HAL
open science

A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data

Xavier Ceamanos, Bjorn Waske, Jon Atli Benediktsson, Jocelyn Chanussot, Mathieu Fauvel, Johannes Sveinsson

To cite this version:

Xavier Ceamanos, Bjorn Waske, Jon Atli Benediktsson, Jocelyn Chanussot, Mathieu Fauvel, et al.. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. International Journal of Image and Data Fusion, 2010, 1 (4), pp.293-307. 10.1080/19479832.2010.485935 . hal-00578897

HAL Id: hal-00578897
https://hal.science/hal-00578897

Submitted on 13 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Journal of Image and Data Fusion
Vol. 1, No. 4, December 2010, 293–307

A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data

Xavier Ceamanos[a]*, Björn Waske[b], Jón Atli Benediktsson[c], Jocelyn Chanussot[d], Mathieu Fauvel[e] and Johannes R. Sveinsson[c]

[a]*Laboratoire de Planétologie de Grenoble, CNRS-UJF, Grenoble, France;* [b]*Institute for Geodesy and Geoinformation, University of Bonn, Bonn, Germany;* [c]*Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland;* [d]*GIPSA-lab, Grenoble INP, Grenoble, France;* [e]*MISTIS team, INRIA Grenoble, Saint Ismier, France*

(*Received 4 January 2010; final version received 2 April 2010*)

Classification of hyperspectral data using a classifier ensemble that is based on support vector machines (SVMs) are addressed. First, the hyperspectral data set is decomposed into a few data sources according to the similarity of the spectral bands. Then, each source is processed separately by performing classification based on SVM. Finally, all outputs are used as input for final decision fusion performed by an additional SVM classifier. Results of the experiments underline how the proposed SVM fusion ensemble outperforms a standard SVM classifier in terms of overall and class accuracies, the improvement being irrespective of the size of the training sample set. The definition of the data sources resulting from the original data set is also studied.

Keywords: classification of remote sensing data; support vector machines; hyperspectral data; decision fusion

Downloaded At: 14:16 22 March 2011

1. Introduction

Hyperspectral data cover a wide spectral range from the visible to the short-wave infrared, resulting in hundreds of data channels. Thanks to this volume of information, it is feasible to deal with applications that require a precise discrimination in the spectral domain. Nevertheless, a large number of features can become a curse in terms of classification accuracy if enough training samples are not available, i.e. the Hughes phenomenon (Hughes 1968). In this context, the use of conventional statistical methods may not be adequate for classifying high-dimensional data and therefore more sophisticated classifiers need to be considered.

In this article, classification of hyperspectral remote sensing data with a classifier ensemble based on support vector machines (SVMs) are investigated. SVMs have been used for the classification of hyperspectral data. Camps-Valls and Bruzzone (2005) demonstrated that SVMs perform equal or better than other classifiers in terms of accuracy. Furthermore, SVMs have shown to be relatively unaffected by the Hughes phenomenon (Pal and Mather 2004). Multiple classifier systems (MCSs), or classifier ensembles, are another machine learning concept, which have been applied to remote sensing data sets latterly (Benediktsson *et al.* 2007). By combining different classification

*Corresponding author. Email: xavier.ceamanos@obs.ujf-grenoble.fr

ISSN 1947–9832 print/ISSN 1947–9824 online
© 2010 Taylor & Francis
DOI: 10.1080/19479832.2010.485935
http://www.informaworld.com

algorithms or variants of the same classifier, MCS can improve classification accuracy in comparison to a single classifier. For example, Ham *et al.* (2005) successfully applied ensemble strategies to hyperspectral data.

Both machine learning concepts can be used together in a reasonable manner. In Waske and Benediktsson (2007), SVMs are used for classifying multisensor imagery using a separately trained SVM classifier for each source followed by a final SVM-based decision fusion. The proposed technique in Waske and Benediktsson (2007) outperforms other classifier ensembles and a conventional SVM, which is applied on the whole data set, in terms of accuracy.

Considering the previously obtained results, it may be desirable to employ this concept for classifying hyperspectral data. In order to define the multiple sources for every classifier, the original hyperspectral data can be split into several smaller data sources, based on the similarity between data channels as it was done in Benediktsson and Kanellopoulos (1999). Afterwards, each individual source undergoes a preliminary classification using a separately trained SVM classifier. The outputs of the individual classifiers are then combined by decision fusion, which is performed by an additional SVM. In order to investigate the capabilities of the proposed method, classification results on two hyperspectral images are compared to those achieved by an ordinary SVM classifier, using the whole data set. Moreover, different decompositions of the hyperspectral data set are probed to study the criteria for an optimal definition of the data sources.

This article is organised as follows. SVM and MCS are reviewed in Sections 2 and 3, respectively. Then, the SVM fusion approach is presented in Section 4, along with an introduction to the data sets in Section 5. Experimental results are given in Section 6, and conclusions are finally drawn in Section 7.

Downloaded At: 14:16 22 March 2011

2. Support vector machines

SVM refers to a set of supervised learning methods aiming at discriminating two classes by fitting an optimal separating hyperplane to the training data within a multidimensional feature space (Vapnik 1998).

Although SVMs were originally proposed to solve linear classification problems, they can be generalised to non-linear decision functions by using the so-called *kernel trick* (Scholkopf and Smola 2002). Using kernels, the data may be mapped onto a higher dimensional space F, thus allowing to work implicitly in F while all the computation is done in the input space. A widely used kernel in remote sensing is the Gaussian radial basis function (RBF) (Scholkopf and Smola 2002).

