Differential weights-based band selection for hyperspectral image classification

Author

Liu, Yun, Wang, Chen, Wang, Yang, Bai, Xiao, Zhou, Jun, Bai, Lu

Published

2017

Journal Title

International Journal of Wavelets, Multiresolution and Information Processing

Version

Accepted Manuscript (AM)

DOI

10.1142/S0219691317500655

DIFFERENTIAL WEIGHTS BASED BAND SELECTION FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Yun Liu, Chen Wang, Yang Wang, Xiao Bai, Jun Zhou, Lu Bai

Abstract

Band selection plays a key role in the hyperspectral image classification since it helps to reduce the expensive cost of computation and storage. In this paper we propose a supervised hyperspectral band selection method based on differential weights, which depict the contribution degree of each band for classification. The differential weights are obtained in the training stage by calculating the sum of weight differences between positive and negative classes. Using the effective one-class Support Vector Machine, the bands corresponding to large differential weights are extracted as discriminative features to make the classification decision. Moreover, label information from training data is further exploited to enhance the classification performance. Finally, experiments on three public datasets, as well as comparison with other popular feature selection methods, are carried out to validate the proposed method.

I. INTRODUCTION

Hyperspectral imaging sensors collect rich information from across the electromagnetic spectrum, which is usually composed of hundreds of contiguous and narrow spaced spectral bands. Due to the phenomenon that certain objects have unique spectral signatures, hyperspectral

Yun Liu and Yang Wang are with the School of Mechanical Engineering and Automation, Beihang University, Beijing, 100191,China

Chen Wang and Xiao Bai are with the School of Computer Science and Engineering, Beihang University, Beijing, 100191,China

Jun Zhou is with the School of Information and Communication Technology, Griffith University, Nathan, Queensland 4111, Australia

Lu Bai is with the School of Information, Central University of Finance and Economics, Beijing,100081,China

imaging is useful in a wide array of applications, such as face recognition,[32], [34], [35] mineral recognition,[16] vegetation study[23] and land cover classification.[20]

One challenging problem in hyperspectral imaging classification is that the high dimensionality of the spectral data leads to expensive costs of computation and storage. The high correlation among spectral bands implies that redundant bands could be removed without decreasing classification performance. Therefore, band dimensionality reduction techniques are often used to address this problem, as well as an important preprocessing step for hyperspectral images. Specifically, there are two kinds of such technics. The first one is feature extraction, which transforms raw hyperspectral images from the original space to a feature space with lower dimensions. Algorithms like Principle Component Analysis (PCA)[1] and Independent Component Analysis (ICA)[6] are suitable for this kind of method. The drawback of feature extraction for hyperspectral image is that the transformed feature space no longer contains spectral information and lose the interpretability of original data. The second kind of method is band selection, which chooses some 'good' bands rather than all of them. Here, the 'good bands' means those bands which are discriminative but not redundant for classification.

Various band selection methods have been proposed and could further be categorized into unsupervised[4], [7], [9], [25], [27] and supervised[2], [15], [24], [26], [30], [33] manners. Unsupervised methods, such as Filter,[9] Wrapper ,[13], [14] are convenient to implement because it requires no labeled data. Filter methods use variable ranking for feature ordering. Typically, correlation criteria, e.g., mutual information,[9] is to score features and then remove those below a threshold. However, because some features appear informative when combined with others but not on their own, they may mistakenly be discarded by Filter methods. Wrapper methods use a predictor wrapped on a search algorithm to find the feature subset with the highest performance for the predictor. For hyperspectral band selection, various criteria, such as Bhattacharyya distance, divergence and Jeffries-Matusita (JM) distance,[13], [14] have been employed as the predictor. In addition to this, Yao et al.[31] described a nonlinear band selection method based on Bayesian kernel which can reduce the uncertainty by computation of posterior label probabilities. Pabitra et al.[17] proposed an unsupervised feature selection algorithm based on measuring similarity between features. Sebastiano et al.[22] proposed a new suboptimal search strategy for feature selection in very high dimensional remote sensing images. Unfortunately, enormous computations would be introduced since a new model is always created to evaluate the current subset each time during

its search procedure. And many hyperspectral imaging classification tasks prioritize the accuracy. Therefore, supervised methods are much preferable in such cases.

Supervised method[3], [10], [11], [28], [29], [31 tries to find the optimal feature subset via the classifier construction. In other words, they make the feature selection in the training process and thereby could choose desired features by assessing classification results. Compared with unsupervised approaches, Embedded methods not only consider the group correlation of features, but also spend less computation time. One of their typical schemes is to use the weights of a classifier to rank the feature importance. Options of classifier include Random Forest,[8] Support Vector Machine(SVM)[12] etc. It is interesting to notice that in these methods the weights of the classifier contain significant information and reveal the importance of features. Intuitively, larger weights indicate more importance that the feature has. On this basis, we argue that the performance could be further boosted by introducing differential weights criteria into the training process.

