PFG 2014 / X, 0027–0039
Stuttgart, February 2014

Article

An AdaBoost Ensemble Classifier System for Classifying Hyperspectral Data

Pouria Ramzi, Farhad Samadzadegan, Tehran, Iran & Peter Reinartz, Oberpfaffenhofen

Keywords: AdaBoost, band clustering, hyperspectral data, multiple classifier systems, support vector machines

Summary: This paper presents a new multiple classifier system based on AdaBoost to overcome the high dimensionality problem of hyperspectral data. The hyperspectral data are first split into a number of band clusters based on the similarities between the contiguous bands, and each band group is considered as an independent data source. The redundant bands in each cluster are then removed using branch and bound technique. Next, a support vector machine (SVM) is applied to each cluster and the outputs are combined using the weights calculated in AdaBoost iterations. Experimental results with AVIRIS and ROSIS datasets clearly demonstrate the superiority of the proposed algorithm in both overall and single class accuracies when compared to other multiple classifier systems. For AVIRIS data, which contains classes with greater complexity and fewer available training samples, the differences between the overall accuracies of the AdaBoost results are significantly higher compared to those of the other methods, and more pronounced than for the other dataset. In terms of class accuracies, the proposed AdaBoost approach also outperforms other methods in most of the classes.

Zusammenfassung: *Klassifizierung von Hyperspektraldaten mit einem multiplen Klassifizierungssystem auf AdaBoost Basis.* In diesem Beitrag wird ein neues multiples Klassifizierungssystem auf AdaBoost Basis entwickelt, um das Problem der hohen Dimensionalität von Hyperspektraldaten zu verringern. Die Hyperspektraldaten werden zunächst in eine Reihe von Spektralkanal-Clustern unterteilt, welche auf Ähnlichkeiten in benachbarten Kanälen beruhen. Jedes Cluster wird als unabhängige Daten-Teilmenge für die weitere Verarbeitung verwendet. Zunächst werden die redundanten Kanäle in jedem Cluster entfernt und jeweils ein Support-Vector-Machine (SVM) Klassifizierungsalgorithmus angewendet. Die Ergebnisse werden gewichtet kombiniert, wobei die jeweiligen Gewichte aus den AdaBoost Iterationen abgeleitet werden. Experimentelle Ergebnisse mit AVIRIS und ROSIS Datensätze zeigen deutlich die Überlegenheit des vorgeschlagenen Algorithmus im Vergleich zu anderen multiplen Klassifizierungssystemen, sowohl bei der Klassifizierungsgenauigkeit einzelner Klassen als auch bei der Gesamtgenauigkeit. Für AVIRIS Daten, welche Klassen mit höherer Komplexität enthalten und für die weniger Trainingsdaten zur Verfügung stehen, sind die Gesamtgenauigkeiten des AdaBoost Verfahrens signifikant höher und deutlich auffallender als bei dem anderen verwendeten Datensatz. In Bezug auf die Genauigkeit einzelner Klassen übertrifft der vorgeschlagene AdaBoost Ansatz die anderen Methoden ebenfalls für die meisten Einzelklassen.

1 Introduction

Hyperspectral sensors make it possible to obtain high-dimensional data with high spectral and spatial resolutions, which makes discrimination among similar land-cover classes easier. However, the large number of bands can become a drawback when classifying hyperspectral data using statistical methods (Landgrebe 2003, Richards 2013). In other words,

DOI: 10.1127/1432-8364/2014/0205 1432-8364/14/0205 $ 3.50

when the number of training samples is relatively small with respect to the number of features, the well-known problem of the curse of dimensionality, also known as the Hughes phenomenon, occurs (Hughes 1968). In this situation, the model is overfitted to the training data and this can lead to poor generalisation capabilities for the classifier. In recent years, extensive research has been performed in the area of hyperspectral image classification (Chi et al. 2008, Xie et al. 2011, Villa et al. 2011).

One of the most popular and well-known classification methods that have been successfully applied on hyperspectral data are support vector machines (SVMs). They were demonstrated to perform better than, or at least equivalently to, many other classifiers when applied to hyperspectral data (Braun et al. 2012, Camps-Valls & Bruzzone 2005, Li et al. 2013, Tarabalka et al. 2010). Another category of classification techniques that is highly capable of classifying hyperspectral data are multiple classifier systems (MCSs) or classifier ensembles, which are used to make highly accurate learning algorithms by combining a set of moderately accurate classifiers (Briem et al. 2002, Ceamanos et al. 2010, Chan & Paelinckx 2008). Among the ensemble methods, one of the most popular ones is AdaBoost introduced by Freund & Schapire (1999), which makes use of a set of component classifiers by changing the weights of training samples during the boosting iterations.

Because SVMs and AdaBoost work in different ways, combining them to benefit from the capabilities of both seems desirable. It has been demonstrated that the effectiveness of ensemble methods, such as AdaBoost, depends on both the accuracies of component classifiers and the diversities between them; in other words, the ensemble classification performance is only improved if the accuracy and diversity of ensemble classifiers are well-balanced (Prasad & Bruce 2008, Valentini & Dietterich 2004). In our experiment, diverse SVM classifiers are trained by splitting the hyperspectral bands into several band groups based on the similarities between the adjacent bands and removing the non-informative bands in each cluster and considering the SVM applied on each set as a component classifier in boosting. Because each band cluster has different spectral properties, the SVM classifiers applied to them seem to be diverse enough to satisfy the ensemble system. One of the main strengths of this technique in comparison to other ensemble methods is that for each iteration in AdaBoost, the system will concentrate on previously misclassified samples and attempt to learn them in different parts of that data with different spectral characteristics. In addition to this, SVM kernel parameters for each data cluster are defined according to its characteristics. It is more reasonable than using just only one pair of kernel parameters for training the whole dataset. The results of this system are subsequently compared with another multiple classifier system utilising majority voting to combine decisions of SVM component classifiers applied on band clusters (which is called the MV-SVM), a single SVM applied on all bands of the datasets (known as the SVM-All in literature), and a single SVM applied on the selected bands from different clusters of datasets (the SVM-Sel in this paper) in terms of overall, average and single class accuracies, kappa parameters and training times.

