IMPROVING SVM CLASSIFICATION ACCURACY USING A HIERARCHICAL APPROACH FOR HYPERSPECTRAL IMAGES

Begüm Demir,and Sarp Ertürk

Laboratory of Image and Signal Processing (KULIS), Electronic and Telecomm. Eng. Dept., Veziroglu Campus, 41040 Kocaeli, Turkey
{begum.demir, sertur}@kocaeli.edu.tr

ABSTRACT

This paper proposes to combine standard SVM classification with a hierarchical approach to increase SVM classification accuracy as well reduce computational load of SVM testing. Support vectors are obtained by applying SVM training to the entire original training data. For classification, multi-level two-dimensional wavelet decomposition is applied to each hyperspectral image band and low spatial frequency components of each level are used for hierarchical classification. Initially, conventional SVM classification is carried out in the highest hierarchical level (lowest resolution) using all support vectors and a one-to-one multiclass classification strategy, so that all pixels in the lowest resolution are classified. In the sub-sequent levels (higher resolutions) pixels are classified using the class information of the corresponding neighbor pixels of the upper level. Therefore, the classification at a lower level is carried out using only the support vectors of classes to which corresponding neighbor pixels in the higher level are assigned to. Because classification with all support vectors is only utilized at the lowest resolution and classification of higher resolutions requires a subset of the support vectors, this approach reduces the overall computational load of SVM classification and provides reduced SVM testing time compared to standard SVM. Furthermore, the proposed approach provides significantly better classification accuracy as it exploits spatial correlation thanks to hierarchical processing.

Index Terms— Hyperspectral image, hierarchical, wavelet transforms, classification, support vector machine

1. INTRODUCTION

Hyperspectral imaging systems provide high-resolution spectral information for a scene in the form of hundreds of narrow spectral band images, and it is possible to classify regions or identify objects within the scene with much higher accuracy compared to standard vision sensors.

An important research topic in hyperspectral imaging is to develop approaches that can provide high classification accuracies. Kernel based hyperspectral image classification algorithms such as support vector machines (SVMs) [1-3] and relevance vector machines (RVMs) [4] have been shown to provide higher classification accuracies than alternative approaches and have therefore become very popular in recent years. Research to even further increase the classification accuracy of conventional SVM and RVM based classification is ongoing.

In [5-8], it is proposed to combine spatial and spectral information of hyperspectral images to provide higher accuracy in hyperspectral image classification. In [5], spatial feature vectors are obtained using either the mean only, or the mean and standard deviation together of a certain neighborhood window of the corresponding feature vector, and kernel matrices corresponding to spatial and spectral feature vectors are computed separately and then combined using different combination approaches. In [6,7] it is proposed to use morphological profiles (MPs), which are obtained by applying opening and closing operations to the first several principle components of the hyperspectral data, for increased classification accuracy. MPs are directly used for neural network classification in [6] and fused with original features for SVM classification in [7]. A two stage classification algorithm is proposed in [8] to combine spatial and spectral information. In [8], it is noted that conventionally the class of any pixel and the class of at least one of its neighbors is the same, and this is referred to as "same class neighborhood property". In [8], hyperspectral images are first classified using SVM classification and the initial classes of each pixel and its eight neighbors are identified in the first stage. Then, each pixel is classified by a binary decision tree based hierarchical classifier using this information in the second stage.

Although the approach proposed in [8] uses a binary decision tree based hierarchical classifier in the second stage to fine-tune the first classification results that are obtained in a conventional way, an entirely hierarchical classifier had not been presented so far. This paper proposes a novel entirely hierarchical classification approach that makes use of the same class neighborhood property but uses an entirely hierarchical approach for the classification of hyperspectral images accomplished using multi-level wavelet decomposition. Multiple hierarchical levels are constructed from the full resolution hyperspectral image up to low resolution, using wavelet decomposition. Classification is performed at the lowest resolution first, and then classification results are disseminated downwards using the same class neighborhood property. It is shown that the proposed approach provides improved classification accuracy as well as reduced SVM classification time.

