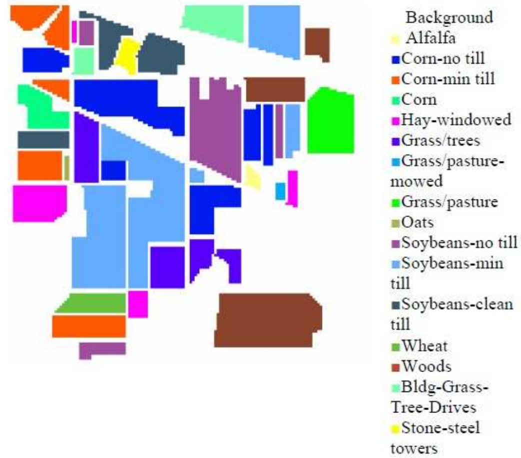



**Fig. 3:** The ground truth of the Indian pine site dataset

| No. | Category |
|---|---|
| 1 | grass |
| 2 | tree |
| 3 | roof |
| 4 | water |
| 5 | road |
| 6 | trail |
| 7 | shadow |

**Fig. 4:** The false-color IR image of a portion of Washington, D.C. Mall image with a size of 205x307 pixels. There are seven categories, grass, tree, roof, water, road, trail, and shadow.

and $\beta$, the scatter matrices of SLDA can be computed. After applying the SLDA and the classifier, the new semi-labels and membership values can be obtained. If the change rate of semi-labels is less than a threshold, then stop the iteration. Otherwise, continue the second part until the change rate of semi-labels is less than a threshold.

## 4 Experimental Results

### *4.1 Experimental Data and Design*

In this study, two real data sets are applied to evaluate the classification performances. They are the Indian Pine site (IPS): a mixed forest/agricultural site in Indiana [6] and the Washington, D.C. Mall hyperspectral image [6] as an urban site.

The IPS dataset was gathered by a sensor known as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). This dataset was obtained from an aircraft operated by the NASA/Jet Propulsion Laboratory and flown at an altitude of 65000ft. The images measured 145145 pixels, with 220 spectral bands measuring approximately 20m across the ground. Figs. 2 and 3 show the grayscale IR image and ground truth of IPS, respectively. There are 16 different land-cover classes available in the original ground-truth image [6]. This study uses 9 categories, Corn-no till (class 1), Corn-min till (class 2), Grass/pasture (class 3), Grass/trees (class 4), Hay-windowed (class 5), Soybeans-no till (class 6), Soybeans-min till (class 7), Soybeans-clean till (class 8), Woods (class 9), because sample sizes of the rest classes are too small [7]-[9].

The second data set, the Washington, D.C. Mall, was obtained from a Hyperspectral Digital Imagery Collection Experiment (HYDICE) from an airborne hyperspectral data flightline over the Washington, DC urban area. Two hundred and ten bands were collected in the 0.4-2.4 $\mu$m region of the visible and infrared spectrum. Some water absorption channels were discarded, resulting in 191 channels. This dataset is available in the student CD-ROM of [6]. This experiment used 7 classes: grass (class 1), tree (class 2), roof (class 3), water (class 4), road (class 5), trail (class 6), and shadow (class 7). Fig. 4 shows the grayscale IR image of a portion of the image and the corresponding seven categories.

This study uses three distinct subsets, $N_i$=20 $< N < d$ (case 1), $N_i$=40 $< d < N$ (case 2), and $d < N_i$=300 $< N$ (case 3), to investigate the influence of training sample size to the dimensionality in these two hyperspectral image data sets. The case 1 is an ill-posed classification situation, which means data dimensionality exceeds the number of independent training samples in every class. The case 2 is a poorly-posed classification situation, which means that data dimensionality is greater than or comparable to the number of (independent) per-class representative training samples, but smaller than the total

**Table 1:** Table I: The Accuracy in Percentages for Two Different Data Sets. The Number in Bracket is the Reduced Dimensionality

| Case | Classifier | FE | Dataset | |
|---|---|---|---|---|
| | | | IPS | DC |
| 1 | Gaussian | LDA | 55.89% (7) | 47.71% (3) |
| | | SLDA | 69.56% (8) | 90.86% (4) |
| | kNN | LDA | 60.22% (8) | 64.43% (6) |
| | | SLDA | 68.78% (8) | 91.00% (6) |
| 2 | Gaussian | LDA | 69.00% (8) | 86.29% (6) |
| | | SLDA | 78.89% (8) | 92.00% (4) |
| | kNN | LDA | 69.00% (8) | 89.29% (6) |
| | | SLDA | 78.67% (8) | 90.43% (6) |
| 3 | Gaussian | LDA | 87.78% (8) | 96.14% (5) |
| | | SLDA | 87.56% (8) | 94.71% (6) |
| | kNN | LDA | 86.00% (8) | 95.86% (6) |
| | | SLDA | 85.56% (8) | 95.29% (6) |


**Fig. 5:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the Gaussian classifier in case 1 ($N_i$ =20), respectively.

