*Journal of **AI** and **D**ata **M**ining*
Vol 3, No 2, 2015, 181-190.
10.5829/idosi.JAIDM.2015.03.02.07

Overlap-based feature weighting: The feature extraction of Hyperspectral remote sensing imagery

M. Imani and H. Ghassemian*

Faculty of Electrical & Computer Engineering, Tarbiat Modares University, Tehran, Iran.

Received 12 April 2015; Accepted 26 August 2015
**Corresponding author: ghassemi@modares.ac.ir (H. Ghassemian).*

Abstract
Hyperspectral sensors provide a large number of spectral bands. This massive and complex data structure of hyperspectral images presents a challenge to traditional data processing techniques. Therefore, reducing the dimensionality of hyperspectral images without losing important information is a very important issue for the remote sensing community. We propose to use overlap-based feature weighting (OFW) for supervised feature extraction of hyperspectral data. In the OFW method, the feature vector of each pixel of hyperspectral image is divided to some segments. The weighted mean of adjacent spectral bands in each segment is calculated as an extracted feature. The less the overlap between classes is, the more the class discrimination ability will be. Therefore, the inverse of overlap between classes in each band (feature) is considered as a weight for that band. The superiority of OFW, in terms of classification accuracy and computation time, over other supervised feature extraction methods is established on three real hyperspectral images in the small sample size situation.

Keywords: *Class Discrimination, Overlap, Feature Weighting, Feature Extraction, Hyperspectral.*

1. Introduction

The high spectral resolution hyperspectral images allow the characterization, identification, and classification of the land covers with improved accuracy, robustness, and more details. A large number of training samples is required for achieving satisfactory accuracy in classification problems. However, the collection of ground reference data (training samples) in real world applications is an expensive and time consuming task and so the number of available training samples might be very limited.

There are different solutions to cope with the small training sample size. Semi-supervised approaches use the ability of unlabeled samples in addition to labeled samples to improve the classification accuracy [1,2]. Advanced classifiers such as kernel-based classifiers are distribution free and do not make assumptions about the density functions of the data [3,4]. Feature reduction is one of the most important solutions for small sample size problem [5-9]. In addition to improving the classification accuracy, feature reduction techniques reduce the computational complexity and also simple the visualization of data. Feature reduction methods are divided into two general groups: feature selection and feature extraction. Feature selection methods select an appropriate subset of features from the original candidate features and maintain the physical meaning of data. Feature extraction methods transform the feature space of data usually with using a projection matrix. Feature reduction techniques can be done supervised [10,11], unsupervised [12,13] or semi-supervised [14]. We assess the supervised feature extraction methods in this paper.

Multiple features such as spectral, texture, and shape features are employed to represent pixels from different perspectives in hyperspectral image classification. The properly combining multiple features results in good classification performance. A patch alignment framework to linearly combine multiple features in the optimal way, which obtains a unified low-dimensional representation of these multiple features for subsequent classification, is introduced in [15]. A pixel in a hyperspectral image can be represented

by both spatial and spectral features. Each view of a feature summarizes a specific characteristic of the studied object from different feature spaces, and also features for different views are complementary to each other. An ensemble manifold regularized sparse low-rank approximation algorithm for multi-view feature dimensionality reduction is proposed in [16].

Linear discriminant analysis (LDA) is a simple and popular method for feature extraction in different pattern recognition applications [17]. LDA maximizes the between-class scatter matrix and minimizes the within-class scatter matrix to increase the class discrimination. Because of singularity of within-class scatter matrix, LDA has weak efficiency when the number of training samples is limited. Generalized discriminant analysis (GDA) is the nonlinear version of LDA, which works in the kernel space [18]. Because of the limitation of rank of between-class scatter matrix, LDA and GDA can extract maximum $c-1$ features where c is the number of classes. Nonparametric weighted feature extraction (NWFE) uses the nonparametric form and weighted mean for calculation of scatter matrices [19]. Thus, NWFE can extract more than $c-1$ features and, moreover, it has good efficiency with small training set. Median-mean line discriminant analysis (MMLDA), which is recently proposed, copes with the negative effect of the class mean caused by outliers with introduction of median–mean line as an adaptive class-prototype [20].

We propose a supervised feature extraction method in this paper that is simple, fast and efficient in small sample size situation. The proposed method is named overlap-based feature weighting (OFW). In a hyperspectral image, the adjacent spectral bands contain redundant information. Thus, we divide the feature vector of each sample of data to some segments in such a way that each segment contains adjacent spectral bands. We consider the weighted mean of spectral bands (original features) in each segment as an extracted feature. If classes have more overlap in a spectral band, then, the discrimination of classes in that band is harder. Thus, the class discrimination ability in each band has reverse relationship with the overlap value between classes in that band.

Therefore, we assign the inverse of overlap between classes in each feature, as a weight for that feature in the weighted mean. Feature extraction methods such as LDA, GDA, NWFE, and MMLDA need to estimate the mean vectors (the first order statistics) and the scatter matrices (the second order statistics). The accurate estimate of statistics needs large enough training set. When the number of training samples is limited, the accurate estimate of mean vectors and covariance matrices cannot be provided, and so, the accuracy of LDA-based methods such as conventional LDA, GDA, NWFE, and MMLDA is decreased. The proposed method, OFW, just uses the original training samples and does not need to estimate the statistics of data. Therefore, it can have good efficiency in small sample size situations compared to LDA-based methods. Moreover, OFW has simple calculations, so, it is fast. The efficiency of OFW is investigated by three real hyperspectral images. The current paper focuses on the following sections: section 2 introduction of proposed method, section 3 the experimental results, and section 4 conclusions.