Following the introduction, the concepts of band clustering of hyperspectral data, SVMs, and MCSs are reviewed, respectively. Then, the proposed AdaBoostSVM technique to classify hyperspectral data is described in section 2. To evaluate the performance of the proposed algorithm and compare it with other classification methods, the results of applying these classifiers to two hyperspectral datasets are presented in section 3. Lastly, in section 4 a general discussion and conclusions are drawn.

1.1 *Band Clustering*

The aim of band clustering for hyperspectral data is that adjacent highly correlated bands should be merged into one group and bands with little redundancy should be separated into different groups (Li et al. 2011). This process guarantees that different subsets (clusters) have spectral heterogeneity for further tasks. In recent years, many studies have proposed dividing hyperspectral bands into different similar groups as a primary step for feature selection. Zhao and colleagues partitioned the hyperspectral data into several groups using correlation coefficients. A non-parametric clustering method was then used to extract the joint spatial-spectral features of the hyperspectral data (Zhao et al. 2011). In Martinez-Uso et al. (2007), the bands are grouped using information measures such as mutual information (MI) or the Kullback–Leibler divergence to reduce data redundancy and non-useful information. To calculate MI, entropy and joint entropy measures of spectral bands are calculated. If a discrete random variable x has the probability density function p(x), the entropy of X is defined as:

$$H(X) = -\sum p(x) \ \log(p(x)) \tag{1}$$

For two discrete random variables X and Y, with joint probability density function p(x, y), the joint entropy of X and Y is defined as:

$$H(X,Y) = -\sum\sum p(x,y)\log(p(x,y)) \tag{2}$$

The MI is used to measure the correlation between two random variables and it is defined as:

$$\begin{aligned} MI(X,Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \end{aligned} \tag{3}$$

The redundancy between two bands is greater when the value of MI is larger (Li et al. 2011).

Guo et al. (2006) further found that when the bands are highly correlated, the grouping based on a simple criterion such as correlation coefficient matrices or mutual information by itself would not be suitable as a similarity measure. The reason is that it can be low because either the two bands present a weak relation (such as it should be desirable) or the entropies of these variables are small (in such a case, the variables contribute with little information). Considering this, in Bigdeli et al. (2013) and Li et al. (2011) mutual information is initially employed to partition the bands into disjoint subspaces and a band selection technique is then employed to search for the optimal combination of bands.

After clustering, each cluster is normally treated as a feature set in an ensemble classification system (Bigdeli et al. 2013, Liao & Moody 1999, Martinez-Uso et al. 2007).

1.2 *SVMs*

The aim of SVMs is to separate two classes by fitting an optimal linear separating hyperplane to the training samples. The optimisation problem is being solved to maximise the margins between the hyperplane. If the samples are not linearly separable in the original space, kernel functions are used to map data into higher dimension where they can be separated with linear decision functions. This space is called Hilbert space (Vapnik & Vapnik 1998). The most widely used kernel in remote sensing is the Gaussian radial basis function (RBF) (Bigdeli et al. 2013, Ceamanos et al. 2010, Samadzadegan et al. 2012, Schölkopf & Smola 2001). For RBF-SVMs, the model parameters include the Gaussian width σ and the regularisation parameter C. Because SVMs are now very common in remote sensing communities, we omit their principles and concepts here. Instead, the reader is referred to Scholkopf & Smola (2001) and Watanachaturaporn & Arora (2004). SVMs are inherently binary classifiers, but they can be extended to solve multiclass problems. One method for this support is based on the combination of binary classifiers. These concepts, one-against-one and one-against-all methods are demonstrated in Melgani & Bruzzone (2004) and in Watanachaturaporn & Arora (2004).

SVMs work satisfactorily when small training sets are available on high-dimensional feature spaces (Pal & Mather 2006) and thus have attracted increased attention in remotely sensed hyperspectral communities. Melgani & Bruzzone (2004) applied SVMs for classification of hyperspectral data and obtained better classification results compared to other common classification methods. In a work presented by Braun et al. (2010), SVMs are utilised to classify vegetation from hyperspectral data and the results are compared with other classifiers such as maximum likelihood and spectral angle mapper. Tarabalka et al. (2010) presented a method for spectral-spatial classification of hyperspectral images using SVMs. These authors' method offered improved accuracy in comparison to some other classification approaches.

1.3 *MCSs*

MCSs represent approaches that use more than one classifier and combine their decisions with the goal of achieving more accurate results. These systems have been recently reviewed in the context of remote sensing and yield satisfactory results when dealing with hyperspectral and multi-source data (Ceamanos et al. 2010, Chan & Paelinckx 2008).

Ensemble classification methods are divided into two primary categories. In the first group, different learning algorithms are applied on the same training set, and their decisions are later combined (Benediktsson & Kanellopoulos 1999). The second approach is based on only one learning algorithm, and the ensemble is created by changing the training set or the feature subsets. The drawback of ensembles using different learning algorithms for analysis of hyperspectral data is that they add greater computational burden to a procedure already complicated by high-dimensional inputs (Chan & Paelinckx 2008). As a result, in most remote sensing research based on combining classifiers, the second concept is utilised (Bigdeli et al. 2013, Camps-Valls & Bruzzone 2005, Ceamanos et al. 2010).

Bagging (Breiman 1996) and boosting (Freund & Schapire 1999) are two main methods of the second approach and are reported to be effective in increasing classification accuracy (Opelt et al. 2006, Viola & Jones 2001). The most popular boosting algorithm is AdaBoost (adaptive boosting) which has been extensively used in different applications such as remote sensing in recent years (Camps-Valls & Bruzzone 2005, Chan & Paelinckx 2008, Frick et al. 2011). The algorithm takes a training set and a distribution or a set of weights over the training set as inputs. AdaBoost then calls the learning algorithm in a series of rounds. With each round, the weights of incorrectly classified examples are increased such that the weak learner is forced to focus on the hard examples in the training set. Finally, the classifiers of different iterations are combined with weighted voting (Freund & Schapire 1999). AdaBoost is initially defined to solve binary problems; however, it can be generalised to perform multiclass classifications. The most straightforward generalisations are known as AdaBoost.M1, AdaBoost.M2, and AdaBoost.MH (Freund & Schapire 1999, Schapire & Singer 1998).