Nonetheless, a disadvantage of SVM classifiers is related to their sensitivity to the choice of the kernel and regularisation parameters. The latter is classically overcome by cross-validation techniques using training data.

Despite the fact that SVMs deal with binary classification problems, several strategies have been developed to overcome multiclass problems as in remote sensing applications (Roli *et al.* 2001). In general, a c-class problem is split into multiple binary problems to finally combine the individual binary SVM classifiers in an ensemble. In the *one-against-one* (OAO) approach, a set of $c(c-1)/2$ individual SVM is trained, one for each pairwise class discrimination. The output of the classifier for each pairwise classes results in an image, that is hereafter referred as *rule image* (Waske and Benediktsson 2007), containing

the distance of each pixel to the decision boundary of the SVM. A score function is computed for each class by adding all positive and negative votes for a specific class. The final class membership for an unknown sample is eventually determined by the simple majority. Although the OAO strategy requires more SVM classifiers than other techniques, it is more appropriate for remote sensing applications because the classification problem is partitioned into several simpler sub-problems. An exhaustive introduction to linear, non-linear and multiclass SVM can be found in Scholkopf and Smola (2002).

One of the main advantages of SVM regarding remote sensing classification is that the approach only considers samples that are close to the class boundaries, the so-called *support vectors*. Using this, SVM works satisfactorily even with small training sets, when classifying high-dimensional hyperspectral imagery (Pal and Mather 2006). SVMs were applied to hyperspectral airborne visible/infrared imaging spectrometer (AVIRIS) data by Melgani and Bruzzone (2004). In that study, results underlined that SVMs are an adequate and effective alternative to traditional approaches for the classification of hyperspectral remote sensing data regardless of the adopted multiclass strategy. The good performances of SVMs are further demonstrated in Dalponte *et al*. (2008) and Koetz *et al*. (2008). In those studies, hyperspectral imagery and light detection and ranging (LIDAR) data are successfully combined using SVM classifiers.

Downloaded At: 14:16 22 March 2011

3. Multiple classifier systems

MCS aim a more accurate classification decision at the expense of an increased complexity by combining classifiers. Many researchers have investigated the strategy of combining the predictions of several classifiers to supply a single classification result (Roli *et al*. 2001, Banfield *et al*. 2007, Benediktsson *et al*. 2007). The resulting classifier (hereafter referred to as an *ensemble*) is generally more accurate than any of the individual classifiers that make up the ensemble. Both theoretical and empirical research has demonstrated that a good ensemble is one where the individual classifiers in the ensemble are accurate as well as they make mistakes on different parts of the input space (Krogh and Vedelsby 1995, Kittler 1998). Two popular methods for creating accurate ensembles are Bagging (Breiman 1996) and Boosting (Schapire 1990).

Figure 1 illustrates the basic framework for a classifier ensemble. First, each classifier is trained using the same training set. Then, for each example, the predicted output of each classifier is combined to produce the output of the ensemble.

Several studies have demonstrated that simple averaging (i.e. majority vote) of the classifier predictions is an effective scheme (Hansen and Salamon 1990, Kittler 1998). With this schema, a given pixel is classified to the class that was chosen by the majority of classifiers. Combining the output of several classifiers is useful only if there is disagreement among them (i.e. their variations are not correlated). Obviously, combining several identical classifiers produces no gain. Hansen and Salamon (1990) proved that if the average error rate is less than 50% and the component classifiers in the ensemble are independent in the production of their errors, the expected error for that example can be reduced to zero as the number of classifiers combined goes to infinity. However, such assumptions rarely hold in practice. Krogh and Vedelsby (1995) later proved that an ideal ensemble consists of highly correct classifiers that disagree as much as possible. Opitz and Shavlik (1996) empirically verified that such ensembles generalise well.

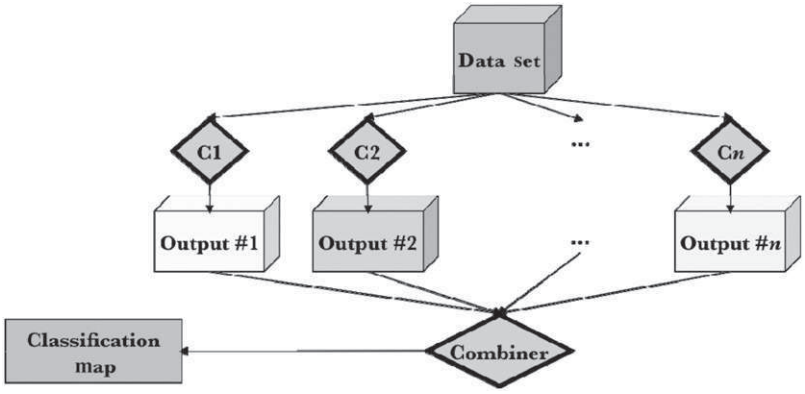

Figure 1. MCS schema.

As a result, methods for creating ensembles centre around producing classifiers that disagree on their predictions and therefore they are as independent as possible. Generally, these methods focus on either different types of classifiers or variations of the same classifier are used in the hope that the resulting classifiers will produce different predictions.