2. Proposed method

The adjacent spectral bands (features) in each pixel of hyperspectral image contain high redundant information. Then, for extraction of m features from d original spectral bands, we divide the feature vector of each sample of data to m segments containing $K=\left\lfloor\frac{d}{m}\right\rfloor$ adjacent spectral bands. Then, the weighted mean of spectral bands in each segment, is considered as an extracted feature for that segment. Let, $\boldsymbol{x}=[x_1 \quad x_2 \cdots \quad x_d]^T$ be the feature vector of a pixel of hyperspectral image and $\boldsymbol{y}=[y_1 \quad y_2 \cdots \quad y_m]^T$ be the extracted feature vector of $\boldsymbol{x}$ where $m<d$. The elements of $\boldsymbol{y}$ are calculated as follows:

$$y_l=\sum_{j=(l-1)K+1}^{lK} w_j x_j \ , \qquad 1\le l\le m-1 \tag{1}$$

$$y_m=\begin{cases}\displaystyle\sum_{j=(m-1)K+1}^{d} w_j x_j \ ; \ mK<d \ (m<\frac{d}{K}) \\ \displaystyle\sum_{j=(m-1)K+1}^{mK} w_j x_j \ ; \ mK=d \ (m=\frac{d}{K})\end{cases} \tag{2}$$

where, w_j is the weight of jth spectral band in the above weighted mean. How to decompose the whole spectral signature has been searched in some literatures such as [21].

To this end, we implemented the simplest possible approach for segmentation of spectral signature of pixels. The calculation of weights is the novelty of our proposed method. In some spectral bands, the difference between classes is more than other bands.

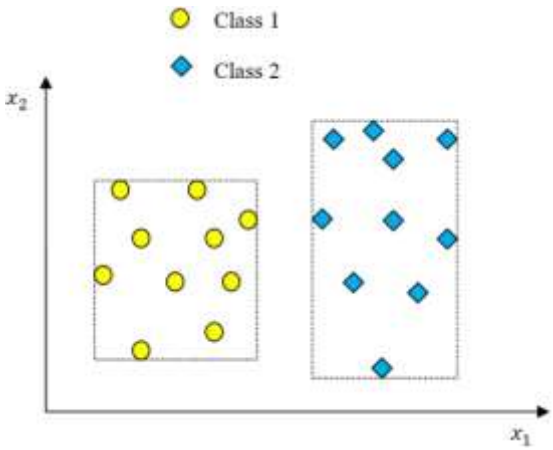

Figure 1. Samples of two classes in a two-dimensional feature space.

The more the overlap between classes is in a spectral band (feature), the harder the class discrimination will be in that spectral band. In other words, the class discrimination ability, in each feature, has reverse relationship with the overlap between classes in that feature.

Figure 1 shows the samples of two classes in a two-dimensional feature space. In band x_1, two classes have not overlap and thus are discriminable from each other, while in the band x_2, classes are overlapped and discrimination between them is hard. To better understanding, see figure 2.

Two classes in band x_j, have no overlap; thus, they are easily separated from each other using a simple line; while these two classes have overlap in band x_l, and so, a complex nonlinear curve is needed to separate them from each other. Therefore, it is obvious that the ability of each spectral band in discrimination between classes has a reverse relationship with the overlap between classes in that band.

Let, $x_{jq}^{i}(j = 1, \dots, d; q = 1, \dots, n_i; i = 1, \dots, c)$ be the jth feature of qth sample of class i where d, c, and n_i are the number of spectral bands (features), the number of classes, and the number of training samples in class i, respectively. The minimum and maximum values of each spectral band in each class are given by:

$$f_{min,ji} = \min_{q=1,\dots,n_i} x_{jq}^{i} \; ; i = 1, \dots, c \; ; \; j = 1, \dots, d \tag{3}$$

$$f_{max,ji} = \max_{q=1,\dots,n_i} x_{jq}^{i} \; ; i = 1, \dots, c \; ; \; j = 1, \dots, d \tag{4}$$

where, $f_{min,ji}$ is the minimum value of feature j in class i and $f_{max,ji}$ is the maximum value of feature j in class i.

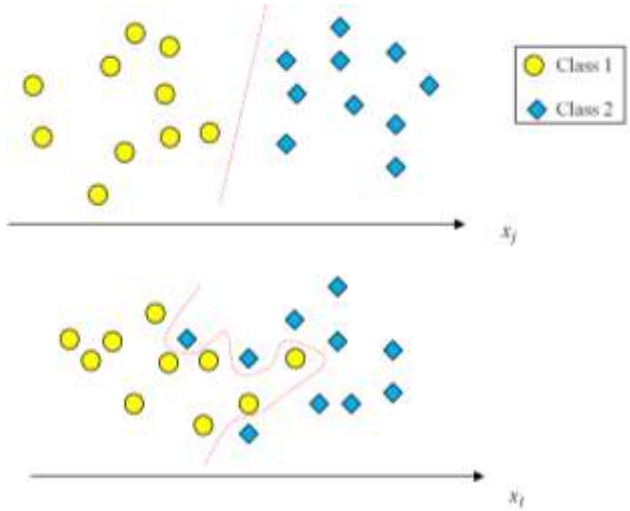

Figure 2. There is not overlap between classes in band x_j, and so two classes are easily separated from each other in x_j while there is overlap between classes in band x_l, and so two classes are hardly separated from each other in x_l.

Two classes i and k $(i, k = 1, \dots, c)$ are not overlapped and are completely separate from each other in band j if:

$$\begin{aligned}&(f_{min,ji} < f_{min,jk}) \,\&\, (f_{min,ji} < f_{max,jk}) \\ &\&\, (f_{max,ji} < f_{min,jk}) \,\&\, (f_{max,ji} < f_{max,jk})\end{aligned} \tag{5}$$

or

$$\begin{aligned}&(f_{min,ji} > f_{min,jk}) \,\&\, (f_{min,ji} > f_{max,jk}) \,\& \\ &(f_{max,ji} > f_{min,jk}) \,\&\, (f_{max,ji} > f_{max,jk})\end{aligned} \tag{6}$$