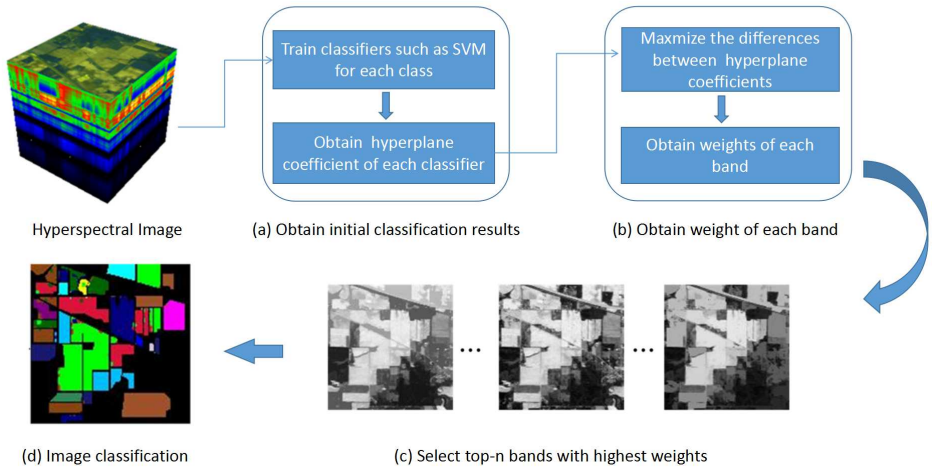


Fig. 1. Key steps of the proposed method: (a) Training one-vs-all classifiers using the original training set and get the hyperplane coefficients. (b) Using these coefficients to construct a optimization function to get weights which indicate the importance of each band for classification.(c) After optimization procedure, bands with large weights are selected for classification.(d)The hyperspectral image classification result using selected bands.

In this paper, we propose a supervised band selection method, which is originated from Embedded methods. When building band selection criteria, most attention is paid to the inter-class discrimination by optimizing differential weights. Specifically, discriminative bands herein

are defined as those that contribute more positive score to the classifier's output of one class than that of other classes in training. To formalize this observation, we use a modified one-class SVM to satisfy differential weights constraint and calculate weights for each band. The bands with large weights are then selected.

Our contributions are as follows: First, we propose a supervised hyperspectral band selection algorithm using differential weight criterion, which takes full advantage of the label information to improve the classification performance. Second, a band weights optimization process is introduced to find the band subset by using the one-class SVM model. Third, the proposed method calculates weight for each band, which gives guidance not only to classification but also to other band analysis approach.

II. Band selection with differential weights

In this section, we detailedly describe the method for hyperspectral band selection. At first, we briefly illustrate the concept of hyperspectral band weighting and describe the notation used in the rest of paper. Then a new band selection criterion based on differential weights is developed. After the criterion is defined, we introduce the one-class SVM model, which is used to solve the optimization of differential weights. Finally, informative bands are selected according to the band weights, and are used to train hyperspectral image classifiers. A summary of the proposed method is shown in Fig. 1

A. Band Weighting

Unlike the searching-based band selection method, we adopt a weighting-based strategy. It prioritizes all bands at the same time, taking each possible group of bands into consideration. This weighting-based band selection method could avoid the problem caused by correlation among these bands when selected one by one. Let (x^i, y^i) be the i-th pixel in the training set, where $x^i \in \mathbb{R}^N$ is the vector of i-th pixel's intensity value of all bands and N is the total number of original bands, $y^i \in \{1, 2, ..., C\}$ is the class label of the i-th pixel and C is the number of classes. The band weight v is a $N \times 1$ vector corresponding to the spectral bands of training samples. The next step is to design a separation criterion to learn the band weight that indicates the importance of each spectral band.

B. Differential Weights

Our motivation is to design a band selection criterion which includes inter-class constraints. So, Firstly, the differential weight is deduced.

As mentioned before, the key criterion to design a supervised hyperspectral band selection method is classification accuracy and it is natural to start by exploring classification result. At the beginning, we train an one-vs-all SVM classifier for each class using the data in this class as positive samples and the rest as negative samples. The objective function is

$$L(w) = \frac{1}{2}\|w\|^2 - \lambda\Sigma_i(y_i f(x_i, w) - 1) \tag{1}$$

where x_i represents the i-th hyperspectral pixel training sample with N bands, and w represents the SVM classification hyperplane. λ is a regularization parameter that controls the model complexity. y_i is the class label, and $f()$ is the classification score for the initial SVM training. It should be noted that other classifiers such as multi-class logistic regression could be used herein as well. After the initial training, each class c has a classification weight w_c.