1.4 *Multiple Classifier Systems with SVMs as Base Classifiers*

The concepts of SVMs and MCSs are based on different ideas. Nevertheless, the two approaches are not exclusive, and combining them in a complementary approach seems desirable (Waske et al. 2010). Several studies have tried to use SVMs in multiple classifier systems (Ceamanos et al. 2010, Li et al. 2008, Valentini & Dietterich 2004). According to these researches, the use of SVMs in ensemble classifier systems seems to be controversial (Li et al. 2008); in several studies, the classification performance of ensembles is better (Bigdeli et al. 2013, Ceamanos et al. 2010), whereas other research, the results of SVM ensembles are not better than that of a single SVM (Wang et al. 2009). In other words, it seems that some considerations must be taken into account when using SVMs as base classifiers in an ensemble system. In Valentini & Dietterich (2004), the effectiveness of ensemble methods depends on the accuracy, diversity and learning characteristics of base learners. Li and colleagues further proved that the ensemble can exhibit a good performance only when the accuracy and diversity of the classifiers are well-balanced (Li et al. 2008). Diversity is one of the most important factors in the success of every ensemble classifier system. Diversity means the errors of different classifiers should be uncorrelated or different classifiers should create errors for different data samples (Li et al. 2008, Kuncheva & Whitaker 2003). The required diversity among the component classifiers in an ensemble system can be satisfied by changing the training samples (Kuncheva & Whitaker 2003), thereby weakening the base classifiers (Wickramaratna et al. 2001), or changing the feature sets (Bigdeli et al. 2013, Ceamanos et al. 2010).

2 Proposed Method

In this paper, an AdaBoost ensemble classification method based on SVMs as base classifiers is proposed to classify the clusters extracted from hyperspectral bands. An overview of our proposed method is illustrated in Fig. 1.

The proposed strategy starts by splitting the hyperspectral data into a few band clusters based on MI between the contiguous bands. Afterwards, the rough hyper-parameters for each data source are calculated using a 5-fold cross validation technique also known as grid search. Because the bands in each cluster are highly correlated, they may contain data redundancies. To reduce the dimension of each cluster, a band selection technique is applied, and only the useful bands are retained in each of the clusters. Among these feature selection techniques, the optimal ones are the exhaustive search and the branch and bound (BB) algorithm (Guyon & Elissee 2003). When the dimensionality of the original feature space is large, the BB algorithm is preferred to the exhaustive search methods. Exhaustive search is only applicable for low-dimension problems (Li et al. 2011) because the number of possible feature sets that need to be searched becomes excessively large as the dimensionality of the original feature space increases. Hence, similar to Li et al. (2011), we utilise the adaptive branch and bound (ABB) algorithm proposed in Nakariyakul & Casasent (2007). Finally, the results of the band selection stage together with the rough kernel parameters for all the clusters lead to the final decision in AdaBoost.

2.1 *Feature/Band Selection*

In our algorithm, the most informative set of bands in each cluster is defined to maximise the classification performance of each ensemble classifier. Because the bands in each cluster are highly correlated, they may contain data redundancies. To reduce the dimension of each cluster, the adaptive branch and bound (ABB) technique is applied, and only the useful bands are retained in each cluster. In the ABB algorithm, one band is pruned each time, and an SVM classifier is trained on the remaining bands. After all the bands are pruned out or the minimum number of bands is reached, the band combination with the highest classification accuracy is chosen as the final result. In this paper, similar to Li et al. (2011), the pruning process is stopped when there are three bands left in each cluster.

2.2 *Proposed AdaBoostSVM System*

The main idea of our proposed AdaBoost-SVM method is similar to that introduced in Li et al. (2008). Nevertheless, there are major differences between these two methods. As mentioned before, the component classifiers in AdaBoost must satisfy the classification accuracy/diversity balance to achieve an increased performance. In Li et al. (2008) the whole

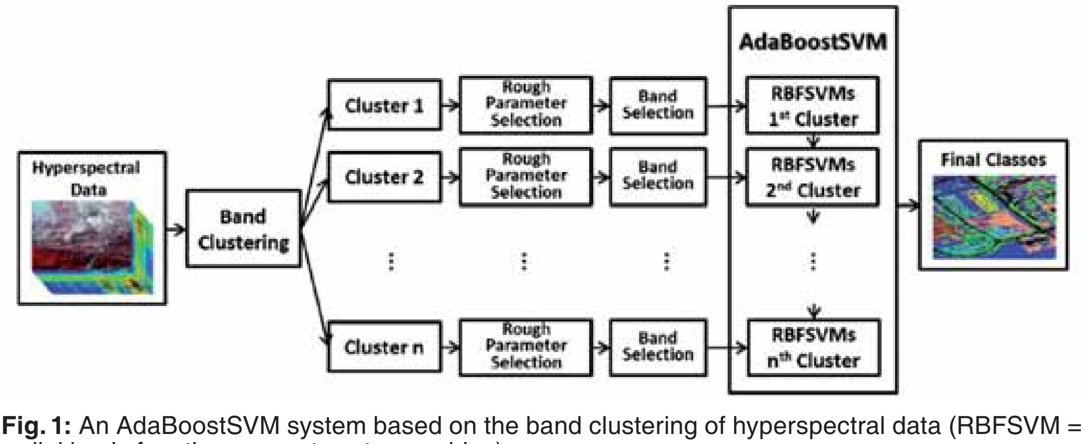

Fig. 1: An AdaBoostSVM system based on the band clustering of hyperspectral data (RBFSVM = radial basis function support vector machine).

dataset is used in different boosting iterations and the required diversity for component classifiers is satisfied by changing the RBF kernel parameter σ in different iterations. However, in our proposed AdaBoostSVM method, the component classifiers are diversified by using different feature subsets in each of the iterations and unlike LI et al. (2008), only the optimum values of kernel parameters for each band cluster are used to make the final decision.