In context of remote sensing, ensemble techniques were used in particular for multisource data sets and hyperdimensional imagery. Briem *et al*. (2002) applied various classifiers on multisource data containing multispectral imagery, synthetic-aperture radar (SAR) data and topographical information. The results clearly demonstrated that ensemble techniques outperform single classifiers in terms of overall accuracy. In Ham *et al*. (2005), a decision-tree-based ensemble method (i.e. random forests) was used for classifying hyperspectral imagery using a limited training sample set. In Waske and Benediktsson (2007), multispectral imagery and multitemporal SAR data were combined by an SVM ensemble. Two SVM classifiers were applied separately on the two data sources. The preliminary outputs were combined by another SVM. The presented approach outperformed all other techniques, including a standard SVM, a boosted decision tree and a majority vote. The concept introduced in Waske and Benediktsson (2007) was modified by a multilevel component, fusing different scales of segmentation, using SVM and a classifier ensemble based on decision trees (DT) (Waske and Van der Linden 2008). In contrast to classifier ensembles, which combine variants of the same classifier, other approaches are based on the combination of different algorithms. Benediktsson and Kanellopoulos (1999), for example, combined neural networks and statistical methods for classifying multisensor and hyperspectral data, respectively. The different data sources underwent a separate classification and the outputs of the individual classifiers were combined by decision fusion to create final classification map.

4. Methods

A SVM-based classifier ensemble for the classification of hyperspecral remote sensing data is introduced. The proposed classification algorithm aims at benefiting from the

Downloaded At: 14:16 22 March 2011

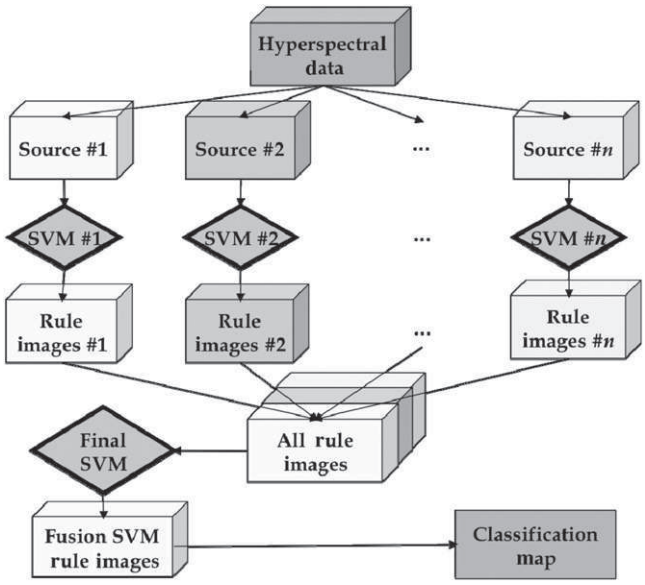

Figure 2. Schematic diagram of the SVM fusion method.

Downloaded At: 14:16 22 March 2011

capabilities of both SVM and MCS concepts. An overview of the classification approach is shown in Figure 2. First, the hyperspectral data set is split into n data sources, i.e. hyperspectral subsets, in a similar way to what was done in Benediktsson and Kanellopoulos (1999). Afterwards, a standard SVM classifier is separately applied to each data source. Finally, the rule images resulting from every SVM are fused (i.e. combined and classified) by an additional SVM in accordance to Waske and Benediktsson (2007). The same training and test sets are used throughout the classification process.

In contrast to Waske and Benediktsson (2007), a single hyperspectral source is now available instead of the multi-source data. Nonetheless, the proposed method aims at exploiting the spectral heterogeneity of a hyperspectral data set by splitting the original input into several similar subsets. As a matter of fact, the usual targets of interest in remote sensing classification problems usually present unique discriminative spectral features along the visible and short-wave infrared spectrum, thus introducing diversity in the hyperspectral data. Hence, it seems reasonable to define a series of subsets of data channels, which encompass the different homogeneous portions along the hyperspectral spectrum, according to the spectral bands similarity. In doing so, the individual SVM classifiers might be suitable for satisfactorily classifying their corresponding hyperspectral subspace. In fact, the kernel of each individual classifier is adjusted according to the corresponding data source properties. Finally, the MCS-inspired fusion step aims for a better classification accuracy by maximising the discriminative information of each subset, while minimising the individual misclassifications. Another reason to motivate the use of classifier ensembles is that individual classifiers may produce diversity if each of them is trained in different input space regions.

The absolute correlation is proposed as a measure of the spectral bands similitude as it was done by Benediktsson and Kanellopoulos (1999). The correlation matrix R is then

computed to identify potential data sources. The elements of R are defined by the absolute correlation r between bands x_i and x_j in an n-dimensional space:

$$r_{x_i x_j} = \left| \frac{n \sum_1^n x_i x_j - \sum_1^n x_i \sum_1^n x_j}{\sqrt{n \sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2} \sqrt{n \sum_1^n x_j^2 - \left(\sum_1^n x_j\right)^2}} \right|. \tag{1}$$

The aim of computing R is to identify ranges of strongly correlated adjacent hyperspectral channels manually.

5. Data sets used

Experiments are conducted on two hyperspectral AVIRIS data sets.

First of all, experiment 1 is carried out on data from Indiana, USA. The AVIRIS Indian Pine data were made available by D.A. Landgrebe at http://cobweb.ecn.purdue.edu/~biehl/MultiSpec. The scene corresponds to the Indian Pine test site in the northwest of Indiana and has 220 data channels and 145×145 pixels. The AVIRIS sensor operates in the visible and mid-infrared wavelength range (i.e. from 0.4 to 2.4 μm) and has 220 data channels coming from four spectrometers. About two-thirds of the scene is agriculture, while the remaining third contains forest or other natural perennial vegetation (see the information classes in Table 1). Data were collected in June 1992 when most of the crops in the test site had not reached their maximum ground cover. The classification of the crops is challenging since the spectral information comes from crop mixtures, soil variations and previous crop residues.