In the multi-class classification problem, the sample is predicted correctly if the sample gets higher score of its ground truth class c than those of other classes. In other words, it means that the i-th sample would be predicted correctly if it satisfies the following inequality:

$$w_c^{\mathrm{T}} x^i > w_{y^i}^{\mathrm{T}} x^i, y^i \neq c \tag{2}$$

This gives us the insight of how to select discriminative bands. Intuitively, a band b is discriminative for a certain class c when the product of this band's intensity value and the classification weight of this certain class is higher than the product with any other classes' weight , i.e.,

$$w_c(b)x^i(b) > w_{y^i}(b)x^i(b), y^i \neq c \tag{3}$$

which could be transformed into the form

$$(w_c(b) - w_{y^i}(b))x^i(b) > 0, y^i \neq c \tag{4}$$

This observation of inter-class discriminability is used to build the band selection criterion and we call $w_c(b) - w_{y^i}(b)$ the differential weights which reveals the importance of each hyperspectral

band. Specifically, we take Eq. 4 as a constraint that each discriminative band should meet. For a certain spectral band, the greater the LHS of Eq. 4 is than 0, the higher probability it has to make the right prediction.

C. Optimization

Before going into the step of calculating band weights, we briefly introduce the one-class SVM,[21] which is the model used herein to solve the optimization process. The classical SVM model is treated in a binary classification setting. For example, class label $y^i \in \{-1, 1\}$, which denotes class A and class B respectively. However, one-class SVM deals with anomaly detection problem, where only data with class label $y^i \in \{1\}$ could be used. The objective function of one-class SVM could be written as:

$$\min_{w_o} \sum_i \max(0, 1 - w_o^{\mathrm{T}} x^i) + \alpha \|w_o\|_1 \tag{5}$$

where w_o is the classification weight vector and α is the regularization parameter. The model learns w_o for positive class to differentiate positive class from non-positive class. Note that the difference between Eq. 5 and the classical SVM is that the class label y^i is ignored since there is only positive class in one-class SVM model. Instead of treating w_o as the classification weight and x^i as the feature vector of i-th data sample, we set them, respectively, as band weight that implies the importance of the band and the discriminative ability of bands that is the indicator of band selection.

Finally, our goal is to learn a $N \times 1$ discriminative weight vector v for each class. Let D^c denote the index of training subset corresponding to class label c. We solve the following optimization problem for each class c:

$$\min_{v} \sum_{i \in D^c} \sum_{y^i \neq c} \max(0, 1 - v^{\mathrm{T}} u^i_{c,y^i}) + \lambda \|v\|_1 \tag{6}$$

where u^i_{c,y^i} represents the criterion:

$$u^i_{c,y^i}(b) = (w_c(b) - w_{y^i}(b)) x^i(b) \tag{7}$$

which is the differential weight times the pixel value which measures its discriminability to other classes of band b. Intuitively, The band b is helpful to make the correct prediction if $u^i_{c,y^i}(b)$ is

greater than 0. Large value of $u^i_{c,y^i}(b)$ indicates that the current band b contains the information we'd like to keep for maintaining classification performance.

In order to find discriminative bands, the constraints built on differential weights are needed. Intuitively, Eq. 6 tries to estimate vector v_c such that the b-th element of v_c represents the discriminative capability of band b to the c class. In other words, band b of the c class is discriminative to other classes if the b-th element of v_c is large. In Eq. 6, the first term is a hinge loss term which measures indiscriminateness of weight vector v_c. For example, when v_c gives big weights to the discriminative bands, the loss term will be low in Eq. 6. The second term is a regularization term, which prevents the optimization from over-fitting. Furthermore, we choose ℓ^1-norm as regularization so that sparsity is given to v_c, which promotes the ability of feature selection. The parameter λ controls the relative importance between hinge loss term and ℓ^1-norm regularization term and can be obtained by cross-validation. Note that Eq. 6 is equivalent to one-class SVM and could be solved by a standard SVM solver.[5]

After obtaining v_c for each class, we select the n bands with the n largest weights. The selected bands contain the most discriminative information and are efficient for classification.

III. Experiment

The proposed method is evaluated on three publicly available datasets. It should be noted that our band selection method is flexible and can be applied to a variety of classifiers although the simple yet effective one-class SVM is used in our method to gain discriminative band weights. Therefore, to validate the property, we discuss the classification accuracy herein by using K Nearest Neighbor(KNN), Linear Discriminant Analysis(LDA) and SVM, respectively. The parameter k in KNN is set to be 3 in the experiment, and the one-against-all strategy is utilized in both LDA and SVM training. Particularly, the regularization parameter in SVM is obtained by 5-fold cross-validation with the widely used radial basis function (RBF) kernel.

Besides, our method is also evaluated by comparing it with three popular band selection technics, i.e., AP,[19] mRMR[18] and SVM-RFE.[12] AP uses affinity propagation to calculate the similarity between bands and search for exemplars as selected bands. mRMR is a supervised band selection method based on representative mutual information. In SVM-RFE, SVM weights, similarly to our method, are also used as feature ranking criterion. However, it should be noted that we further exploit the trained classifier in the proposed method by using differential weights

to build the separation criterion. Finally, the classification results with full bands are reported as well.