2. SUPPORT VECTOR MACHINE BASED CLASSIFICATION

SVM classification has become a very popular kernel based classification algorithm in hyperspectral image classification because it can provide high classification accuracy [1-3]. SVM tries to find the optimal separating hyperplane that maximizes the margin between the closest training sample and the separating hyperplane, and uses boundary pixels (support vectors) to create a decision surface. SVM estimates a classification function using training data from two classes: $(\mathbf{x}_1, y_1),...,(\mathbf{x}_f, y_f) \in \Re^n \times \{\pm 1\}$. In kernel based SVM classification, the original input space is mapped to a higher-dimensional (Hilbert) feature space ($\phi : \Re^n \to H$) using a kernel function. A kernel function is a function that corresponds to an inner product in some expanded feature space. The classification function is obtained by solving the convex optimization problem:

$$\begin{aligned} &\text{maximize}: \sum_{u=1}^{f} \alpha_u - \frac{1}{2} \sum_{u=1}^{f} \sum_{v=1}^{f} \alpha_u \alpha_v y_u y_v K(\mathbf{x}_u, \mathbf{x}_v) \\ &\text{subject to: } \sum_{u=1}^{f} \alpha_u y_u = 0 \text{ and } 0 \le \alpha_u \le C \end{aligned} \quad (1)$$

where C controls the trade-off between complexity (number of support vectors) and data miss-fit (number of non-separable points), and is chosen by the user (i.e. set a priory). The kernel function $K(\mathbf{x}_u, \mathbf{x}_v) = \phi(\mathbf{x}_u)\phi(\mathbf{x}_v)$ does actually not require a direct knowledge of the transform function $\phi(.)$. Here, α_u and α_v are Lagrange multipliers. Each non-zero α_u indicates that the corresponding $\mathbf{x}_u$ is a support vector. The non-linear classifier for a sample $\mathbf{x}$ can then be expressed as

$$f = \text{sgn}\left(\sum_{u=1} \alpha_u y_u K(\mathbf{x}_u, \mathbf{x}) + b \right) \quad (2)$$

Kernel functions used in SVM must satisfy Mercer's condition which requires the kernel to be a continuous symmetric kernel of a positive integral operator. Popular kernels implementing this condition are the Linear kernel $K(\mathbf{x}_u, \mathbf{x}) = \mathbf{x}_u \cdot \mathbf{x}$, the Polynomial kernel defined as $K(\mathbf{x}_u, \mathbf{x}) = (\gamma\, \mathbf{x}_u \cdot \mathbf{x})^d$ and the Radial Basis Function (RBF) kernel $K(\mathbf{x}_u, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_u - \mathbf{x}\|^2)$, where d and γ are kernel parameters. For multiclass SVM, it is possible to combine multiple binary classifiers. The one-against-one approach is utilized in this paper for multiclass SVM because it provides fast training. In the one-against-one method, $K(K-1)/2$ binary classifiers are trained and $K(K-1)/2$ binary tests are required to make a final decision, where K is the total number of classes. Each outcome gives one vote to the winning class, and the class with the most votes is selected as the final result.

3. HIERARCHICAL HYPERSPECTRAL IMAGE CLASSIFICATION

In the proposed method, SVM training is applied as conventional to the original training data to obtain support vectors of all classes. SVM classification (testing), on the other hand, is applied to the hyperspectral data in a hierarchical way. For this purpose, multi-level wavelet decomposition is applied to each hyperspectral image band in the spatial domain, to obtain the hierarchical levels of different resolutions. The highest level with have the lowest spatial resolution and the lowest level will be of full spatial resolution. The proposed classification approach is depicted in Figure 1.

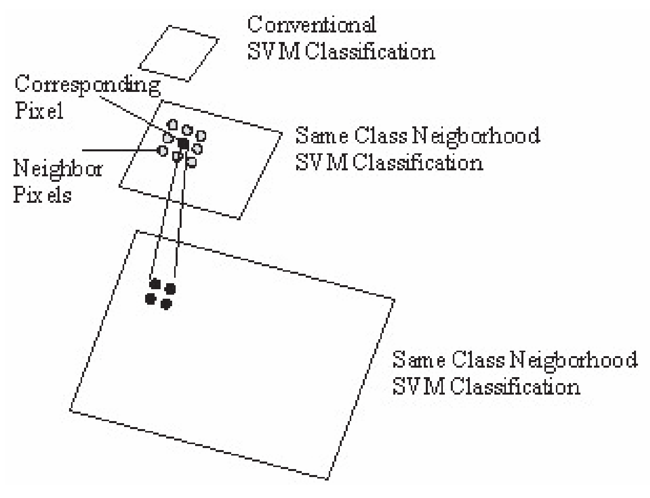


Figure 1. Proposed hierarchical classification approach.

The proposed hierarchical approach can be summarized in the following steps:

1- Multi-level wavelet decomposition is applied to all hyperspectral image bands in the spatial domain, in order to obtain the hierarchical levels of different spatial resolution.

2- Hyperspectral pixels of the highest level (lowest spatial resolution) are classified using conventional SVM. All support vectors of all class are used to perform one-against-one multi-class classification using all classes. Each pixel at the highest level is assigned the corresponding class label.