Otherwise, two classes i and k have overlap and the value of overlap between them in feature j is calculated as follows:

$$\begin{aligned}(OV_{ik})_j = \big|&\max(f_{min,ji}, f_{min,jk}) \\ &- \min(f_{max,ji}, f_{max,jk})\big| \; ; \\ i = 1, \dots, c; k = 1, \dots, c; \; j = 1, \dots, d\end{aligned} \tag{7}$$

where, $(OV_{ik})_j$ is the overlap value of class i and class k in feature j. The overlap between all pairs of classes is calculated as follows:

$$OV_j = \frac{1}{2} \sum_{k=1}^{c} \sum_{i=1}^{c} (OV_{ik})_j \; ; \; j = 1, \dots, d \tag{8}$$

The class discrimination ability has reverse relationship with the overlap value between classes. Thus, the weight associated with each feature in the weighted mean in (1) and (2) is calculated by:

$$w_j = \frac{1}{OV_j}; \; j = 1, \dots, d \tag{9}$$

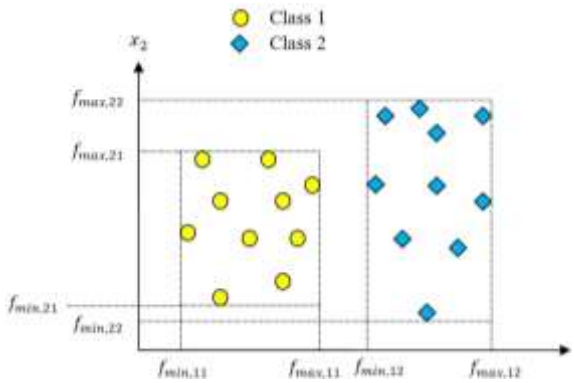

Figure 3. An example of determination of overlap between two classes.

Figure 3 shows an example of determination of overlap between two classes. In band x_1, we have:

$$\left(f_{min,11} < f_{min,12}\right) \,\&\, \left(f_{min,11} < f_{max,12}\right)$$

$$\&\, \left(f_{max,11} < f_{min,12}\right) \,\&\, \left(f_{max,11} < f_{max,12}\right)$$

Thus, in band x_1, classes are not overlapped while in band x_2, classes are overlapped and the overlap value between them is given by:

$$(OV_{12})_2 = \left|\max\left(f_{min,21}, f_{min,22}\right) - \min\left(f_{max,21}, f_{max,22}\right)\right|$$

$$= \left|f_{min,21} - f_{max,21}\right|$$

3. Experiments and discussion

In this section, we assessed the performance of proposed method, OFW, compared to some supervised feature extraction methods such as LDA, NWFE, GDA, and MMLDA using three real hyperspectral images: Indian, university of Pavia, and KSC datasets. The Indian Pines is provided by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over Northwestern Indiana. Indian image comprises 224 spectral bands, which are initially reduced to 200 by removing water absorption bands. This image has 145×145 pixels and 16 classes which 10 interesting classes of it are chosen for our experiments. The university of Pavia dataset is collected by the Reflective Optics System Imaging Spectrometer (ROSIS). The number of spectral bands in the original recorded image is 115 from which 103 bands are selected for analysis of data after the removal of noisy bands. This urban image has nine classes and 610×340 pixels. The KSC dataset is provided by AVIRIS over the Kennedy Space Center, Florida. After removing water absorption and low SNR bands, 176 bands are used for the analysis of data. The KSC image has 512×614 pixels and 13 classes.

Support vector machine (SVM) and Gaussian maximum likelihood (ML) are used as classifier to assess the performance of feature extraction methods. The polynomial with degree 3 with default parameters defined in LIBSVM [22] is used as kernel function in SVM classifier. We used some measures for assessment of classification accuracy: Average accuracy, average reliability, and kappa coefficient [23]. The reliability in a class is the number of testing samples that are correctly classified divided to the overall samples, which are classified in that class. We used the McNemars test [24] for assessment of statistical significance of differences in the classification results. The sign of Z_{12} indicates whether classifier 1 is more accurate than classifier 2 ($Z_{12} > 0$) or vice versa ($Z_{12} < 0$). The difference in classification accuracy between two classifiers is statistically significant if $|Z_{12}| > 1.96$. We used 16 training samples per class in our experiments to investigate the performance of feature extraction methods in small sample size situation. The training samples are chosen randomly from entire scene. We used the reminded samples as testing samples. We did each experiment 10 times and the average results are reported here.

Figures 4, 5, 6 show the average classification accuracy versus the number of extracted features with 16 training samples by a) SVM, b) ML classifiers for Indian, Pavia, and KSC datasets respectively. The accuracy and reliability of classes obtained by 16 training sample and SVM classifier for Indian (with 9 extracted features), Pavia (with 8 extracted features), and KSC (with 10 extracted features) are represented in tables 1, 2, and 3. The ground truth map (GTM) and classification maps for Indian and Pavia datasets are shown in figures 7, and 8 respectively. The highest classification accuracies achieved by 16 training samples for all feature extraction methods and hyperspectral images are shown in table 4. Table 5 shows the McNemars test results for different cases. The comparison of computation time of feature extraction processes is done in table 6.

We can see from the obtained results that OFW works better than other methods almost in all cases (only for Pavia urban image with ML classifier, GDA has better performance than other feature extraction methods). Popular feature extraction methods such as LDA, NWFE, GDA, and MMLDA calculate the scatter matrices and maximum the between-class scatter matrix and

minimum the within-class scatter matrix. The proposed method, OFW, calculates the weighted mean of adjacent spectral bands and considers the inverse of overlap between classes in each feature as a weight for that feature. LDA, NWFE, GDA, and MMLDA methods need to calculate the first and second order statistics of data (mean vectors and covariance matrices) while OFW does not need to estimate theses statistics. Therefore, when a limited number of training samples is available, the accurate estimations of mean vectors and scatter matrices cannot be provided. In these conditions, OFW is superior to LDA-based methods. However, with increasing the number of training samples, the accurate estimations of scatter matrices are obtained and the performance of LDA-based methods is improved.

The performance of OFW is compared with LDA in different training sample size for Indian dataset by SVM classifier and 9 extracted features and the results are shown in figure 9.