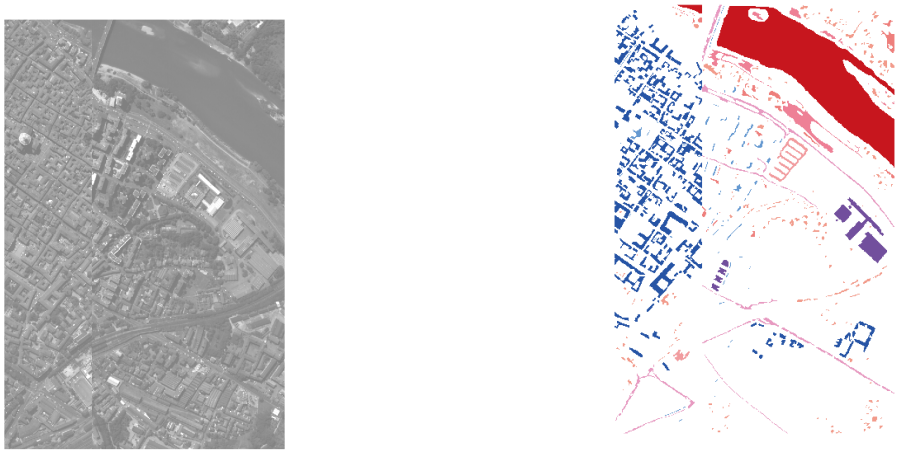
A. Pavia Centre

Pavia Centre is a section of the scene of Pavia, nothern Italy, which was acquired by the ROSIS sensor during a flight campaign in 2002. Fig. 2(a)shows the sample band and Fig. 2(b)shows the ground truth of the land cover class. It is a 1096x1096 pixels image with spatial resolution of 1.3 meters and each pixel has 102 spectral bands. The dataset is categorized into nine classes. In Pavia Centre, some samples without information are manually discarded. Then we randomly choose 15 percent of the remaining samples for each class as training set and the rest as test set. See Table. I for detail.

Fig. 3 shows experimental results of various band selection methods with different classifiers. It can be seen that our method outperforms the others in most cases. Especially, the SVM classifier achieves the highest classification accuracy slightly. The reason may be that SVM with RBF kernel, as a non-linear classifier, is more powerful than linear classifiers like LDA. In general, our method has consistent performance no matter what classifier is used herein. It could verify the independence of our method on classifiers to some extent. When the number of bands is small, SVM-RFE may be competitive with ours. However, as the number of bands increases, specifically when more than 15 bands are selected in this case, our method could achieve better performance. On the other hand, it can be easily seen that our method outperforms AP in any number of selected bands, which is reasonable since it makes the inter-class discrimination by optimizing the inter-class criterion. Moreover, classification accuracy of our method with only 40 bands is very close to that with full bands, which demonstrates heavy redundancy of hyperspectral images in classification application.

B. Indian Pines

The second dataset we used is Indian Pines, which is shown in Fig. 4(a). The ground truth land cover class is shown in Fig. 4(b). Indian Pines is a section of the scene acquired by AVIRIS sensor in 1992 over the Indian Pines test site, North-western Indiana. It is a 145*145 pixels image and each pixel has 224 spectral bands covering wavelength ranging from 400nm to 2500nm. We remove water absorption bands: [104-108], [150-163], 220. Furthermore, seven classes are



(a) (b)

Fig. 2. Pavia Centre: (a) Sample band of the hyperspectral image. (b) Ground truth of the land cover class.

TABLE I

NUMBER OF TRAINING AND TESTING SAMPLES FOR EACH OBJECT CLASS IN PAVIA CENTRE

#	Class	Training	Testing
1	Water	124	800
2	Trees	123	797
3	Asphalt	122	794
4	Self-Blocking Bricks	120	788
5	Bitumen	120	788
6	Tiles	189	1071
7	Shadows	71	405
8	Meadows	124	800
9	Bare Soil	123	797
	TOTAL	1118	6338

discarded since only few samples are available for them, leaving nine classes available in the dataset. We randomly choose 15 percent of the samples for each class as training set and the rest as test set. See Table. II for detail.