3- Hyperspectral pixels at sub-sequent lower levels (higher resolutions) are classified using same class neighborhood SVM classification. For a pixel of the lower level, the corresponding pixel of the higher level is identified and the eight neighbors of this pixel are noted. The pixel at the lower level is classified to belong to one of the classes of these nine pixels at the higher level in accordance with the same-class neighborhood property. Again, SVM is used for classification, however, this time only support vectors corresponding to the classes of these nine pixels are utilized, so that the pixel at the lower level will be classified into one of the classes to which these nine pixels belong to.

4- Similarly, same class neighborhood SVM classification is utilized for lower levels, until the full resolution hyperspectral image (lowest level) has been classified.

Thanks to hierarchical processing, the proposed approach makes use of spatial information through the same class neighborhood property whilst obtaining classification maps of lower levels from previous higher levels. Because each pixel at high-resolution is classified using support vectors of the classes of its neighbors of the previous level (low resolution), the proposed algorithm provides reduced SVM testing time compared to standard SVM applied at full resolution that uses all support vectors of each classes in SVM testing.

4. EXPERIMENTAL RESULTS

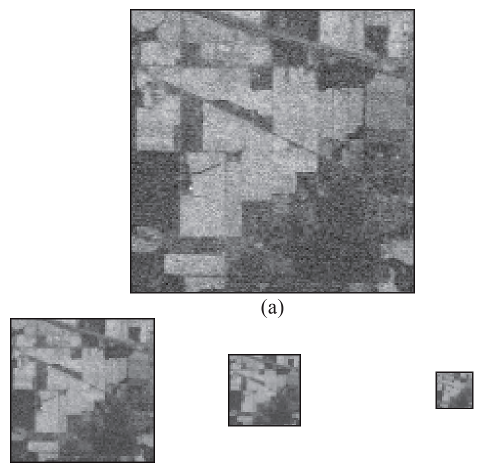
The performance of the proposed approach is presented using a sample hyperspectral image taken over northwest Indiana's Indian Pine test site in June 1992 [9]. The Indian Pine hyperspectral data consists of 145×145 pixels with 220 bands. The number of bands is initially reduced to 200 by removing bands covering water absorption and noisy bands. Although the original ground truth includes 16 classes, nine classes that have a higher number of samples have been selected in this paper and the total number of samples corresponding to each class is shown in Table I.

SVM classification is used with RBF kernel in the presented results. The penalty parameter of SVM is set to 40 a priory and the γ parameter of the RBF kernel is tested between [0.1-2] using a five fold cross validation.

TABLE I. NUMBER OF SAMPLES (NOS) FOR EACH CLASS OF THE INDIAN PINE DATA

Class	NoS
Corn-no till	1434
Corn-min till	834
Grass/Pasture	497
Grass/Trees	747
Hay-windrowed	489
Soybean-no till	968
Soybean-min till	2468
Soybean-clean till	614
Woods	1294
Total	9345

In the presented results, 2D wavelet decomposition with three levels is applied to each hyperspectral image band. Daubechies wavelets with one vanishing moment are used to implement the 2D wavelet decomposition. Figure 2 shows the hierarchical images obtained for a sample band of the hyperspectral image.



(a) (b) (c) (d)

Figure 2. (a) Original hyperspectral image band at full resolution (band #146), (b) low frequency components obtained after one-level wavelet decomposition (c) low frequency components obtained after two-level wavelet decomposition (d) low frequency components obtained after three-level wavelet decomposition

Table II shows direct SVM (SVM) and proposed hierarchical SVM (H-SVM) classification accuracy (CA) results together with SVM testing times for %10 training data rate (TDR) and %50 TDR. TDR shows the rate of samples of each class that are used as initial training data. SVM testing time is provided in seconds and obtained on an Intel Core 2 Duo CPU 7300, 1.7 GHz, and 2GB memory notebook. MATLAB® is used to implement the techniques.

Experimental results show that proposed algorithm significantly improves conventional SVM classification accuracy and reduces the computational time of testing. It is seen that the classification accuracy is increased by about

6% in case of a TDR of 10% demonstrating the case with a low number of training data. The classification accuracy is increased by about 3% in case of 50% TDR. The proposed approach therefore provides more gain in case of low amount of training data, which is important because it is mostly the case in remote sensing applications. It is furthermore seen that the gain in computational time increases with TDR; because more support vectors are obtained in case of increased training data amount, which increases the computational load of conventional SVM.

Figure 3 shows the original ground truth image of the Indian Pine data as well as classification maps of direct SVM and the proposed H-SVM. It is seen that the proposed approach particularly reduces incorrect classifications that occur at central locations of the fields, thanks to spatial relations exploited through the hierarchical processing and same class neighborhood property.