The proposed AdaBoost algorithm takes a training set (X) with labels (Y) in the form $\{(x_1,y_1),\ldots,(x_N,y_N)\}$ and a distribution or a set of weights over the training set w_i^1 as inputs. The weights are initially set to 1. Then, in each boosting iteration the rough values of the SVM kernel parameters for one of these clusters are first calculated using a 5-fold cross-validation technique. The search range for C is (1 – 200) and for σ is (0 – 20). This cross validation and training of the SVMs are performed using LibSVM (CHEN & LIN 2011). To handle multiclass problems, a one-against-one strategy is preferred to a one-against-all strategy based on comparisons performed in HSU & LIN (2002). Next, the most informative set of bands in that cluster are defined using the ABB technique.

Because the performance of RBF-SVMs strongly depends on σ (LI et al. 2008, VALENTINI & DIETTERICH 2004), using a rough value of C, the exact value of σ is determined in an iterative manner for each SVM applied to one subset of bands. This approach is much faster than finding the best values of both parameters at the same time while using techniques as cross-validations. To do this, for each cluster j, the kernel parameter σ is decreased slightly, and an RBF-SVM is applied on the training set. The training error ($\varepsilon^j_{\sigma_t}$) for each of these classifiers ($h^j_{\sigma_t}$) is calculated (4).

$$h^j_{\sigma_t}: \varepsilon^j_{\sigma_t} = \sum_{i=1}^{N} w_i^j \left[h^j_{\sigma_t}(x_i) \neq y_i \right] \tag{4}$$

The classifier with the lowest training error is selected as the component classifier for this cluster (5).

$$h_j: \varepsilon_j = \min(\varepsilon^j_{\sigma_t}) \tag{5}$$

If this training error is less than 0.5, the weights of incorrectly classified training examples for this cluster will then be increased such that the base classifier is forced to focus on the hard examples in the training set of the next band clusters (6).

$$w_i^{j+1} = \frac{w_i^j \exp(-\alpha_j \left[h_j(x_i) \neq y_i \right])}{C_j} \tag{6}$$

where C_j is a normalization constant: $C_j = \sum_{i=1}^{N} w_i^{j+1} = 1$. The weight of the component classifiers are then calculated by (7).

$$\alpha_j = \frac{1}{2} \ln\left(\frac{1-\varepsilon_j}{\varepsilon_j} \right) \tag{7}$$

And finally, after all boosting iterations, the classifiers $h_1,\ldots,h_T$ are combined with weighted voting (8).

$$f(x) = \max\left(\sum_{j=1}^{\text{no. of clusters (T)}} \alpha_j h_j(x) \right) \tag{8}$$

As mentioned earlier, there are many modifications to generalise AdaBoost algorithm to support multiclass problems. These methods are evaluated in this research work and did not observe a meaningful difference in their performance. Thus, in the rest of this paper, AdaBoost.M2 has been used since it is faster and easier to implement.

3 Experimental Results

In this section, we evaluate our proposed methods by two real hyperspectral datasets. Our experimental analysis is organized into two main experiments. In the first experiment, the effectiveness of our algorithm is evaluated in comparison to some other classification strategies applied to hyperspectral data. The second experiment aims at analyzing the effect of training sample size on the performance of the utilised methods.

3.1 *Datasets*

To evaluate the potential of the proposed methods, two hyperspectral datasets which are acquired by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor and the ROSIS-3 (Reflective Optics System Imaging Spectrometer) sensor are used. The first dataset of AVIRIS is known for the complexity of the conveyed classification problem which covers an area of mixed agriculture and forest landscape in the Indian Pine area and was collected in June 1992 (Watanachaturaporn & Arora 2004). The size of this image data is 145×145 pixels and the scene comprises 220 spectral channels with the nominal spectral resolution of 10 nm, a moderate spatial resolution of 20 m by pixel, and 16bit radiometric resolution. After an initial screening, several spectral bands were removed from the dataset due to noise and water absorption phenomena, leaving a total of 200 channels to be used in the experiments (Fig. 2a, b). The second dataset was acquired from Pavia University using the ROSIS-3 sensor during a flight campaign over Pavia, northern Italy. For the ROSIS data, the number of spectral bands is 115. After removing some bands due to noise, 103 bands remain. This dataset exhibits 610×340 pixels with 1.3 metre per pixel spatial resolution (Fig. 2c, d). For AVIRIS and ROSIS datasets, field-surveyed maps consist of 16 and 9 classes respectively and one unclassified class.

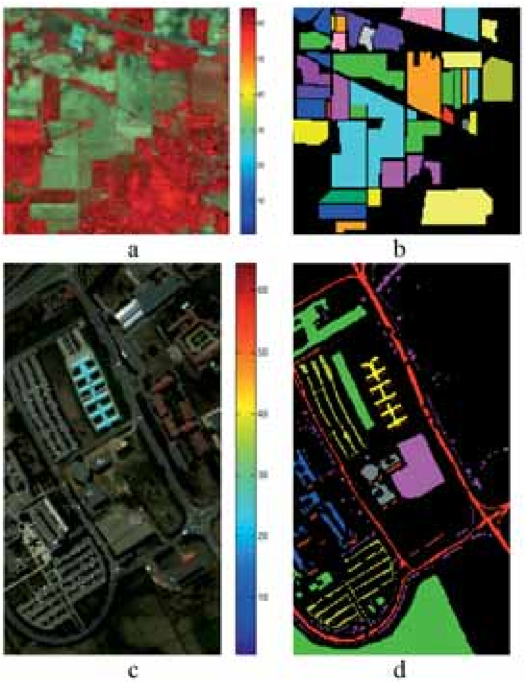

Fig. 2: Hyperspectral datasets. (a) Pseudo-colour AVIRIS image, (b) AVIRIS Indian Pine ground truth, (c) Pseudo-colour ROSIS Pavia image, (d) ROSIS Pavia ground truth.