Fig.5 shows the classification accuracy using various band selection methods and different classifers. Our method proposed in the letter obviously outperforms AP and mRMR. Similarly

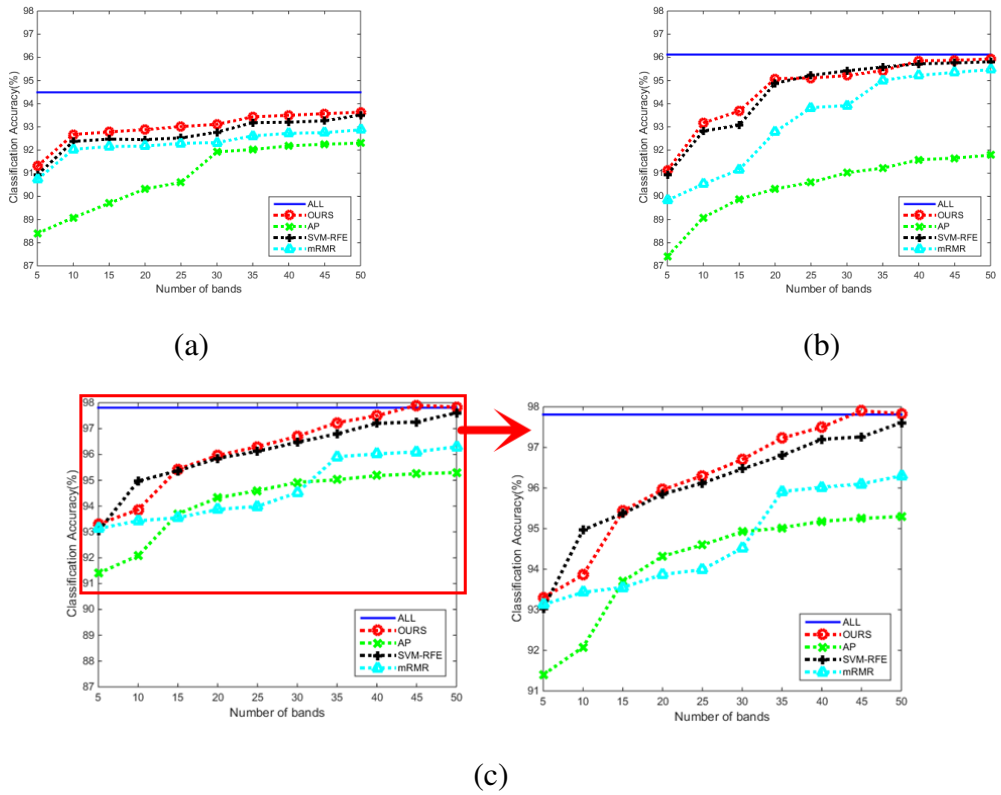


Fig. 3. Classification accuracy on Pavia Centre with (a)KNN, (b)LDA, and (C)SVM.

to the Pavia Centre dataset, SVM has the slightly highest performance among the three classifiers. Besides, the relative performance is also consistent given a classifier since our method is adaptable to a variety of classifiers. We can see that SVM-RFE is competitive to our method when using the KNN classifer, while our method performs slightly better than SVM-RFE for SVM. This is partly because the KNN classifier could suffer from over-fitting and be quite sensitive to noise. Moreover, in the SVM case, the performance of our method achieves better than other band selection methods as the number of bands increases. When the number of selected bands is 15, we report the selected bands by different methods in Table. IV.

C. APHI

The third dataset is call APHI(Airborne Push Hyperspectral Imager), which is shown in Fig. 6(a). The ground truth is shown in Fig. 6(b). Images in this dataset has the size of 210*150*64, covering 455nm to 805nm wavelength range. Some land-cover classes are dis-

(a) (b)

Fig. 4. Indian Pines: (a) Sample band of the hyperspectral image. (b) Ground truth of the land cover class.

TABLE II

NUMBER OF TRAINING AND TESTING SAMPLES FOR EACH OBJECT CLASS IN INDIAN PINES

#	Class	Training	Testing
1	Paddy	200	400
2	Bamboo	100	200
3	Tea	50	150
4	Pachyrhizus	100	250
5	Caraway	50	200
6	Water	200	400
	TOTAL	700	1600

carded because of there are only very few training samples. In APHI, the surface covers are Paddy,Bamboo,Tee, Pachysandras, Caraway and Water, Which are labeled as C4, T6, T7, V2, V13, and W2. Table. III shows the number of training and testing samples we choose from APHI dataset.

Fig. 7 shows classification results of various band selection methods with different classifiers. Among three classification methods, the SVM classifier shows the highest classification accuracy. Our band selection method outperforms the others, especially for AP method. Same as the first two datasets, SVM classifer has the best performance among the there classifiers. Moreover, when using SVM classifier, as the number of bands increased, specifically when more than 20