Moreover, the OFW method uses the simple calculations to obtain weight for each feature and calculates the weighted mean. Thus, it is faster than LDA, NWFE, GDA, and MMLDA methods which need to calculate scatter matrices. After OFW, LDA is faster than others. MMLDA, because of calculation of median-mean line, and GDA, because of calculations in kernel space, are slower than LDA. NWFE is the slowest method because it needs to calculate the weighted mean of all training samples to estimate the nonparametric scatter matrices.

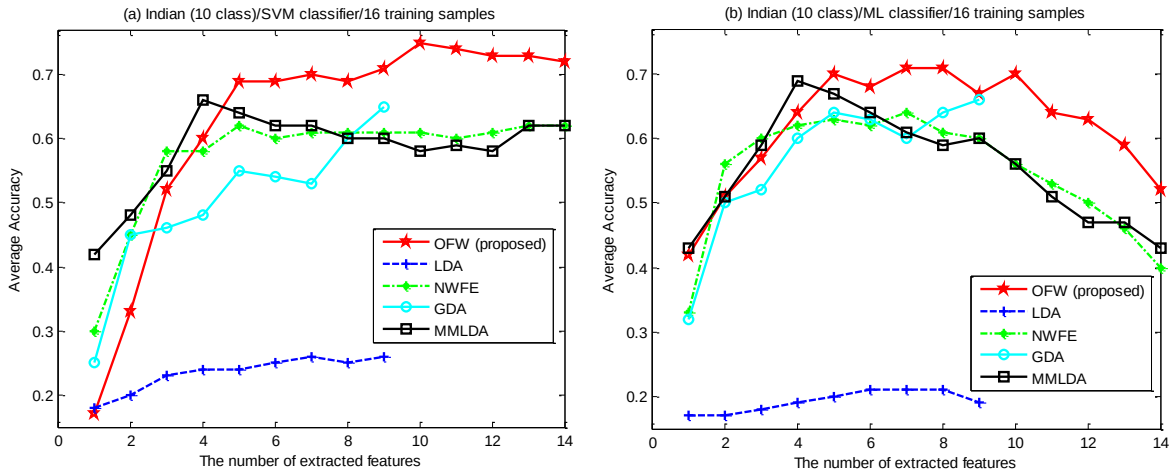

Figure 4. Average classification accuracy versus the number of extracted features obtained by a) SVM, b) ML classifiers for Indian dataset.

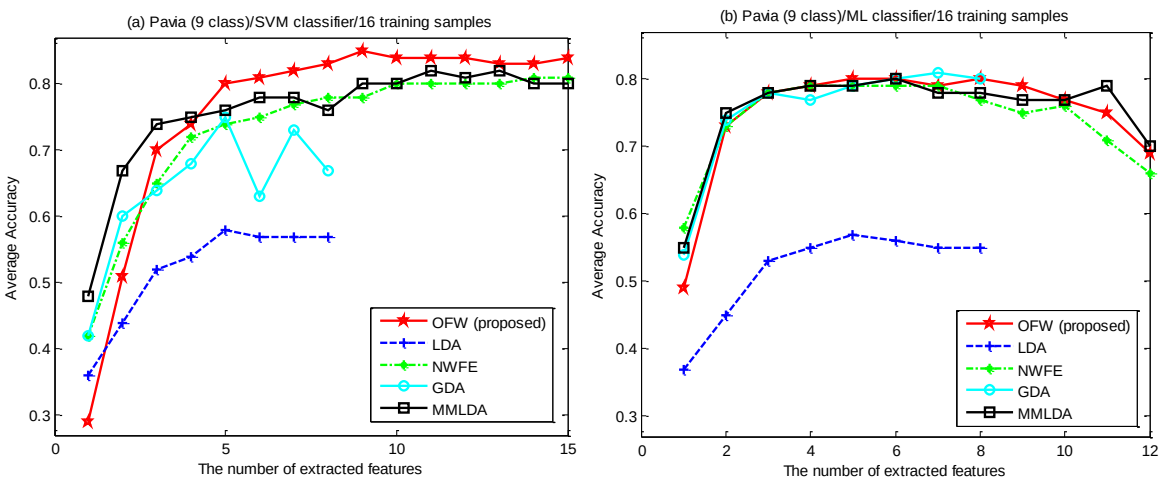

Figure 5. Average classification accuracy versus the number of extracted features obtained by a) SVM, b) ML classifiers for Pavia dataset.

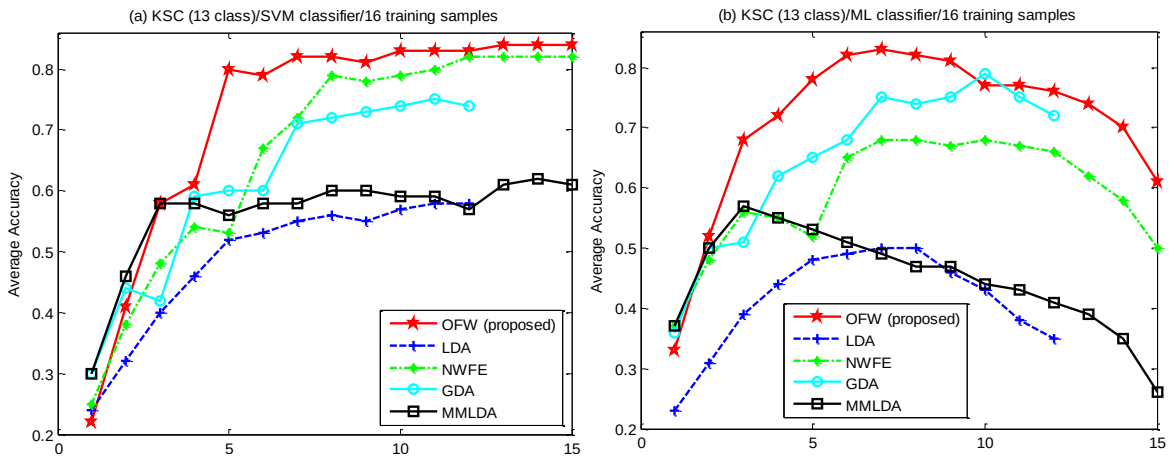

Figure 6. Average classification accuracy versus the number of extracted features obtained by a) SVM, b) ML classifiers for KSC dataset.