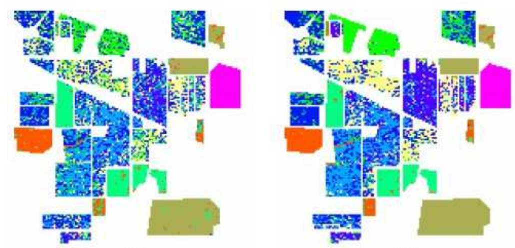



**Fig. 6:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the Gaussian classifier in case 2 ($N_i$ =40), respectively.

number of representative samples. In case 3, there are enough independent training samples. MultiSpec [6] was used to select training and testing samples (100 testing samples per class) randomly in all experiments [**?**].

In the experiments, the classification performances by applying LDA and SLDA with the Gaussian and *k*NN classifiers are compared. The grid search with *k*-fold cross validation is used to find the proper parameters $\alpha$ and $\beta$ within a set $\{0, 0.1, 0.2, \ldots, 2\}$.

## 4.2 Results

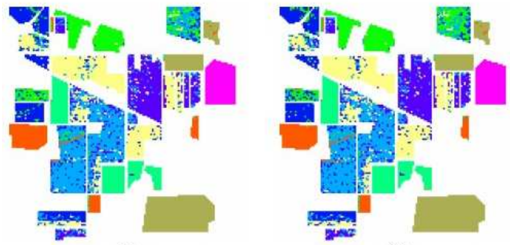
In the experimental results, the classification accuracies in percentages for two different hyperspectral image data sets, the Indian Pine site (IPS) and the Washington, D.C. Mall (DC) hyperspectral image, are showed in Table I. Note that LDA only can extract *L*–1 (number of classes minus one) features. For example, the Indian Pine site (IPS) dataset has 9 categories, so LDA only can extract 8 features. Hence, the classification performances by applying LDA and SLDA with 1 to *L*-1 extracted features are computed. The highest classification accuracies among 1 to *L*-1 extracted features by applying LDA and SLDA are showed in Table I. The numbers in brackets are the corresponding reduced dimensionalities.

As result of experiment we can know that the number of dimensionality with the classification accuracies has weak correlation. The classification performances by applying our proposed SLDA are higher than the classification performances by applying LDA in cases 1 and 2. However, in case 3, the accuracies by applying our proposed method are similar to that by applying LDA. Hence, in the small sample size problem, the proposed method can yield a better classification performance than LDA. Especially in the case 1, an ill-posed classification situation, the proposed method can improve the performance of LDA significantly. This is because the following two reasons. First, the number of training samples of SLDA is larger than of LDA, since the semi-samples is used by applying SLDA. Second, the within- and between-class scatter matrices are modified by the scatter matrices of FLDA with the estimations of posteriors.

Figs. 5-7 show the classification maps of the IPS dataset by applying LDA and SLDA with the Gaussian classifier in three cases, respectively. Similarly, Figs. 8-10 show the classification maps of the IPS dataset by applying LDA and SLDA with the *k*NN classifier in three cases, respectively. Since the number of training samples in each class and the number of all training samples are both less than the dimensionality in case 1, the number of independent samples is insufficient. Hence, the estimations of covariance matrices may be biased. That is the Fig. 1 (a) is the poorest. Moreover, in case 1 and case 2, the proposed method SLDA can not only increase the classification performance but also decrease the imprecise estimation of samples, especially, “Grass/pasture,” “Soybeans-clean till,” “Woods,” “Grass/trees,” and “Hay-windowed” for IPS dataset no matter which classifier is used.

## 4.3 Conclusions

In this paper, a semi-supervised LDA (SLDA) is proposed. The scatter matrices of SLDA are combined

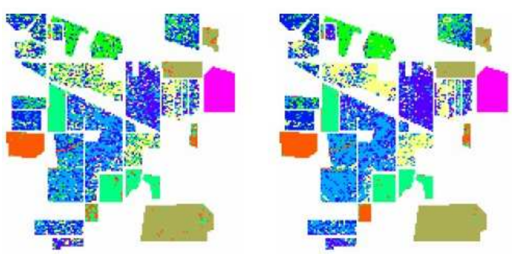


(a) (b)

**Fig. 7:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the Gaussian classifier in case 3 ($N_i$ =300), respectively.

(a) (b)

**Fig. 8:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the *k*NN classifier in case 1 ($N_i$ =20), respectively.



(a) (b)

**Fig. 9:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the *k*NN classifier in case 1 ($N_i$ =40), respectively.

(a) (b)

**Fig. 10:** (a) and (b) are the classification maps of the IPS dataset by applying LDA and SLDA with the *k*NN classifier in case 1 ($N_i$ =300), respectively.

from the scatter matrices of LDA and semi-scatter matrices of FLDA. The posteriors after applying the Gaussian classifier and defined by *k*NN classifier were used to determine the membership values of the scatter matrices of FLDA. The experimental results obtained from two different hyperspectral image datasets, the Indian Pine site (a mixed forest/agricultural site in Indiana) and the Washington, D.C. Mall hyperspectral image (an urban site in Washington, D.C.) confirm that the proposed SLDA improves the classification accuracies when sampling size is small.