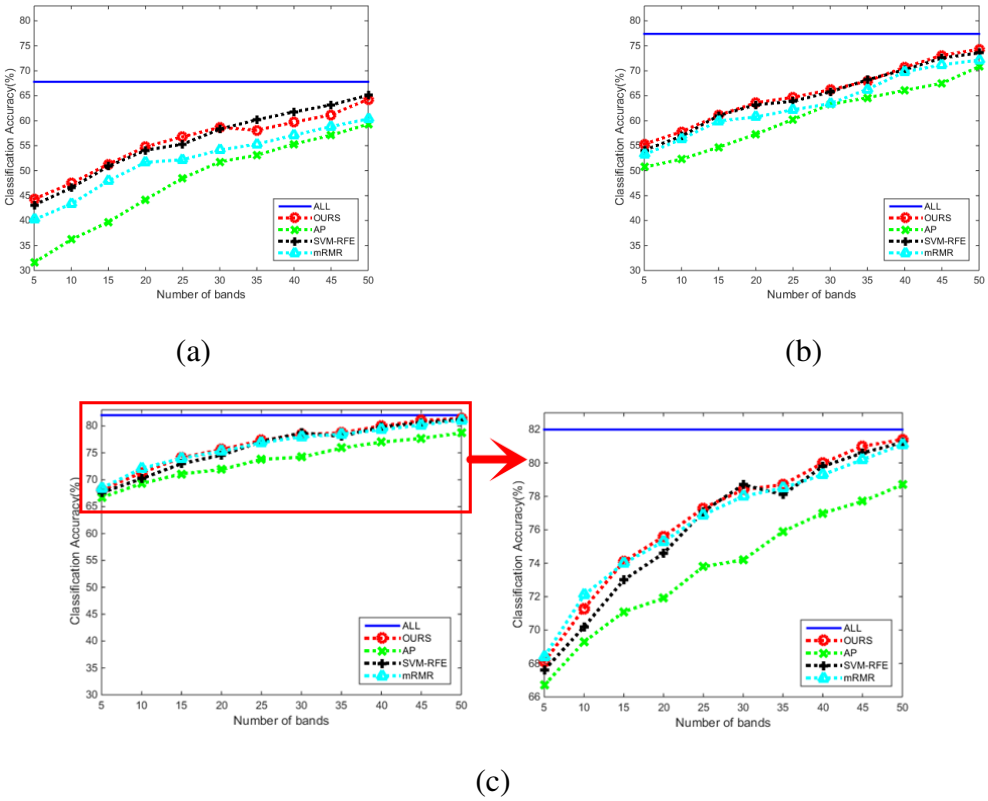


Fig. 5. Classification accuracy on Indian Pines with (a)KNN, (b)LDA, and (c)SVM.

bands are selected, our method can obtain could achieve better performance.

D. Influence of Training Set Size

In this subsection the influence of number of training samples on our band selection method is evaluated. We evaluate our method compared with SVM-RFE and mRMR, namely the supervised band selection methods. Fig.6. shows the SVM based classification performance with different numbers of training samples, where the Indian Pines dataset and 50 bands are used herein. It can be observed that classification accuracy of all three methods is low when only using a few of training samples, and then becomes higher with the increase of training set size. This is related to the over-fitting problem encountered in supervised learning when there are not enough training samples. Nevertheless, it is remarkable that even though training set is small, our method still performs better than SVM-RFE and mRMR because of its improved generalization.

(a) (b)

Fig. 6. APHI dataset: (a) Sample band of the hyperspectral image. (b) Ground truth of the land cover class.

TABLE III

NUMBER OF TRAINING AND TESTING SAMPLES FOR EACH OBJECT CLASS IN APHI

#	Class	Training	Testing
1	Corn-no till	214	1214
2	Corn-mil till	125	704
3	Grass/pasture	72	411
4	Grass/trees	110	620
5	Hay-windrowed	72	406
6	Soybean-no till	146	823
7	Soybean-mil till	368	2087
8	Soybean-clean till	89	504
9	Woods	190	1075
	TOTAL	1214	6882

TABLE IV

COMPARISON OF SELECTED BANDS WITH DIFFERENT METHODS IN INDIAN PINES

Methods	15 selected bands
OURS	10,23,39,50,59,80,90,95,102,116,120,129,132,188,196
AP	1,7,15,25,48,67,86,90,99,119,129,142,182,197,206
SVM-RFE	31,39,45,60,61,66,93,96,103,115,119,130,145,172,199
mRMR	37,53,67,68,69,82,87,98,101,112,115,119,127,137,186