As shown in Table 1, the study site contains very heterogeneous classes in terms of their sizes. Hence, a stratified random sampling is performed to generate a training set. To investigate the performances of the SVM ensemble based on the number of training

Downloaded At: 14:16 22 March 2011

Table 1. Indian Pine land cover classes and available reference samples.

Class #	Land cover class	Labelled samples
1	Alfalfa	54
2	Corn-no till	1434
3	Corn-minimum till	834
4	Corn	234
5	Grass/pasture	497
6	Grass/trees	747
7	Grass/pasture-mowed	26
8	Hay-windrowed	489
9	Oats	20
10	Soybeans-no till	968
11	Soybeans-minimum till	2468
12	Soybeans-clean till	614
13	Wheat	212
14	Woods	1294
15	Building–grass–trees–drives	380
16	Stone–steel towers	95
Total		10,366

samples, three different training sample sets are generated with 1/16, 1/8 and 1/4 of all samples for each class. The remaining reference samples are used as independent test sets. To decrease variations in the classification process, all experiments are repeated 10 times on random selected training and testing sets.

The area used in experiment 2 is a 2048×614 pixel region surrounding the Hekla volcano in Iceland. Hekla is one of the most active volcanoes in Iceland. AVIRIS data from the area were collected on 17 June 1991, which was a cloud-free day. During the data collection, AVIRIS spectrometer no. 4 was not working properly and thus, once all data from that spectrometer were deleted, 157 data channels were left.

Twenty-four land cover classes were defined in the area, and 34,158 reference samples were selected. Training and test sets were defined by experts who collected the data. Approximately 50% of the reference samples were used for training, while the rest were used to test the classifiers (Table 2).

Due to the large number of training samples, some other cases were also considered. Similar to the method adopted in experiment 1, training sets containing 25, 50 and 100 samples per class are defined in order to test the SVM fusion when classifying using a small number of training samples. These three last experiments are repeated 10 times on randomly selected training and testing sets in order to decrease variations in the results due to training choice.

Downloaded At: 14:16 22 March 2011

Table 2. Hekla land cover classes and available reference samples.

Class no.	Land cover class	Training samples	Testing samples
1	Andesite lava 1991	1413	1349
2	Andesite lava 1980	434	414
3	Andesite lava 1970	978	922
4	Lichen-covered basalt lava 1878	1674	1023
5	Andesite lava with sparse moss cover	528	495
6	Andesite Lava with sparse thick cover	2057	1958
7	Andesite Lava with birch bushes	806	775
8	Hyaloclastite formation I	728	676
9	Hyaloclastite formation II	342	342
10	Hyaloclastite formation III	651	620
11	Volcanic tephra	1654	1608
12	Lava-covered with tephra and scoria	350	350
13	Scoria	275	275
14	Firn and glacier ice	242	216
15	Snow	528	484
16	Lava with sparse moss cover	1008	1008
17	Old unvegetated lava 1a	1102	1064
18	Old unvegetated lava 1b	816	782
19	Old unvegetated lava 1c	368	345
20	Andesite lava 1991b	246	164
21	Andesite lava 1980b	748	748
22	Rhyolite	202	202
23	Unvegetated andesite lava	276	253
24	Hyaloclastite formation IV	341	341
Total		17,491	16,667

6. Experiments

6.1 *Data sets pre-processing*

Figure 3 shows R as it is introduced in Section 4 for the Indian Pine data set. Values range from 0 to 1, corresponding to low (dark blue) or high (dark red) correlation values, respectively. Very poor correlated channels (channels 104–108, 149–163 and 220) are eliminated due to water absorption, resulting in a final 200-channel data set (Tadjudin and Landgrebe 1999) that is used in all experiments.

By analysing Figure 3, several strongly correlated subsets of adjacent bands are observed. Various decompositions are thus considered with $n = 4$, 5 and 6 data sources. We refer to each decomposition case as n-SVM, since an individual SVM is used to classify each of the n sources that are defined. Figure 3 highlights the 5-SVM case, defining sources #1 (data channels 1–35), #2 (data channels 36–79) and #3 (data channels 80–103) that encompass strongly correlated adjacent channels. By contrast, the subset defined by channels 109–219, which was previously divided by water absorption bands (i.e. data channels 149–163), is considered as two separate sources (i.e. source #4, data channels 109–148, and #5, data channels 164–219). In the 4-SVM case, however, one single source is defined for this spectral range. Finally, source #2 is split into two sources for the 6-SVM case (data channels 36–59 and 60–79, respectively). In fact, Figure 3 shows that the resulting data sources may be more significant than the original source #2.

The correlation matrix R is also computed (Figure 4) for the Hekla data set in order to identify potential data sources. Data channels 1–2, 106–110 and 151–157 are removed because of noise due to water absorption, eventually resulting in a 143-band data set, which is used throughout the experiments.