Table 1. Accuracy and reliability of classes of Indian dataset obtained by SVM classifier and 9 extracted features.

	class		**OFW (proposed)**		**LDA**		**NWFE**		**GDA**		**MMLDA**	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Corn-no till	1434	0.64	0.68	0.19	0.32	0.60	0.56	0.59	0.39	0.49	0.43
2	Corn-min till	834	0.64	0.40	0.24	0.18	0.55	0.40	0.58	0.51	0.63	0.38
3	Grass/pasture	497	0.93	0.67	0.32	0.14	0.50	0.40	0.83	0.58	0.74	0.50
4	Grass/trees	747	0.74	0.81	0.31	0.28	0.78	0.83	0.72	0.83	0.59	0.75
5	Hay-windrowed	489	0.99	1.00	0.20	0.31	0.95	1.00	0.99	0.99	0.97	0.98
6	Soybeans-no till	968	0.72	0.52	0.09	0.21	0.48	0.39	0.38	0.43	0.45	0.49
7	Soybeans-min till	2468	0.50	0.80	0.31	0.42	0.49	0.68	0.50	0.72	0.48	0.63
8	Soybeans-clean till	614	0.57	0.52	0.23	0.11	0.46	0.47	0.44	0.48	0.30	0.38
9	Woods	1294	0.89	0.95	0.28	0.61	0.68	0.83	0.89	0.92	0.90	0.90
10	Bldg-Grass-Tree-Drives	380	0.53	0.48	0.40	0.11	0.64	0.39	0.56	0.57	0.51	0.50
	Average Acc. and Average Rel.		**0.71**	**0.68**	**0.26**	**0.27**	**0.61**	**0.59**	**0.65**	**0.64**	**0.60**	**0.59**
	Kappa coefficient		**0.63**		**0.16**		**0.53**		**0.56**		**0.52**	

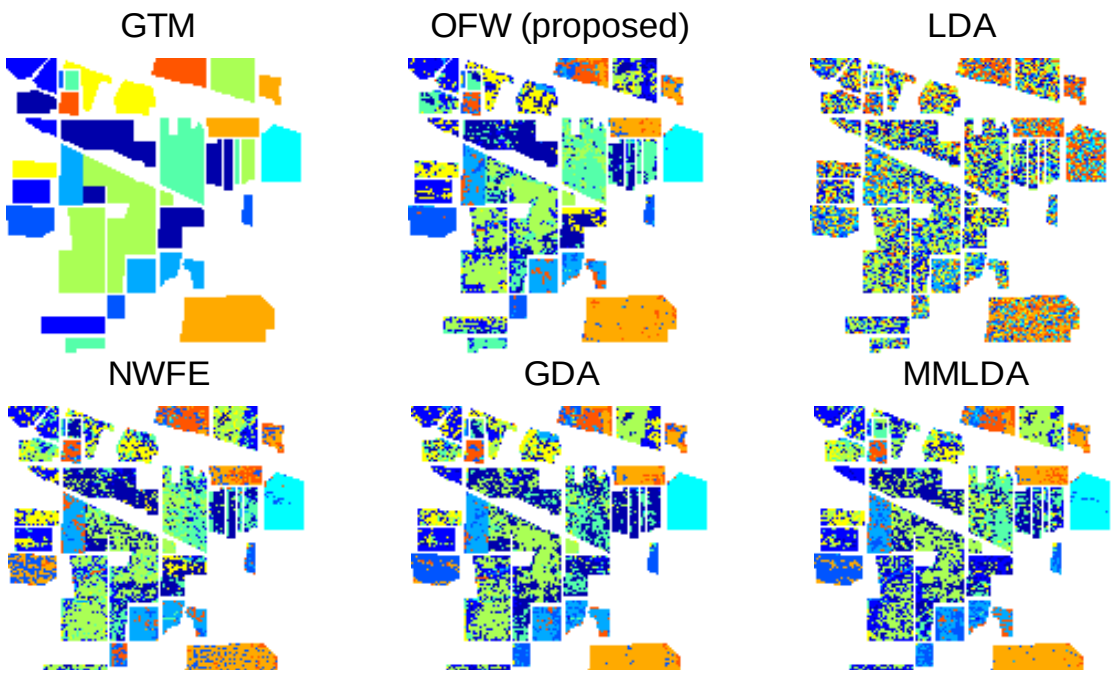

Figure 7. GTM and classification maps for Indian dataset obtained by SVM classifier and 9 extracted features.

Table 1. Accuracy and reliability of classes of Indian dataset obtained by SVM classifier and 9 extracted features.

	class		OFW (proposed)		LDA		NWFE		GDA		MMLDA	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Corn-no till	1434	0.64	0.68	0.19	0.32	0.60	0.56	0.59	0.39	0.49	0.43
2	Corn-min till	834	0.64	0.40	0.24	0.18	0.55	0.40	0.58	0.51	0.63	0.38
3	Grass/pasture	497	0.93	0.67	0.32	0.14	0.50	0.40	0.83	0.58	0.74	0.50
4	Grass/trees	747	0.74	0.81	0.31	0.28	0.78	0.83	0.72	0.83	0.59	0.75
5	Hay-windrowed	489	0.99	1.00	0.20	0.31	0.95	1.00	0.99	0.99	0.97	0.98
6	Soybeans-no till	968	0.72	0.52	0.09	0.21	0.48	0.39	0.38	0.43	0.45	0.49
7	Soybeans-min till	2468	0.50	0.80	0.31	0.42	0.49	0.68	0.50	0.72	0.48	0.63
8	Soybeans-clean till	614	0.57	0.52	0.23	0.11	0.46	0.47	0.44	0.48	0.30	0.38
9	Woods	1294	0.89	0.95	0.28	0.61	0.68	0.83	0.89	0.92	0.90	0.90
10	Bldg-Grass-Tree-Drives	380	0.53	0.48	0.40	0.11	0.64	0.39	0.56	0.57	0.51	0.50
	Average Acc. and Average Rel.		**0.71**	**0.68**	**0.26**	**0.27**	**0.61**	**0.59**	**0.65**	**0.64**	**0.60**	**0.59**
	Kappa coefficient		**0.63**		**0.16**		**0.53**		**0.56**		**0.52**	