3.2 *Experimental Setup*

Before the classification, some pre-processing should be performed on these data clusters. Data sources should be scaled to the range (0 – 1). This eases the tuning of the SVM kernel parameters (Chen & Lin 2011).

In all the experiments, approximately 10% of the randomly selected samples in each class are considered as training set and the rest are used for evaluation. To decrease variations in the classification process, all of the experiments are repeated 10 times on randomly selected samples, and the results are averaged. For the MV-SVM, the SVM-All, and the SVM-Sel approaches, the kernel parameters are defined using grid search 5-fold cross validations. To evaluate the performance of different methods, the Kappa coefficient and the overall accuracy are usually used (Congalton et al. 1991). Moreover, producer's accuracy is utilised to measure the accuracy of each class.

In addition to overall and single class accuracies, the computational time spent on each method is presented in seconds. It is based on a Pentium IV machine with a 2.20 GHz Dual Core Processor and 4 GB of memory.

3.3 *Output Results on First Dataset*

The first step is to perform band clustering based of the MI between adjacent bands. Depending on the local minima values, various decompositions can be considered Fig. 3.

We have evaluated different decompositions and found that the one with seven clusters outperforms others in terms of classification performance. Details of these clusters and overall accuracies of different component classifiers in our AdaBoostSVM system are given in Tab. 1.

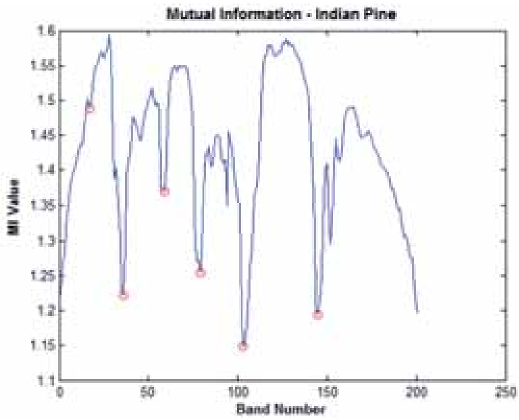

Fig. 3: Band clustering of the Indian Pine dataset based on MI calculations. Red circles define the bands for each cluster.

The results of the proposed AdaBoostSVM and the ensemble classification system utilising majority voting (the MV-SVM) algorithms as well as the SVM-All and the SVM-Sel are compared in Tab. 2 in terms of the accuracies of each of the classes, the overall accuracy and the Kappa parameter. As can be seen, although in some classes such as classes #4 and #15 the accuracies of the AdaBoost-SVM are less than those of the SVM-Sel and MV-SVM, in almost half of the classes, the improvement of accuracies by the AdaBoost-SVM is significantly higher. The classification maps obtained using these four methods are illustrated in Fig. 4. It is obvious that the map achieved by the AdaBoostSVM appears more

Tab. 1: Final band clustering results on the Indian Pine dataset.

Clusters	1	2	3	4	5	6	7
Bands	1–15	16–36	37–59	60–79	80–102	103–146	147–200
Accuracy (%)	56.19	64.13	63.78	69.56	69.01	66.78	71.22

Tab. 2: Results of applying the proposed method on Indian Pine dataset (MV = majority voting, Sel = applied to selected bands (%)).

Class	Colour	Land Cover Class	Samples	AdaBoost SVM	MV SVM	SVM-Sel (133 bands)	SVM-All (202 bands)
1		Alfalfa	54	95.45	100.00	90.63	53.06
2		Corn-no till	1 434	88.84	80.03	81.87	80.32
3		Corn-minimum till	834	85.91	87.26	78.16	65.38
4		Corn	234	71.12	86.92	85.37	53.55
5		Grass/pasture	497	95.47	94.78	85.68	90.82
6		Grass/trees	747	96.65	89.09	94.20	98.21
7		Grass/pasture-mowed	26	100.00	100.00	93.75	47.82
8		Hay-windrowed	489	98.46	98.46	96.59	98.35
9		Oats	20	100.00	100.00	91.67	55.56
10		Soybeans-no till	968	90.82	91.42	81.97	88.06
11		Soybeans-minimum till	2 468	88.75	67.50	88.83	87.08
12		Soybeans-clean till	614	84.52	87.41	80.29	78.84
13		Wheat	212	98.82	99.40	97.38	99.48
14		Woods	1 294	97.00	90.07	90.13	94.76
15		Building-grass-trees-drives	380	74.44	84.70	81.57	49.42
16		Stone-steel towers	95	97.77	89.80	92.98	74.12
Kappa				0.89	0.78	0.85	0.82
OA				91.03	81.52	86.53	84.15
Time				345 s	296 s	84 s	208 s

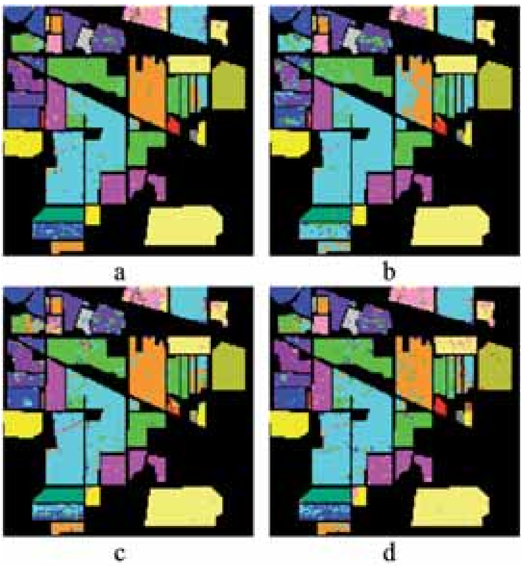

Fig. 4: Classification maps of the AVIRIS Indian Pine dataset. (a) AdaBoostSVM, (b) MV-SVM, (c) SVM-Sel, (d) SVM-All.