Downloaded At: 14:16 22 March 2011

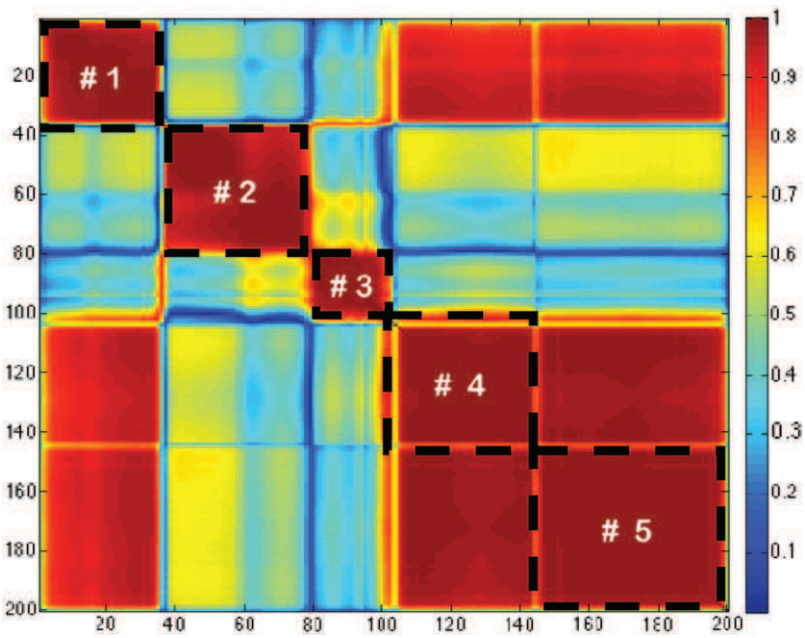

Figure 3. Correlation matrix of the Indian Pine data set.
Note: Black boxes highlight the 5-SVM case in experiments.

Two decompositions are considered for experiment 2. First, Figure 4 shows the 3-SVM case corresponding to the definition of three data sources. As shown in Figure 4, data source #1 (data channels 3–29) is uncorrelated to both data sources #2 (data channels 30–105) and #3 (data channels 111–150). Subsets #2 and #3 are defined since data channels within the latter present a greater inner correlation in comparison to the correlation to channels in subset #2. Second, a 4-SVM case is also considered in order to investigate the impact of defining a greater number of sources with a lower degree of uncorrelation. Now, the former data source #2 is divided into two subsets separated by data channel 53.

Once all the data sources are defined for both experiments, data are standardised by using the mean and variance of each band in order to make all features statistically equal. Then, the data are scaled to the range [0, 1]. This procedure eases the tuning of the SVM kernel parameters (Chen and Lin 2001).

Classification by the SVM fusion ensemble is performed on all the data configurations that have been previously introduced. In addition to the proposed approach, a conventional SVM classifier using the same training and testing sets is applied on both the 200-band Indian Pine image and the 143-band Hekla data set for comparison.

SVM with RBF kernel and OAO multiclass strategy are considered in all experiments, as explained in section 2. In order to evaluate the kernel and the regularisation parameters of the SVM classifiers, a fivefold cross-validation is performed on training data with a user-defined range. All SVM classifications are implemented using LIBSVM (Chen and Lin 2001).

In the proposed SVM fusion ensemble, the rule images, which are obtained from the classification of every data source by a SVM classifier, are standardised and scaled in the same fashion as it is done for the original data. Again, a fivefold cross-validation is performed to determine the SVM parameters.

Downloaded At: 14:16 22 March 2011

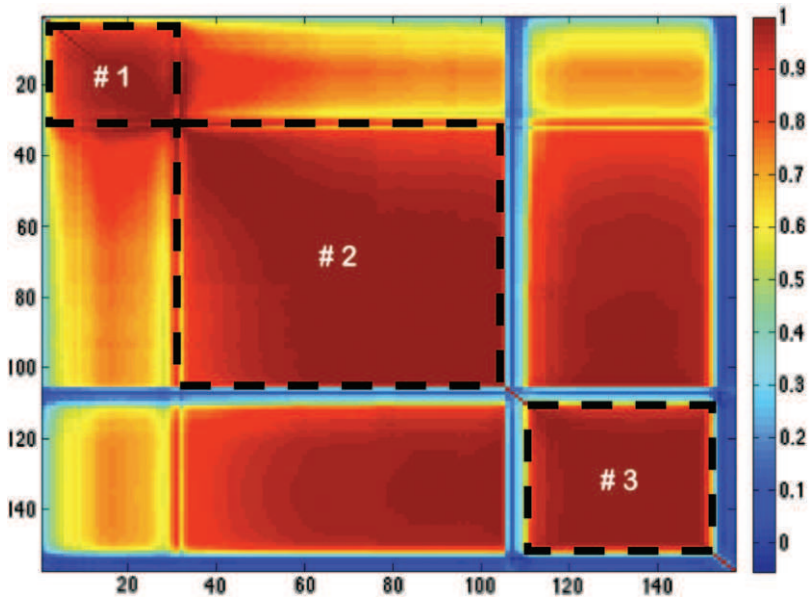

Figure 4. Correlation matrix of the Hekla data set.
Note: Black boxes highlight the 3-SVM case in experiments.

As mentioned previously, all classifications based on the SVM fusion are performed 10 times, using different randomly selected training sets. Afterwards, the accuracies are averaged for a final assessment of the results.

6.2 *Classification results*

The overall results in Figure 5 for the Indian Pine data clearly demonstrate that the proposed SVM ensemble outperforms a standard SVM in terms of accuracy, irrespective of the number of training samples as well as the number of defined subspaces. The improvement of the overall accuracy varies between 1.1% and 2.4%. Using 1/4 of the samples, an overall test accuracy of 88.8% is obtained with the standard SVM. Then, the results are improved by the SVM fusion up to 2.1% in the 6-SVM case. As expected, the results clearly underline the positive impact of increasing the number of training samples. Whereas the overall accuracy for the 4-SVM is 81.8%, using 1/16 of the available samples, the classification accuracy increases up to 89.8%, using 1/4 of all samples. Regarding the number of feature subsets decompositions, the 6-SVM achieves the best accuracies (up to 90.8%) in contrast to the 4-SVM classifier which obtains the smallest improvement. Nevertheless, 4-SVM results still show improvement than those obtained by the standard SVM. In addition to the previous results, Figure 6 compares the Indian Pine's ground truth (i.e. training and testing samples) to the classification map that is obtained by the 6-SVM classifier.