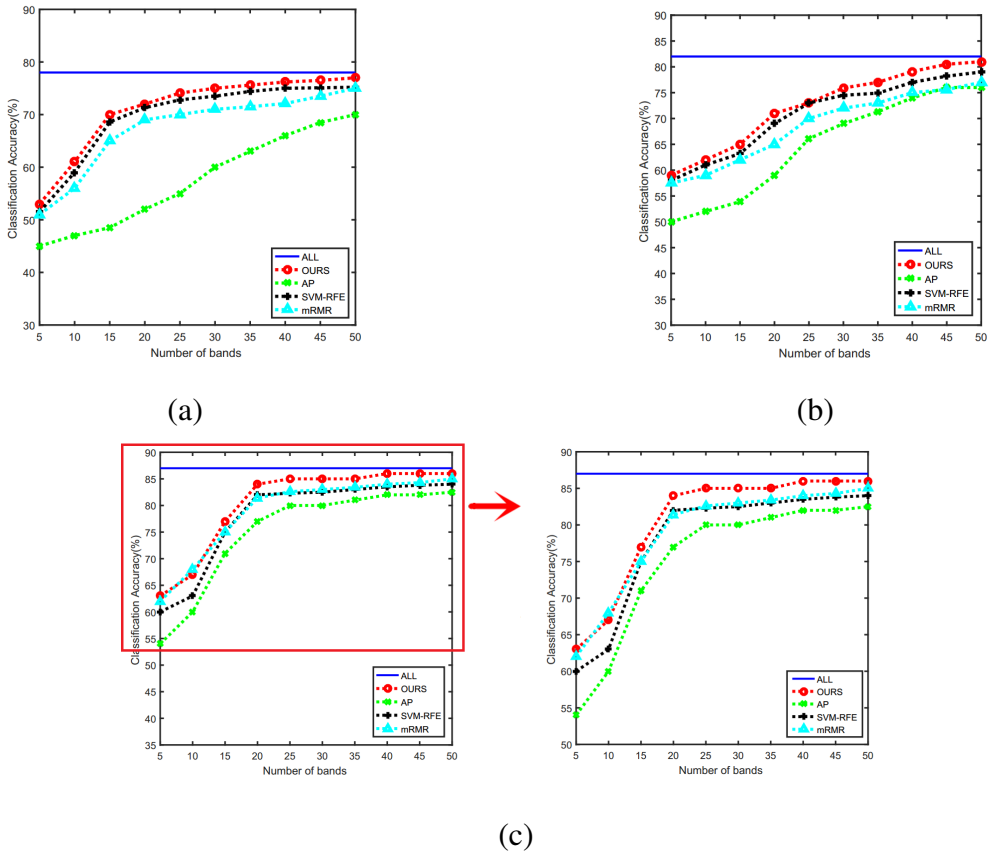


Fig. 7. Classification accuracy on APHI with (a)KNN, (b)LDA, and (C)SVM.

IV. Conclusion

In this letter we propose a new method for hyperspectral band selection. The differential weights are introduced to reveal the importance of each band in the selection criterion for classification. Experimental results in three widely used datasets show its effectiveness compared with other popular methods. The success of proposed method is due to the following reasons: 1) discriminative band is defined explicitly, which fully takes the discrimination among classes into account; 2) a modified one-class SVM is used to formalize the optimization problem so that discriminative weight vector for each class can be efficiently solved.

References

[1] A. Agarwal, T. El-Ghazawi, H. El-Askary and J. Le-Moigne, Efficient hierarchical-pca dimension reduction for hyperspectral imagery, *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, IEEE2007, pp.

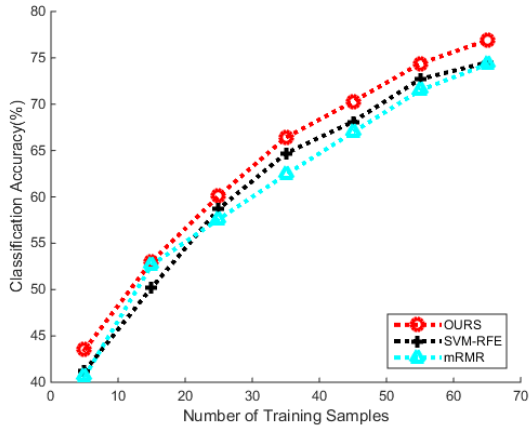


Fig. 8. Influence on performance of size of training set in Indian Pines. The same number of training samples per class is tested.

353–356.

[2] R. Archibald and G. Fann, Feature selection and classification of hyperspectral images with support vector machines, *IEEE Geosci. Remote Sensing Lett.* **4** (2007) 674–677.

[3] X. Bai, Z. Guo, Y. Wang, Z. Zhang and J. Zhou, Semi-supervised hyperspectral band selection via spectral-spatial hypergraph model2015.

[4] P. Bajcsy and P. Groves, Methodology for hyperspectral band selection, photogrammetric engineering and remote2004.

[5] C.-C. Chang and C.-J. Lin, Libsvm: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**(3) (2011) p. 27.

[6] Q. Du and H. Yang, Similarity-based unsupervised band selection for hyperspectral image analysis, *Geoscience and Remote Sensing Letters, IEEE* **5**(4) (2008) 564–568.

[7] X. Geng, K. Sun, L. Ji and Y. Zhao, A fast volume-gradient-based band selection method for hyperspectral image, *Geoscience and Remote Sensing, IEEE Transactions on* **52**(11) (2014) 7111–7119.

[8] K. Ghariani, N. Chehata, A. Le Bris and P. Lagacherie, Agricultural field delimitation using active learning and random forests margin, *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, IEEE2014, pp. 1717–1720.

[9] B. Guo, S. R. Gunn, R. Damper and J. Nelson, Band selection for hyperspectral image classification using mutual information, *Geoscience and Remote Sensing Letters, IEEE* **3**(4) (2006) 522–526.