Table 2. Accuracy and reliability of classes of Pavia dataset obtained by SVM classifier and 8 extracted features.

	class		OFW (proposed)		LDA		NWFE		GDA		MMLDA	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Asphalt	6631	0.90	0.87	0.37	0.70	0.78	0.86	0.18	0.53	0.73	0.80
2	Meadows	18649	0.67	0.91	0.33	0.91	0.48	0.85	0.63	0.79	0.54	0.86
3	Gravel	2099	0.71	0.63	0.44	0.30	0.73	0.53	0.80	0.45	0.65	0.48
4	Trees	3064	0.87	0.61	0.78	0.75	0.83	0.63	0.86	0.64	0.92	0.64
5	Painted metal sheets	1345	0.99	0.98	0.90	1.00	0.98	0.98	0.97	0.99	0.99	1.00
6	Bare Soil	5029	0.77	0.46	0.78	0.22	0.75	0.31	0.47	0.30	0.66	0.33
7	Bitumen	1330	0.80	0.80	0.37	0.22	0.81	0.64	0.46	0.13	0.65	0.43
8	Self-Blocking Bricks	3682	0.73	0.80	0.35	0.32	0.67	0.80	0.62	0.82	0.70	0.76
9	Shadows	947	1.00	0.97	0.78	0.71	1.00	0.99	1.00	1.00	1.00	1.00
	Average Acc. and Average Rel.		**0.83**	**0.78**	**0.57**	**0.57**	**0.78**	**0.73**	**0.67**	**0.63**	**0.76**	**0.70**
	Kappa coefficient		**0.70**		**0.36**		**0.57**		**0.47**		**0.58**	

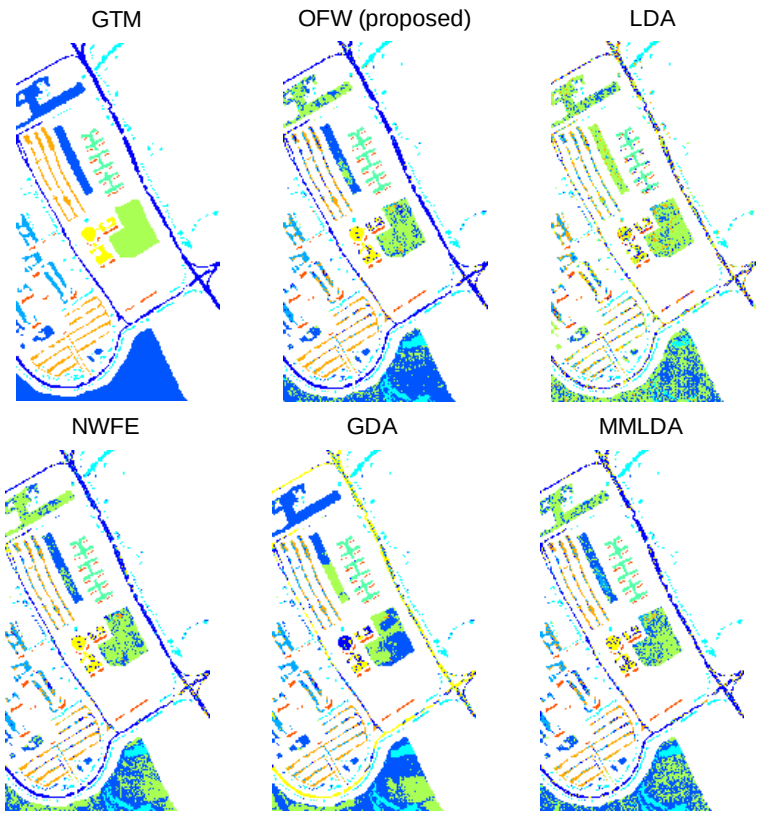

Figure 8. GTM and classification maps for Pavia dataset obtained by SVM classifier and 8 extracted features.

Table 3. Accuracy and reliability of classes of KSC dataset obtained by SVM classifier and 10 extracted features.

	lass		OFW (proposed)		LDA		NWFE		GDA		MMLDA	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Scrub	761	0.95	0.91	0.51	0.84	0.92	0.90	0.77	0.91	0.71	0.86
2	Willow swamp	243	0.89	0.92	0.51	0.54	0.92	0.86	0.84	0.84	0.65	0.50
3	Cabbage palm hammock	256	0.90	0.79	0.62	0.34	0.89	0.75	0.80	0.76	0.74	0.68
4	Cabbage palm/oak hammock	252	0.54	0.59	0.29	0.17	0.31	0.35	0.40	0.45	0.24	0.36
5	Slash pine	161	0.68	0.59	0.33	0.22	0.65	0.44	0.66	0.44	0.45	0.33
6	Oak/broadleaf hammock	229	0.56	0.75	0.36	0.33	0.43	0.73	0.42	0.35	0.48	0.32
7	Hardwood swamp	105	0.90	0.85	0.70	0.37	0.87	0.85	0.85	0.84	0.50	0.41
8	Graminoid marsh	431	0.80	0.83	0.58	0.58	0.82	0.77	0.43	0.74	0.45	0.38
9	Spartina marsh	520	0.94	0.87	0.55	0.53	0.88	0.85	0.89	0.83	0.86	0.73
10	Cattail marsh	404	0.87	0.99	0.75	0.96	0.76	0.89	0.75	0.84	0.23	0.39
11	Salt marsh	419	0.98	0.90	0.74	0.86	0.98	0.93	0.97	0.84	0.94	0.96
12	Mud flats	503	0.83	0.86	0.54	0.71	0.83	0.95	0.81	0.62	0.52	0.56
13	Water	927	0.98	0.98	0.99	0.99	0.99	0.98	0.97	1.00	0.94	0.96
	Average Acc. and Average Rel.		**0.83**	**0.83**	**0.57**	**0.57**	**0.79**	**0.79**	**0.74**	**0.73**	**0.59**	**0.57**
	Kappa coefficient		**0.86**		**0.59**		**0.83**		**0.75**		**0.62**	

Table 4. Highest classification accuracies achieved by 16 training samples (numbers in parentheses represent the number of features which obtain the highest average accuracies in experiments).