Regarding the class accuracies in Figure 7, similar results are obtained. The SVM ensemble classifiers outperform all standard SVM class accuracies or at least achieve similar results. As in the case of the overall accuracies, the 6-SVM approach provides the best average accuracies. In fact, it improves the accuracy results for most classes, especially the largest classes such as classes #2, #11 and #14, thus underlining the benefits of splitting the data set into several sources. In particular, small classes #1 (alfalfa, 14 training samples) and #9 (oats, 5 training samples) are poorly classified by a standard SVM with accuracies of 69% and 67%, respectively. With the SVM fusion, however, these individual accuracies are improved by at least 9%. In contrast to the good average performance of the 6-SVM approach, these two classes obtain their best results when the SVM fusion is

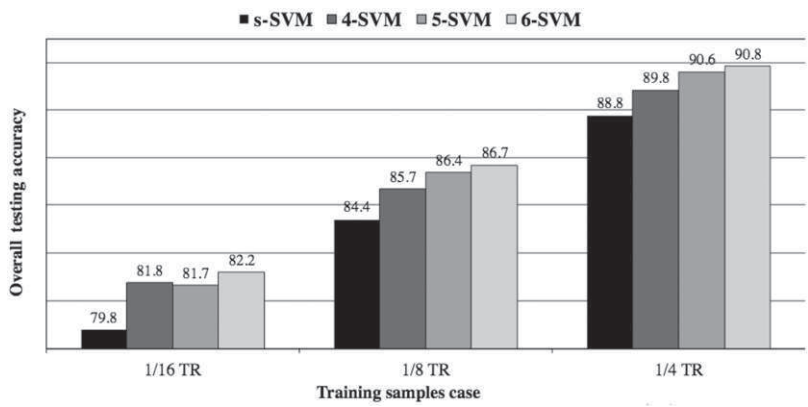

Figure 5. Overall test accuracy for the Indian Pine data set using the SVM fusion ensemble and a standard SVM classifier.

Downloaded At: 14:16 22 March 2011

performed with four individual sources, achieving accuracies up to 78% and 93%, respectively. Furthermore, the analysis of the obtained accuracies shows that the SVM fusion achieves more balanced results as measured by the standard deviation of the classification accuracy. The standard deviation for the class-specific accuracies is 10.4 when classified by a standard SVM. In contrast, the variance drops down to 7.4, using the proposed method.

On the other hand, results for the Hekla data set corroborates the good performances of the SVM fusion in terms of accuracy (Figure 8) that have been shown by the Indian Pine experiment. For the classification of the Hekla data, the improvement of the overall accuracy varies between 2.3% and 6.3%, being greater than for the Indian Pine case. Using all training samples, an overall test accuracy of 90.0% was achieved with a standard SVM. Then, the results are improved by the SVM fusion up to 3.5% in the 4-SVM case. Again, the results demonstrate the positive impact of larger training sets. The overall accuracy for the 4-SVM using 25 samples per class is 73.1%, while classification accuracy

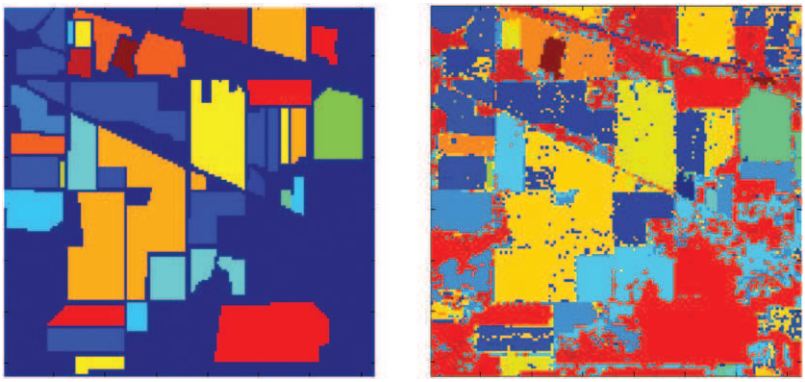

Figure 6. Ground truth and 6-SVM classification map for the Indian Pine data set.

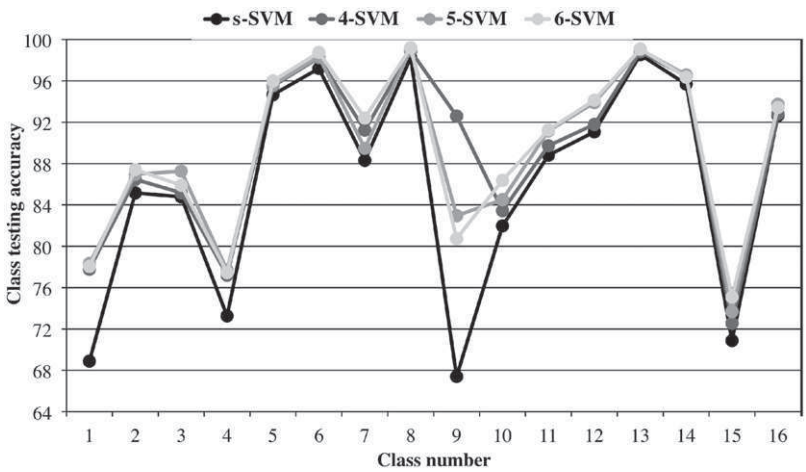

Figure 7. Class test accuracy for the Indian Pine data set using the SVM fusion ensemble and a standard SVM classifier.