5. CONCLUSIONS

This paper presents a novel hierarchical classification approach which improves SVM classification accuracy and reduces computational time of SVM classification. Multi-level wavelet decomposition is applied to each hyperspectral image band to enable hierarchical processing. Pixels of the highest level (lowest resolution) are classified using all support vectors, while pixels at sub-sequent lower levels (higher resolution) are classified using same class neighborhood SVM classification, that only uses support vectors of classes to which the corresponding neighbors of the higher level belong to. The proposed approach increases classification accuracy and reduces classification time.

6. REFERENCES

[1] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778-1790, Aug 2004.

[2] G. Camps-Valls, and L. Bruzzone, “Kernel-based methods for hyperspectral image classification,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351-1362, June 2005.

[3] J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, “Support vector machine classifiers as applied to AVIRIS data,” *in Summaries 8th JPL Airborne Earth Science Workshop*, JPL Pub. 99-17, 1999, pp. 217–227.

[4] B. Demir, and S. Ertürk,“Hyperspectral image classification using relevance vector machines,” *IEEE Geoscience and Remote Sensing Letters*, vol.4, no.4, pp.586-590, Oct. 2007.

[5] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, “Composite kernels for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, Jan 2006.

[6] J. A. Palmason, J. A. Benediktsson, J. R. Sveinsson, and J. Chanussot, “Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis,” *in Proc. IGARSS*, 2005, Seoul, Korea, pp. 176–179.

[7] M. Fauvel, J. A. Benediktsson, J. Chanussot and J. R. Sveinsson, “Spectral and spatial classification of hyperspectral data using SVMs and morphological profile,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804-3814, Nov 2008.

[8] A. Keshavarz, H. Ghassemian, H. Dehghani, “Hierarchical classification of Hyperspectral images by using SVMs and ‘same class neighborhood property’,”*in Proc. IGARSS*, 2005, Seoul, Korea, pp.3219-3222.

[9] D. Landgrebe, ‘AVIRIS NW Indiana’s Indian Pines 1992 data set,’http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html

TABLE II. CLASSIFICATION ACCURACY (CA) AND SVM TESTING TIME (t) RESULTS OF STANDARD SVM AND PROPOSED H-SVM FOR 10 % TDR AND 50 % TDR

Method	10 % TDR		50 % TDR	
	CA	t (second)	CA	t (second)
SVM	83.02	6.54	93.77	27.73
H-SVM	**89.11**	**4.87**	**97.15**	**9.83**

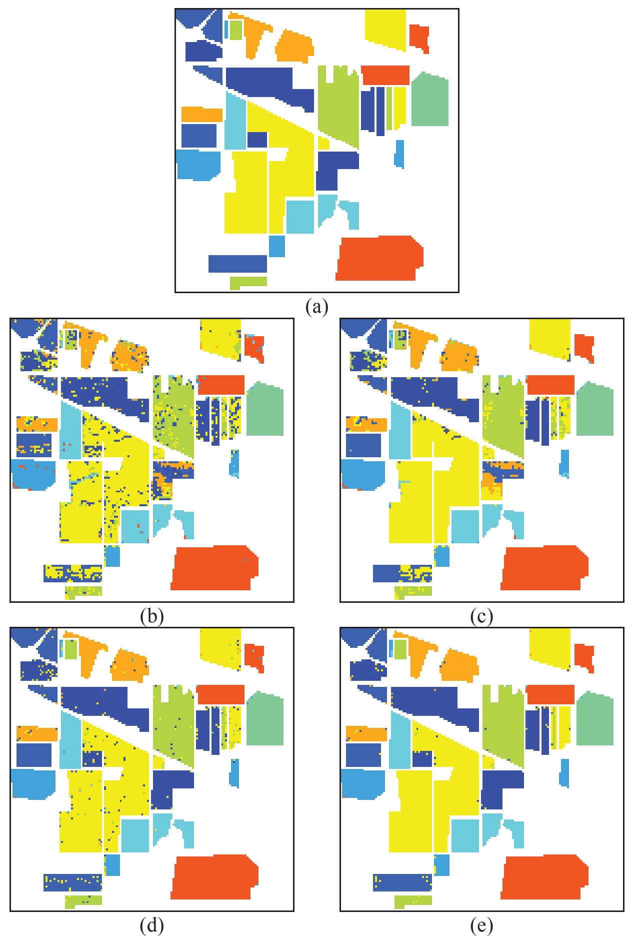


Figure 3. (a) Original ground truth image (b) classification map of direct SVM for a TDR of 10%, (c) classification map of proposed H-SVM for a TDR of 10%, (d) classification map of direct SVM for a TDR of 50%, (e) classification map of H-SVM for a TDR is 50%.