Dataset	Classifier	OFW (proposed)	LDA	NWFE	GDA	MMLDA
Indian	SVM	**0.75 (10)**	0.26 (7)	0.62 (5)	0.65 (9)	0.66 (4)
	ML	**0.71 (7)**	0.21 (6)	0.64 (7)	0.66 (9)	0.69 (4)
Pavia	SVM	**0.85 (9)**	0.58 (5)	0.81 (14)	0.75 (5)	0.82 (11)
	ML	0.80 (5)	0.57 (5)	0.79 (4)	**0.81 (7)**	0.80 (6)
KSC	SVM	**0.84 (13)**	0.58 (11)	0.82 (12)	0.75 (11)	0.62 (14)
	ML	**0.83 (7)**	0.50 (7)	0.68 (7)	0.79 (10)	0.57 (3)

Table 5. McNemars test results (Z_{rc} denotes each case of table where r is the row and c is the column).

Indian/SVM classifier/16 training samples/ 9 extracted features

	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	55.64	16.28	11.26	16.91
LDA	-55.64	0	-45.62	-49.19	-45.35
NWFE	-16.28	45.62	0	-5.90	0.54
GDA	-11.26	49.19	5.90	0	6.88
MMLDA	-16.91	45.35	-0.54	-6.88	0

Indian/ML classifier/16 training samples/ 7 extracted features

	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	62.41	15.16	19.60	15.08
LDA	-62.41	0	-54.47	-51.39	-54.19
NWFE	-15.16	54.47	0	6.15	0.10
GDA	-19.60	51.39	-6.15	0	-6.08
MMLDA	-15.08	54.19	-0.10	6.08	0

Pavia /SVM classifier/16 training samples/ 8 extracted features

	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	93.00	46.11	56.76	41.14
LDA	-93.00	0	-60.79	-36.21	-61.98
NWFE	-46.11	60.79	0	23.90	-2.65
GDA	-56.76	36.21	-23.90	0	-24.25
MMLDA	-41.14	61.98	2.65	24.25	0

Pavia /ML classifier/16 training samples/ 5 extracted features

	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	73.49	7.93	-9.72	2.47
LDA	-73.49	0	-69.13	-82.25	-72.50
NWFE	-7.93	69.13	0	-17.80	-5.64
GDA	9.72	82.25	17.80	0	11.91
MMLDA	-2.47	72.50	5.64	-11.91	0

KSC / SVM classifier/16 training samples/ 10 extracted features					
	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	30.71	6.03	17.66	29.21
LDA	-30.71	0	-26.81	-17.83	-3.54
NWFE	-6.03	26.81	0	11.53	25.19
GDA	-17.66	17.83	-11.53	0	16.57
MMLDA	-29.21	3.54	-25.19	-16.57	0

KSC / ML classifier/16 training samples/ 7 extracted features					
	OFW (proposed)	LDA	NWFE	GDA	MMLDA
OFW (proposed)	0	36.83	24.02	16.09	37.01
LDA	-36.83	0	-18.83	-27.59	0.02
NWFE	-24.02	18.83	0	-12.91	22.60
GDA	-16.09	27.59	12.91	0	30.03
MMLDA	-37.01	-0.02	-22.60	-30.03	0

Table 6. Comparison of computation time of feature extraction process obtained by 16 training samples and 6 extracted features for Indian dataset.

Indian, 16 training samples, 6 extracted features The comparison of computation time in feature extraction processes					
	OFW (proposed)	LDA	NWFE	GDA	MMLDA
Computation time (seconds)	0.24	0.56	88.34	0.67	0.66

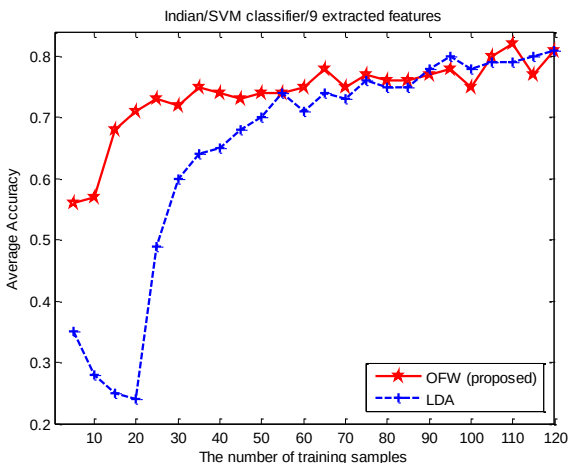

Figure 9. Comparison of OFW with LDA in different training sample size for Indian dataset obtained by SVM classifier and 9 extracted features.

4. Conclusion

The overlap-based feature weighting (OFW) is proposed for feature extraction of hyperspectral images in this paper. In the proposed method, the feature vector of each pixel is divided into some segments and the weighted mean of features in each segment is calculated as an extracted feature. The weight for each feature is obtained by calculation of overlap between classes in that feature. In the OFW method, there is no need to calculate the statistics of data. As a result, OFW is a simple, fast, and efficient for feature extraction of high dimensional data in small sample size situations. The superiority of OFW compared to some popular feature extraction methods is shown for Indian, Pavia, and KSC datasets using limited training samples.