Downloaded At: 14:16 22 March 2011

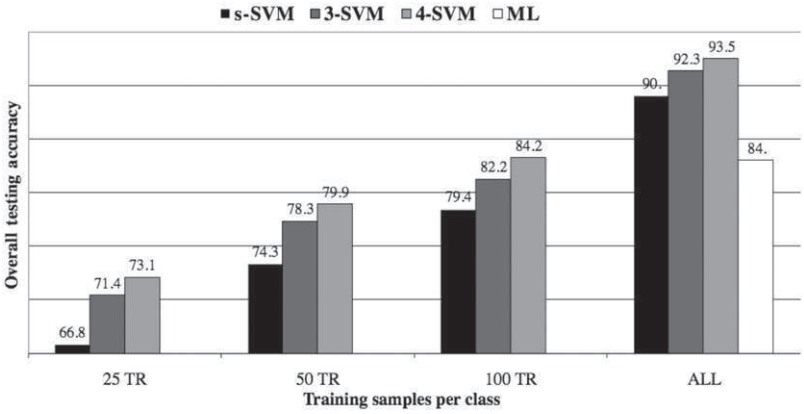

Figure 8. Overall test accuracy for the Hekla data set using the SVM fusion ensemble, a standard SVM classifier and a maximum likelihood (ML) classifier.

increases up to 84.2% using 100 samples per class. Regarding the number of data sources, experiment 2 reaffirms the benefits of a higher number of samples. Certainly, the 4-SVM achieves the best overall accuracy (up to 93.5%) in contrast to the 3-SVM classifier which shows the smallest improvement. The fact that the SVM fusion obtains equal or better accuracies than a standard SVM by using half of the training samples is interesting and it might be useful to reduce fieldwork to be carried out by experts in order to collect reference data. Figure 8 also shows how all SVM-based results outperform a classical maximum likelihood classifier, whose results are poorer mainly due to the Hughes phenomenon.

Figure 9 shows results similar to Figure 7 concerning class accuracies for the Hekla data set. The SVM fusion classifiers outperform all standard SVM class accuracies or, at least, achieves similar results. In general, the 4-SVM approach provides the best average accuracies. As in experiment 1, several classes (e.g. #8, #9, #13 and #20) are significantly improved. Again, the results demonstrate the benefits of splitting the data set into several sources and fusing the outputs. In particular, difficult classes like #9 (hyaloclastite formation II, 342 training samples) and #20 (andesite lava 1991b, 246 training samples) are poorly classified by a standard SVM, while the improvement by 21% or 13% is achieved by the SVM fusion approach. Similar to the Indian Pine case, the SVM ensemble achieved more balanced results for the Hekla data in terms of the standard deviation of the class-specific classification accuracy, as shown in Figure 9.

Concerning the 10 independent runs that are done in each experiment, standard deviation measures demonstrated that classifications considering a few training samples are slightly affected by the choice of training data. Nevertheless, these measures are negligible since they are below the accuracy improvements that have been presented in this section.

7. Discussion and conclusion

In this article, the problem of classifying hyperspectral data was addressed and an approach that is based on the fusion of multiple SVM classifiers was proposed.

Downloaded At: 14:16 22 March 2011

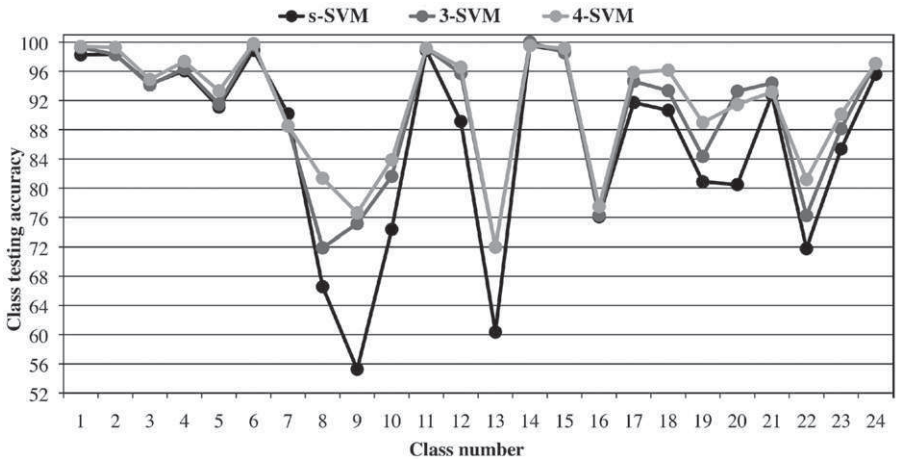

Figure 9. Class test accuracy for the Hekla data set using the SVM fusion ensemble and a standard SVM classifier.

Downloaded At: 14:16 22 March 2011

The proposed method relies on the decision fusion of individual SVM classifiers that are trained on different feature subspaces emerging from a single hyperspectral data set. The separate data sources are defined according to the correlation of the spectral bands. In doing so, each individual SVM classifier may exploit the heterogeneity of the data along the hyperspectral spectrum.

In experiments, different feature space decompositions were performed in order to evaluate the convenience of defining a lower number of data sources for the sake of a greater uncorrelation between them. Then, all the SVM fusion results were compared to those obtained by a standard SVM classifier dealing with the whole hyperspectral data set. The SVM fusion ensemble always outperformed the standard SVM classifier in terms of the overall and class accuracies. Regarding the different number of subspaces, our experimental results underline that the overall and class accuracies increase with a higher number of individual sources. Furthermore, experiments on the Hekla data set demonstrated that the SVM fusion ensemble provides similar accuracies to those yielded by the standard SVM classifier using twice as many training samples. The latter makes the SVM fusion approach particularly interesting in terms of small training sample sets.