[10] Z. Guo, X. Bai, Z. Zhang and J. Zhou, A hypergraph based semi-supervised band selection method for hyperspectral image classification, *2013 IEEE International Conference on Image Processing* (2013) 3137–3141.

[11] Z. Guo, H. Yang, X. Bai, Z. Zhang and J. Zhou, Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models, *IGARSS*, 2013.

[12] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene selection for cancer classification using support vector machines, *Machine learning* **46**(1-3) (2002) 389–422.

[13] R. Huang and M. He, Band selection based on feature weighting for classification of hyperspectral data, *Geoscience and*

Remote Sensing Letters, IEEE **2**(2) (2005) 156–159.

[14] A. Ifarraguerri and M. W. Prairie, Visual method for spectral band selection, *Geoscience and Remote Sensing Letters, IEEE* **1**(2) (2004) 101–106.

[15] A. Le Bris, N. Chehata, X. Briottet and N. Paparoditis, A random forest class memberships based wrapper band selection criterion: Application to hyperspectral, *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*, IEEE2015, pp. 1112–1115.

[16] I. Makki, R. Younes, C. Francis, T. Bianchi and M. Zucchetti, A survey of landmine detection using hyperspectral imaging, *ISPRS Journal of Photogrammetry and Remote Sensing* **124** (2017) 40–53.

[17] P. Mitra, C. A. Murthy and S. K. Pal, Unsupervised feature selection using feature similarity, *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (2002) 301–312.

[18] H. Peng, F. Long and C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **27**(8) (2005) 1226–1238.

[19] Y. Qian, F. Yao and S. Jia, Band selection for hyperspectral imagery using affinity propagation, *IET Computer Vision* **3**(4) (2009) 213–222.

[20] R. Roscher and B. Waske, Shapelet-based sparse representation for landcover classification of hyperspectral images, *IEEE Transactions on Geoscience and Remote Sensing* **54**(3) (2016) 1623–1634.

[21] B. Schölkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett, New support vector algorithms, *Neural computation* **12**(5) (2000) 1207–1245.

[22] S. B. Serpico and L. Bruzzone, A new search algorithm for feature selection in hyperspectral remote sensing images, *IEEE Trans. Geoscience and Remote Sensing* **39** (2001) 1360–1367.

[23] D. Stratoulias, I. Keramitsoglou, P. Burai, L. Csaba, A. Zlinszky, V. Tóth and H. Balzter, A framework for lakeshore vegetation assessment using field spectroscopy and airborne hyperspectral imagery, *Earth Observation for Land and Emergency Monitoring* (2017) p. 159.

[24] Y. Y. Tang, Y. Tao and E. C. M. Lam, New method for feature extraction based on fractal behavior, *Pattern Recognition* **35** (2002) 1071–1081.

[25] Y. Y. Tang, L.-T. Tu, J. Liu, S.-W. Lee, W.-W. Lin and I.-S. Shyu, Offline recognition of chinese handwriting by multifeature and multilevel classification, *IEEE Trans. Pattern Anal. Mach. Intell.* **20** (1998) 556–561.

[26] Y. Y. Tang, L. Yang and J. Liu, Characterization of dirac-structure edges with wavelet transform, *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* **30 1** (2000) 93–109.

[27] Y. Y. Tang and X. You, Skeletonization of ribbon-like shapes based on a new wavelet function, *IEEE Trans. Pattern Anal. Mach. Intell.* **25** (2003) 1118–1133.

[28] C. Yan, X. Bai, P. Ren, L. Bai, W. Tang and J. Zhou, Band weighting via maximizing interclass distance for hyperspectral image classification, *IEEE Geoscience and Remote Sensing Letters* **13**(7) (2016) 922–925.

[29] C. Yan, X. Bai and J. Zhou, Band weighting and selection based on hyperplane margin maximization for hyperspectral image classification, *IGARSS*, 2015.

[30] H. Yang, Q. Du, H. Su and Y. Sheng, An efficient method for supervised hyperspectral band selection, *Geoscience and Remote Sensing Letters, IEEE* **8**(1) (2011) 138–142.

[31] F. Yao and Y. Qian, Band selection based gaussian processes for hyperspectral remote sensing images classification, *2009 16th IEEE International Conference on Image Processing (ICIP)* (2009) 2845–2848.

[32] T. Zhang, B. Fang, Y. Y. Tang, G. He and J. Wen, Topology preserving non-negative matrix factorization for face recognition, *IEEE Transactions on Image Processing* **17** (2008) 574–584.

[33] T. Zhang, B. Fang, Y. Y. Tang, Z. Shang and B. Xu, Generalized discriminant analysis: A matrix exponential approach, *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* **40 1** (2010) 186–97.

[34] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li and F. Lang, Multiscale facial structure representation for face recognition under varying illumination, *Pattern Recognition* **42** (2009) 251–258.

[35] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang and X. Liu, Face recognition under varying illumination using gradientfaces, *IEEE Transactions on Image Processing* **18** (2009) 2599–2606.