References

[1] Dobigeon, N., Tourneret, J.-Y. & Chang, C.-I (2008). Semi-Supervised Linear Spectral Unmixing Using a Hierarchical Bayesian Model for Hyperspectral Imagery. IEEE Transactions on Signal Processing, vol. 56, no. 7, pp. 2684-2695.

[2] Yang, L., et al. (2014). Semi-Supervised Hyperspectral Image Classification Using Spatio-Spectral Laplacian Support Vector Machine. IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 3, pp. 651-655.

[3] Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 6, pp.1351–1362.

[4] Cortes, C., & Vapnik, V. (1995). Support vector networks. Machine Learning, vol. 20, no. 3, pp. 273–297.

[5] Imani, M. & Ghassemian, H. (2015). Feature space discriminant analysis for hyperspectral data feature reduction. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 102, pp. 1-13.

[6] Ghamary Asl, M., Mobasheri, M. R. & Mojaradi, B. (2014). Unsupervised Feature Selection Using Geometrical Measures in Prototype Space for Hyperspectral Imagery. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 7, pp. 3774-3787.

[7] Imani, M. & Ghassemian, H. (2015). Feature Extraction Using Weighted Training Samples. IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 7, pp. 1387-1391.

[8] Kamandar, M. & Ghassemian, H. (2013). Linear Feature Extraction for Hyperspectral Images Based on Information Theoretic Learning. IEEE Geoscience and Remote Sensing Letters, vol. 10, no. 4, pp. 702 - 706.

[9] Kuncheva, L. I. & Faithfull, W. J. (2014). PCA Feature Extraction for Change Detection in Multidimensional Unlabeled Data. IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 1, pp. 69-80.

[10] Wen, J., et al. (2013). Neighborhood Preserving Orthogonal PNMF Feature Extraction for Hyperspectral Image Classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 2, pp. 759-768.

[11] Imani, M. & Ghassemian, H. (2014). Band Clustering-Based Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples. IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 8, pp. 1325 - 1329.

[12] Cariou, C., Chehdi, K. & Le Moan, S. (2011). BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing. IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 3, pp. 565-569.

[13] Fan, W., Bouguila, N. & Ziou, D. (2013). Unsupervised Hybrid Feature Extraction Selection for High-Dimensional Non Gaussian Data Clustering with Variational Inference. IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 7, pp. 1670-1685.

[14] Izquierdo-Verdiguier, E., et al. (2014). Semisupervised Kernel Feature Extraction for Remote Sensing Image Analysis. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 9, pp. 5567-5578.

[15] Zhang, L., et al. (2012). On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 3, pp. 879-893.

[16] Zhang, L., et al. (2015). Ensemble Manifold Regularized Sparse Low-Rank Approximation for Multiview Feature Embedding. Pattern Recognition, vol. 48, pp. 3102–3112.

[17] Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition. California, USA. San Diego: Academic Press Inc.

[18] Baudat, G. & Anouar, F. (2000). Generalized discriminant analysis using a kernel approach. Neural Computation, vol. 12, no. 10, pp. 2385–2404.

[19] Kuo, B. C. & Landgrebe, D. A. (2004). Nonparametric weighted feature extraction for classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 5, pp. 1096-1105.

[20] Xu, J., et al. (2014). Median–mean line based discriminant analysis. Neurocomputing, vol. 123, pp. 233–246.

[21] Zortea, M., Haertel, V. & Clarke, R. (2007). Feature Extraction in Remote Sensing High-Dimensional Image Data. IEEE Geoscience and Remote Sensing Letters, vol. 4, no. 1, pp. 107-111.

[22] Chang, C. & Linin, C. (2008). LIBSVM—A Library for Support Vector Machines. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[23] Cohen, J. (1960). A coefficient of agreement from nominal scales. Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46.

[24] Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. Photogrammetric Engineering and Remote Sensing, vol. 70, no. 5, pp. 627–633.

نشریه هوش مصنوعی و داده کاوی

وزن دهی ویژگی بر مبنای هم‌پوشانی برای استخراج ویژگی تصاویر سنجش از دور ابرطیفی

مریم ایمانی* و حسن قاسمیان

دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران.

تاریخ دریافت: ۲۰۱۵/۰۴/۱۲؛ تاریخ پذیرش: ۲۰۱۵/۰۸/۲۶

چکیده:

سنجنده‌های ابرطیفی، داده را به طور همزمان در صدها باند طیفی مجاور و باریک دریافت می‌کنند. به دلیل مسئله نفرین ابعاد، استفاده از همه‌ی باندهای طیفی، ضرورتا منجر به بهبود نتایج نمی‌شود. استخراج ویژگی برای تحلیل تصاویرابرطیفی به یک موضوع مهم در طبقه بندی این داده‌ها تبدیل شده است. ما در این مقاله پیشنهاد داده‌ایم تا از وزن‌دهی ویژگی بر مبنای هم‌پوشانی برای استخراج ویژگی نظارت شده‌ی تصاویر ابرطیفی استفاده کنیم. در روش پیشنهادی، بردار ویژگی هر پیکسل از تصویر به چند بخش تقسیم بندی می‌شود و میانگین وزن‌دار باندهای طیفی مجاور در هر بخش به عنوان یک ویژگی استخراج شده در نظر گرفته می‌شود. در هر باند طیفی که هم‌پوشانی میان کلاس‌ها کمتر باشد، توانایی تفکیک کلاس‌ها در آن باند بیشتر خواهد بود. بنابراین، معکوس مقدار هم‌پوشانی در هر باند، به عنوان وزنی برای آن باند طیفی در نظر گرفته می‌شود. برتری روش پیشنهادی از نظر دقت و سرعت نسبت به سایر روش‌های استخراج ویژگی نظارت شده در شرایطی که از مجموعه نمونه‌ی آموزشی کوچکی استفاده می‌شود برای سه داده‌ی ابرطیفی واقعی در آزمایش‌ها نشان داده شده است.

کلمات کلیدی: تفکیک کلاسی، هم‌پوشانی، وزن‌دهی ویژگی، استخراج ویژگی، ابرطیفی