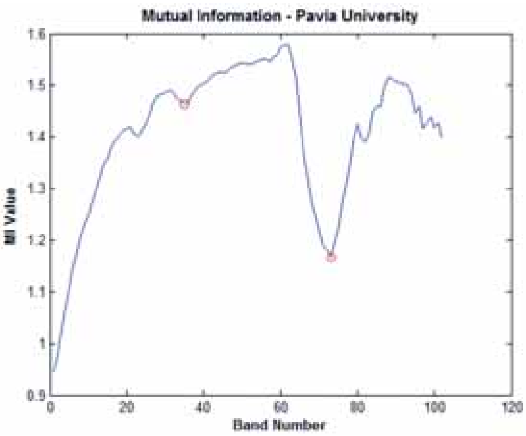

Fig. 5: Band clustering of the Pavia dataset based on the MI.

Tab. 3: Final band clustering results on Pavia dataset.

Clusters	1	2	3
Bands	1 – 38	39 – 73	74 – 103
Accuracy (%)	82.39	88.99	91.34

homogeneous than the others, which is indicative of superior classification performance.

In the last row of Tab. 2, the computational time spent on each method to fulfil is also expressed in seconds. The AdaBoostSVM as it is performed in a couple of iterations takes a longer time, 345 seconds, but not much more than the other methods.

3.4 *Output Results on Second Dataset*

The same experiments are repeated on the second dataset (the ROSIS Pavia University). These 103-band data are divided into three clusters using the MI method (Fig. 5).

The number of spectral bands contained in each of the clusters is given in Tab. 3. The four classification methods are later applied on these three data sources.

The classification results in terms of each class and overall accuracies together with the computational times are given in Tab. 4.

Similar to the other dataset, the results of the AdaBoostSVM on the Pavia University dataset exhibit greater accuracies than those for the MV-SVM, the SVM-Sel and the SVM-All. In most of the nine classes, the AdaBoost-SVM outperforms the other methods. This result is particularly evident for classes exhibiting lower accuracies. But the overall differences are not as high as in the first experiment. The classification maps of these methods are illustrated in Fig. 6.

4 Conclusions and Discussion

The classification of hyperspectral data was addressed and evaluated using ensemble classifier systems based on AdaBoost using SVMs as base classifiers. The proposed method relies on combining SVM classifiers that are trained on different data clusters defined according to the correlation matrix of the spectral bands. This algorithm has been compared with an ensemble classification system using majority voting to combine SVMs applied on band clusters (the MV-SVM), a single SVM applied on selected bands from the whole datasets (the SVM-Sel) and a single SVM applied on all of the spectral bands (the SVM-All). All the techniques have been evaluated using two standard hyperspectral datasets. In both datasets, the AdaBoostSVM outperforms the other three methods in terms of accuracy. For classification of the Indian Pine data, the differences between the overall accuracy of the AdaBoostSVM and those of the other methods vary between 4.50% and 9.51%, being greater for the Pavia University dataset. As observed, the overall and single class accuracies achieved in the Indian Pine dataset are lower than that of the Pavia data, which may be attributable to two reasons.

First, the size of the Indian Pine dataset is small compared to the other dataset; therefore, fewer training samples are available for some classes, which decrease the classification performance. This finding has been observed and reported by some other research-

Tab. 4: Results of applying the proposed method on the Pavia University dataset (%).

Class	Colour	Land Cover Class	Samples	AdaBoost SVM	MV SVM	SVM-Sel (63 bands)	SVM-All (103 bands)
1		Asphalt	6 631	94.14	89.86	93.64	94.27
2		Meadows	18 649	98.27	94.01	97.21	95.00
3		Gravel	2 099	80.78	88.35	85.00	81.10
4		Trees	3 064	95.40	95.58	95.11	95.09
5		Painted metal sheets	1 345	99.26	100	98.88	100
6		Bare Soil	5 029	90.57	94.65	90.34	92.87
7		Bitumen	1 330	87.47	89.37	89.47	73.43
8		Self-Blocking Bricks	3 682	91.31	82.08	89.31	84.33
9		Shadows	947	99.76	100	99.74	99.76
Kappa				0.93	0.90	0.92	0.92
OA				94.79	92.35	93.69	92.53
Time				445 s	370 s	117 s	183 s

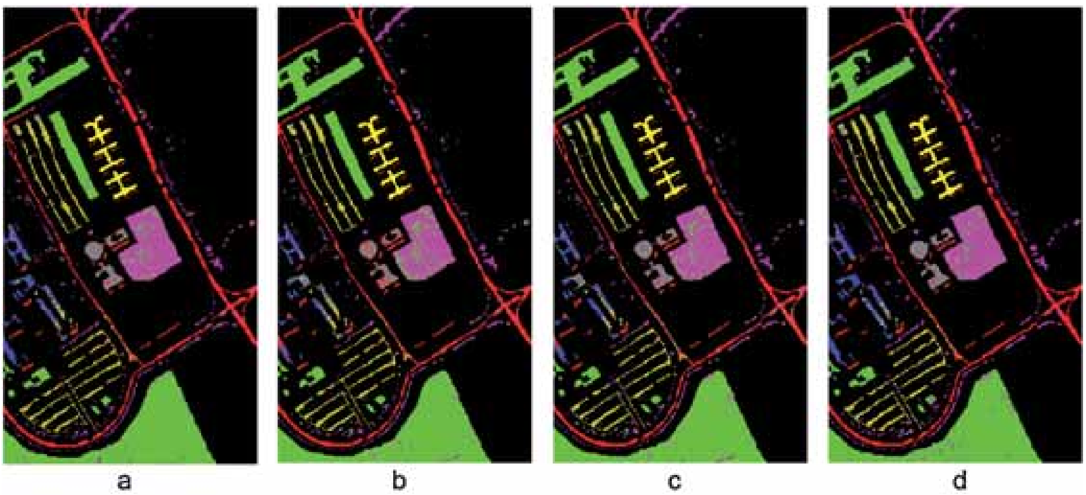

Fig. 6: Classification maps of the ROSIS Pavia University dataset. (a) AdaBoostSVM, (b) MV-SVM, (c) SVM-Sel, (d) SVM-All.

ers such as CEAMANOS et al. (2010) too. Second, classification of the Indian Pine dataset is inherently difficult because this area is covered with mixed agriculture and forestry landscapes, and the data were collected at a time when most of the crops in the test site had not reached maximum ground cover (CEAMANOS et al. 2010). Some classes are homogeneous areas that can be clearly defined; therefore, the accuracies achieved for these classes are high using all methods. By contrast, classes such as #2 (Corn-no till), #3 (Corn-minimum till), #4 (Corn), #11 (Soybeans-minimum till), #12 (Soybeans-clean till), and #15 (Building-grass-trees-drives) can be misclassified in certain parts because they have subclasses (no till, minimum till, clean till) or consist of different elements (Building-grass-trees-drives).