In the future, we will apply optimal algorithms to find proper parameters $\alpha$ and $\beta$ for SLDA. Moreover, for applying SLDA to the support vector machines (SVMs), we will design a mathematical model to define the posteriors after applying SVMs. Another consideration is to change the neighborhood system. In this study the neighborhood system is for the hyperspectral image classification problem. For non-image classification problem, the neighborhood system can be design by the concept of *k*NN. We will investigate these researches in the feature.

## Acknowledgement

The authors would like to thank two anonymous reviewers for their suggestions and Dr. Landgrebe for providing the Indian Pines site and The Washington DC Mall data sets. This work was supported in part by the National Science Council, Taiwan, under Grant NSC 100-2628-E-142-001-MY3.

## References

[1] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., San Diego, CA: Academic, 1990.

[2] C.-H. Li, B.-C. Kuo, and C.-T. Lin, “LDA-based Clustering Algorithm and Its Application to an Unsupervised Feature Extraction,” *IEEE Transactions on Fuzzy Systems*, **19**, 152-163, Feb. 2011.

[3] H.-S. Chu, C.-H. Li, B.-C. Kuo, and C.-T. Lin, “A Semisupervised Feature Extraction Method Based on Fuzzy-type Linear Discriminant Analysis,” *IEEE International Conference on Fuzzy Systems*, 1927-1932, June 27-30, 2011, Taipei, Taiwan.

[4] Q. Jackson and D. A. Landgrebe, “Adaptive Bayesian Contextual Classification Based on Markov Random Fields,” *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 2454-2463, 2002.

[5] C.-H. Li, B.-C. Kuo, C.-T. Lin, and C.-S. Huang, “A Spatial-Contextual Support Vector Machine for Remotely Sensed Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, 1-16, 2011.

[6] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, Hoboken, NJ: Chichester, 2003.

[7] B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel Nonparametric Weighted Feature Extraction for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 1139-1155, April 2009.

[8] J.-M. Yang, P.-T. Yu, and B.-C. Kuo, "A Nonparametric Feature Extraction and Its Application to Nearest Neighbor Classification for Hyperspectral Image Data," *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 1279-1293, March 2010.

[9] J.-M. Yang, B.-C. Kuo, P.-T. Yu, and C.H. Chuang, "A Dynamic Subspace Method for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2840-2853, July 2010.

---

**Cheng-Hsuan Li** received the B.S. and M.S. degrees from National Chung Hsing University, Taichung, Taiwan, in 2001 and 2003, respectively, and the Ph.D. degree in Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu, Taiwan. He is currently an assistant Professor with the Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taichung, Taiwan. His research interests include pattern recognition, data mining, and machine learning.

**Hsin-Hua Ho** received the B.S. degree from National Chung Hsing University, Taichung, Taiwan, in 2002, and the M.S. degree from National Taichung University of Education, Taichung, Taiwan in 2006. She is currently working toward the Ph.D. degree in Department of Electrical Engineer, National Chung Hsing University, Taichung, Taiwan. Her research interests include pattern recognition and machine learning.

**Bor-Chen Kuo** (S'01-M'02) received the B.S. degree in mathematics education and the M.S. degree in statistics and assessment from National Taichung Teachers College, Taichung, Taiwan, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 2001. From 2002 to 2005, he was an Associate Professor with National Taichung Univer-sity of Education, Taichung, where he is currently a Professor and the Director of the Graduate Institute of Educational Measurement and Statistics. His research interests include pattern recognition, feature extraction and classification of high-dimensional data, computerized adaptive testing, and nonparametric functional estimation. Dr. Kuo is a member of the Chinese Association of Psychological Testing AdCom and the Editor of the Journal of Educational Measurement and Statistics, Taiwan. He was the recipient of an Outstanding and Excellence Research Award from the R.O.C. Education and Research Society in 2009.

**Jin-Shiuh Taur** (SM'05) received the B.S. and M.S. degrees in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1987 and 1989, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1993. He was a member of the Technical Staff with Siemens Corporate Research, Inc. He is currently a Professor with the National Chung Hsing University, Taichung, Taiwan. His research interests include neural networks, pattern recognition, computer vision, and fuzzy-logic systems. Dr. Taur was the recipient of the 1996 IEEE Signal Processing Society's Best Paper Award.

**Hui-Shan Chu** received the M.S. degree from National Taichung University of Education, Taichung, Taiwan in 2011. She is currently working toward the Ph.D. degree in Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan. Her research interests include pattern recognition and machine learning.

**Min-Shian Wang** received the B.S. degree from the Department of Statistics, Feng Chia University, Taichung, Taiwan, in 2006, and the M.S. degree in the Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taichung, Taiwan, in 2011. She is currently an executive assistant with the Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education. Her research interests include educational measurement, pattern recognition, and machine learning.