One reason for the success of the proposed approach might be the combination of two powerful machine learning concepts: SVM and MCS. On the one hand, SVMs have shown to perform well with high-dimensional data. On the other hand, the conceptual advantage of classifier ensembles has been widely demonstrated. Even though the accuracy of each classifier in the ensemble may decrease due to individual errors, the utilisation of many classifiers, as well as the final result of the decision fusion, eliminates the misclassifications, improving the results of final classifications, as shown by the presented experiments.

Another reason for the success of the proposed approach may be the differences between spectral bands along the hyperspectral image spectrum. For instance, using the conventional SVM for the whole heterogeneous data requires the definition of one single kernel function, which may not be adapted to the whole diversity of information. It might be more adequate to take advantage of this heterogeneity by splitting the data into a few distinct subsets, defining a kernel function, which is adapted for each data source (i.e. a

portion of the hyperspectral image), separately and fuse the derived outputs. Using this, each single SVM is built according the properties of its corresponding feature set and thus the possible discriminative information can be extracted.

Concerning the different decomposition of the hyperspectral data set, the experiments underline that a greater number of subspaces may be more relevant for an optimal data processing than a high degree of non-correlation between data sources. As a matter of fact, the latter assumption is corroborated by the results, which show that the overall accuracies, as well as the class accuracies, increase with a higher number of individual sources. Again, the cause might be that classifier ensembles become more advantageous with more classifiers. The optimal number of data sources as well as the automatic decomposition of the original data set is a topic of our current research.

The previous statements are confirmed by the presented experimental results. Overall, the proposed method can be considered attractive for the classification of hyperspectral remote sensing data. It relies on the original concept of SVM and therefore the strategy can be simply integrated in existing applications. In the future research, a deeper study on the impact of the definition of the data sources (i.e. measure of the spectral bands similarity and number of subsets) on the overall accuracy will be investigated. Furthermore, the reduction of the computational complexity, as well as classification problems with few training samples, need to be investigated.

Downloaded At: 14:16 22 March 2011

Acknowledgements

This research was partially supported by the Research Fund of the University of Iceland.

References

Benediktsson, J.A., Chanussot, J., and Fauvel, M., 2007. Multiple classifier systems in remote sensing: from basics to recent developments. *In*: M. Haindl, J. Kittler and F. Roli, eds. *Multiple Classifier Systems*. Heidelberg, Germany: Springer, 501–512.

Benediktsson, J.A. and Kanellopoulos, I., 1999. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (3), 1367–1377.

Banfield, R., *et al*., 2007. A comparison of decision tree ensemble creation techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 173–180.

Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24 (2), 123–140.

Briem, G.J., Benediktsson, J.A., and 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. and Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 43 (6), 1351–1362.

Chen, C.-C. and Lin, C.-J., 2001. LIBSVM: A library for support vector machines. Available from: http://www.csie.ntu.edu.tw/~cjlin/libsvm [Accessed 10 May 2010].

Dalponte, M., Bruzzone, L., and Gianelle, D., 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions in Geoscience and Remote Sensing*, 46 (5), 1416–1427.

Kittler, J., 1998. Combining classifiers: a theoretical framework. *Pattern Analysis and Applications*, 1, 18–27.

Koetz, B., *et al*., 2008. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management*, 256 (3), 263–271.

Krogh, A. and Vedelsby, J., 1995. Neural network ensembles, cross validation, and active learning. *In*: G. Tesauro, D. Touretzky and T. Leen, eds. *Advances in neural information processing systems*, Vol. 7, Cambridge, UK: MIT Press, 231–238.

Ham, J., *et al*., 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 492–501.

Hansen, L. and Salamon, P., 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 993–1001.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, IT-14, 55–63.

Melgani, F. and Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (8), 1778–1790.

Opitz, D. and Shavlik, J., 1996. Actively searching for an effective neural-network ensemble. *Connection Science*, 8 (3/4), 337–353.

Pal, M. and Mather, P.M., 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems*, 20, 1215–1225.

Pal, M. and Mather, P.M., 2006. Some issues in the classification of DAIS hyperspectral data. *International Journal of Remote Sensing*, 27 (14), 2895–2916.

Roli, F., Giacinto, G., and Vernazza, G., 2001. Methods for designing multiple classifier systems. *Proceedings of the second international workshop on multiple classifier systems*. Cambridge, UK, 78–87.

Schapire, R., 1990. The strength of weak learnability. *Machine Learning*, 5 (2), 197–227.

Scholkopf, B. and Smola, A.J., 2002. *Learning with kernels, support vector machines, regularization, optimization and beyond*. Cambridge: MIT Press.

Tadjudin, S. and Landgrebe, D.A., 1999. Covariance estimation with limited training samples. *IEEE Transactions on Geoscience and Remote Sensing*, 37 (4), 2113–2118.

Vapnik, V.N., 1998. *Statistical learning theory*. New York: Wiley.

Waske, B. and Benediktsson, J.A., 2007. Fusion of support vector machines for classification of multisensor data. *IEEE Transactions in Geoscience and Remote Sensing*, 45, 3858–3866.

Waske, B. and Van der Linden, S., 2008. Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 46 (5), 1457–1466.

Downloaded At: 14:16 22 March 2011