Similar results for the Indian Pine classes have also been reported by CAMPS-VALLS & BRUZZONE (2005), LI et al. (2013), and TARABALKA et al. (2010). By contrast, because the Pavia University dataset is less complex and for all classes, sufficient training samples are available, the classification performance for this dataset is also high.

Among our datasets, the classification of the Pavia University is more time consuming than that of Indian Pine, which is due to the dimension of data; the computing time increases significantly with a greater number of bands and training samples. Nevertheless, the time is important only when the classification must be performed in a short period of time, such as in emergency situations or online applications, or repeated frequently. For the application in this study, the computational time does not seem to be significant.

To conclude, the proposed method seems to be applicable and beneficial for the classification of hyperspectral data since it results in higher classification accuracies and relies on common and extensively-used concepts of SVM and AdaBoost, and can be simply integrated in existing remote sensing applications. In future research, we shall evaluate the performance of other band clustering and band selection methods and will utilise spatial features and other data sources in our classification process. Furthermore, we will analyse the applicability of the proposed method to handle high-dimensional datasets like multi-temporal time series. The optimisation of our algorithm to reduce the computational burden and classification problems will also be investigated.

Acknowledgements

The authors would like to thank Prof. D. LANDGREBE and Prof. P. GAMBA for making the hyperspectral datasets available, and C.C. CHEN and C.J. LIN for the LibSVM codes. Furthermore, we would also like to thank the associate editor who handled our paper and the two anonymous reviewers for their constructive comments and suggestions.

References

BENEDIKTSSON, J.A. & KANELLOPOULOS, I., 1999: Classification of Multisource and Hyperspectral Data Based On Decision Fusion. – IEEE Transactions on Geoscience and Remote Sensing **37** (3):1367–1377.

BIGDELI, B., SAMADZADEGAN, F. & REINARTZ, P., 2013: Band Grouping Versus Band Clustering In SVM Ensemble Classification Of Hyperspectral Imagery. – Photogrammetric Engineering and Remote Sensing **79** (6): 523–534.

BRAUN, A.C., WEIDNER, U. & HINZ, S., 2010: Support Vector Machines for Vegetation Classification A Revision. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2010** (4): 273–282.

BRAUN, A.C., WEIDNER, U. & HINZ, S., 2012: Kernel Composition with the One-Against-One Cascade for Integrating External Knowledge into SVM Classification. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2012** (4): 371–384.

BREIMAN, L., 1996: Bagging Predictors. – Machine Learning **26** (2): 123–140.

BRIEM, G.J., BENEDIKTSSON, J.A. & SVEINSSON, J.R. 2002: Multiple Classifiers Applied to Multisource Remote Sensing Data. – IEEE Transactions on Geoscience and Remote Sensing **40** (10): 2291–2299.

CAMPS-VALLS, G. & BRUZZONE, L., 2005: Kernel-based Methods For Hyperspectral Image Classification. – IEEE Transaction on Geoscience and Remote Sensing **43** (4):1351–1362.

CEAMANOS, X.R., WASKE, B., BENEDIKTSSON, J.A., CHANUSSOT, J., FAUVEL, M. & SVEINSSON, J., 2010: A Classifier Ensemble Based on Fusion of Support Vector Machines for Classifying Hyperspectral Data. – International Journal of Image and Data Fusion **1** (4): 293–307.

Chan, J.C.W. & Paelinckx, D., 2008: Evaluation of Random Forest and Adaboost Tree-based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery. – Remote Sensing of Environment **112** (6): 2999–3011.

Chen, C.C. & Lin, C.J., 2011: LIBSVM: A Library for Support Vector Machines. – ACM Transactions on Intelligent Systems and Technology (TIST) **2** (3).

Chi, M., Feng, R. & Bruzzone, L., 2008: Classification of Hyperspectral Remote-sensing Data with Primal SVM for Small-sized Training Dataset Problem. – Advances in Space Research **41** (11): 1793–1799.

Congalton, R.G., Fenstermaker, L.K., Jensen, J.R., McGwire, K. & Tinney, L.R., 1991: Remote Sensing And Geographic Information System Data Integration: Error Sources And Research Issues. – Photogrammetric Engineering and Remote Sensing **57** (6): 677–687.

Freund, Y. & Schapire, R., 1999: A Short Introduction to Boosting. – Journal of the Japanese Society for Artificial Intelligence **14** (54): 771–780.

Frick, A., Steffenhagen, P., Zerbe, S., Timmermann, T. & Schulz, K., 2011: Monitoring of the Vegetation Composition in Rewetted Peatland with Iterative Decision Tree Classification of Satellite Imagery. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2011** (3): 109–122.

Guo, B., Gunn, S.R., Damper, R.I. & Nelson, J.D.B., 2006: Band Selection For Hyperspectral Image Classification Using Mutual Information. – IEEE Geoscience and Remote Sensing Letters **3** (4): 522–526.

Guyon, I. & Elissee, A., 2003: An introduction to variable and feature selection. – Journal of Machine Learning Resources **3** (2003): 1157–1182.

Hsu, C.W. & Lin, C.J., 2002: A Comparison of Methods for Multi-class Support Vector Machines. – IEEE Transactions on Neural Networks **13** (2002): 415–425.

Hughes, G.F., 1968: On the Mean Accuracy of Statistical Pattern Recognition. – IEEE Transactions on Information Theory (IT) **14:** 55–63.

Kuncheva, L.I. & Whitaker, C.J., 2003: Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. – Machine Learning **51** (2): 181–207.

Landgrebe, D.A., 2003: Signal Theory Methods in Multispectral Remote Sensing. – Wiley-Inter-Science, New York, NY, USA.

Li, J., Bioucas-Dias, J.M. & Plaza, A., 2013: Spectral-Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning. – IEEE Transactions on Geoscience and Remote Sensing **51** (2): 844–856.

Li, S., Wu, H., Wan, D. & Zhu, J., 2011: An Effective Feature Selection Method for Hyperspectral Image Classification Based On Genetic Algorithm and Support Vector Machine. – Knowledge-Based Systems **24** (1):40–48.

Li, X., Wang, L. & Sung, E., 2008: AdaBoost with SVM-based Component Classifiers. – Engineering Applications of Artificial Intelligence **21** (5).

Liao, Y. & Moody, J., 1999: Constructing Heterogeneous Committees Using Input Feature Grouping. – Advances in Neural Information Processing Systems (NIPS) Conference: 921–927, MIT Press, Cambridge, MA, USA.

Martinez-Uso, A., Pla, F., Sotoca, J.M. & Garcia-Sevilla, P., 2007: Clustering Based Multispectral Band Selection Using Mutual Information. – 18th International Conference on Pattern Recognition (ICPR 2006): 760–763, Hong Kong, China.

Melgani, F. & Bruzzone, L., 2004: Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. – IEEE Transaction on Geoscience and Remote Sensing **42** (8): 1778–1790.

Nakariyakul, S. & Casasent, D.P., 2007: Adaptive Branch And Bound Algorithm For Selecting Optimal Features. – Pattern Recognition Letters **28** (12): 1415–1427.

Opelt, A., Fussenegger, M., Pinz, A. & Auer, P., 2006: Generic Object Recognition with Boosting. – IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (3): 416–431.

Pal, M. & Mather, P.M., 2006: Some Issues in the Classification of DAIS Hyperspectral Data. – International Journal of Remote Sensing **27** (14): 2895–2916.

Prasad, S. & Bruce, L.M., 2008: Decision Fusion with Confidence-Based Weight Assignment for Hyperspectral Target Recognition. – IEEE Transactions on Geoscience and Remote Sensing **46** (5): 1448–1456.

Richards, J., 2013: Remote Sensing Digital Image Analysis. – 5th ed., Springer-Verlag, Berlin.

Samadzadegan, F., Hasani, H. & Schenk, T., 2012: Simultaneous Feature Selection and SVM Parameter Determination in Classification of Hyperspectral Imagery Using Ant Colony Optimization. – Canadian Journal of Remote Sensing **38** (2): 139–156.

Schapire, R.E. & Singer, Y., 1998: Improved Boosting Algorithms Using Confidence-rated Predictions. – Eleventh annual conference on Computational learning theory, COLT' 98, to appear in Machine Learning: 80–91, New York, NY, USA.

Schölkopf, B. & Smola, A.J., 2001: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. – MIT Press, Cambridge, MA, USA.

Tarabalka, Y., Fauvel, M., Chanussot, J. & Benediktsson, J., 2010: SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images. – IEEE Geoscience and Remote Sensing Letters **7** (4): 736–740.

Valentini, G. & Dietterich, T.G., 2004: Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods. – The Journal of Machine Learning Research **5** (12): 725–775.

Vapnik, V.N. & Vapnik, V., 1998: Statistical Learning Theory. – Wiley, New York, NY, USA.

Villa, A., Benediktsson, J.A., Chanussot, J. & Jutten, C., 2011: Hyperspectral Image Classification With Independent Component Discriminant Analysis. – IEEE Transactions on Geoscience and Remote Sensing **49** (12): 4865–4876.

Viola, P. & Jones, M., 2001: Fast and Robust Classification Using Asymmetric AdaBoost and a Detector Cascade. – Advances in Neural Information Processing System **14**.

Wang, S.J., Mathew, A., Chen, Y., Xi, L.F., Ma, L. & Lee, J., 2009: Empirical Analysis of Support Vector Machine Ensemble Classifiers. – Expert Systems Applications **36** (3): 6466–6476.

Waske, B., Linden, S.V.D., Benediktsson, J.A., Rabe, A. & Hostert, P., 2010: Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data. – IEEE Transaction on Geosciences and Remote Sensing **48** (10): 2880–2889.

Watanachaturaporn, P. & Arora, M.K., 2004: Support Vector Machines for Classification of Multi- and Hyperspectral Data. – Varshney, P.K. & Arora, M.K. (eds.): Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data: 237–255, Springer-Verlag, Berlin.

Wickramaratna, J., Holden, S.B. & Buxton, B.F., 2001: Performance Degradation in Boosting. – Multiple Classifier Systems: Second International Workshop, MCS 2001: 11–21, Cambridge, UK.

Xie, H., Heipke, C., Lohmann, P., Soergel, U., Tong, X. & Shi, W., 2011: A New Binary Encoding Algorithm for the Simultaneous Region-based Classification of Hyperspectral Data and Digital Surface Models. – PFG – Photogrammetrie, Fernerkundung, Geoinformation **2011** (1): 17–33.

Zhao, Y., Zhang, L. & Kong, S.G., 2011: Band Subset Based Clustering and Fusion for Hyperspectral Imagery Classification. – IEEE Transactions on Geoscience and Remote Sensing **49** (2): 747–756.

Addresses of the Authors:

Pouria Ramzei & Prof. Dr. Farhad Samadzadegan, University of Tehran, Department of Photogrammetry and Remote Sensing, Faculty of Engineering, North Amirabad, Tehran, Iran, Fax +98-21-88008841, e-mail: {pramzi}{samadz}@ut.ac.ir

Prof. Dr.-Ing. Peter Reinartz, German Aerospace Centre (DLR), Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, D-82234 Oberpfaffenhofen, Germany, Tel. +49-8153-28-2757, e-mail: peter.reinartz@dlr.de

Manuskript eingereicht: August 2013
Angenommen: November 2013