Spectral-Spatial Classification of Hyperspectral Images Using Hierarchical Optimization

Yuliya Tarabalka, James Tilton

SPECTRAL-SPATIAL CLASSIFICATION OF HYPERSPECTRAL IMAGES USING HIERARCHICAL OPTIMIZATION

Yuliya Tarabalka and James C. Tilton

NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
e-mail: yuliya.tarabalka@nasa.gov

ABSTRACT

A new spectral-spatial method for hyperspectral data classification is proposed. For a given hyperspectral image, probabilistic pixelwise classification is first applied. Then, hierarchical step-wise optimization algorithm is performed, by iteratively merging neighboring regions with the smallest Dissimilarity Criterion (DC) and recomputing class labels for new regions. The DC is computed by comparing region mean vectors, class labels and a number of pixels in the two regions under consideration. The algorithm is converged when all the pixels get involved in the region merging procedure. Experimental results are presented on two hyperspectral remote sensing images acquired by the AVIRIS and ROSIS sensors. The proposed approach improves classification accuracies and provides maps with more homogeneous regions, when compared to previously proposed classification techniques.

Index Terms— Hyperspectral imaging, hierarchical segmentation, classification, support vector machines.

1. INTRODUCTION

In hyperspectral imagery, each pixel is represented by a detailed spectrum of the received light. Since different substances exhibit different spectral signature, hyperspectral imagery is a well-suited technology for accurate image classification. However, a large number of spectral channels presents challenges to image analysis.

An extensive literature is available on classification of hyperspectral images [1, 2]. Recent studies have shown the advantage of considering the correlations between spatially adjacent pixels for accurate image classification, i.e., applying spectral-spatial classification [3, 4]. One of the recently proposed approaches consists in performing image segmentation (partitioning of the image into homogeneous regions) and then using the identified regions as adaptive neighborhoods for all the pixels within these regions [5]. However, the accuracy of segmentation results strongly depends on the chosen criterion of region homogeneity. In order to mitigate this dependence, we recently proposed to perform probabilistic classification for selecting the most reliably classified pixels as markers, or region seeds, for region growing [4]. This technique led to a significant improvement of classification accuracies when compared to previously proposed methods. The drawback of this method is that the selection of markers strongly depends on the performance of the initial classifier: non-marked regions disappear in the final classification map, while if a marker is classified to the wrong class, the whole region grown from this marker risks to be wrongly classified.

In this work, we propose to use a Hierarchical Step-Wise Optimization (HSWO) method for including spatial dependencies into a classification procedure. HSWO is a segmentation approach, which iteratively merges pairs of the most similar spatially adjacent regions, and generates at its output a hierarchical set of image segmentations [6]. We propose to use supervised classification results for computing more accurately a sequence of region merges and for defining a convergence criterion, leading to a single spectral-spatial classification map. Thus, a new *Classification and Hierarchical Optimization (CaHO)* method for hyperspectral images is proposed. First, probabilistic pixelwise classification of the input image is performed. Then, at each iteration two neighboring regions with the smallest Dissimilarity Criterion (DC) are merged, and a class label for a new region is computed. The DC between regions is defined as a function of region statistical features, a number of pixels in the considered regions and their class labels. When all image pixels get involved in region merging, the algorithm converges, resulting in a spectral-spatial classification map.

The paper is organized as follows. The next section presents a new *CaHO* method. Experimental results are presented and discussed in Section 3 . Finally, conclusions are drawn in Section 4.

2. PROPOSED METHOD

On the input a B-band hyperspectral image is given, which can be considered as a set of n pixel vectors $\mathbf{X} = \{\mathbf{x}_j \in \mathbb{R}^B, j = 1, 2, ..., n\}$. The objective is to compute a classification map $\mathbf{L} = \{L_j, j = 1, 2, ..., n\}$, where each pixel $\mathbf{x}_j$ is assigned to one of K thematic classes (i.e., has a class label L_j). The proposed *CaHO* method, illustrated in Fig. 1, is composed of two main steps:
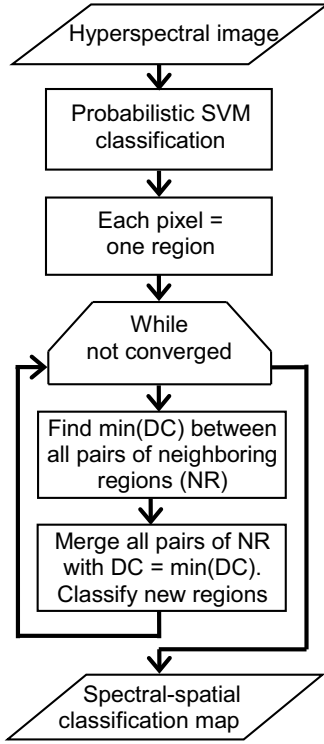
Fig. 1. Flowchart of the proposed *CaHO* approach. "DC" means Dissimilarity Criterion.

2.1. Probabilistic pixelwise classification

The aim of the first step is to compute a classification map $\mathbf{L} = \{L_j, j = 1, 2, ..., n\}$ for a given hyperspectral image, where each pixel has a unique class label, and class probabilities for each pixel $\{P(L_j = k|\mathbf{x}_j), k = 1, ..., K\}, j = 1, 2, ..., n$. We propose to perform probabilistic Support Vector Machines (SVM) classification for this purpose, which is extremely well suited for classifying hyperspectral data [1]. We refer the reader to [1] and [7] for details on the SVM method, and to [4] for details on how class probabilities are estimated using pairwise coupling of binary probability estimates.

2.2. Hierarchical optimization

At this step, regularization of the classification map obtained at the previous step is performed, by applying a new hierarchical optimization approach as follows:

    1) Initialize the optimization by labeling each image pixel as a separate region. Each one-pixel region R_i has a class label $L(R_i)$ and a K-dimensional vector of class probabilities $\{P_k(R_i) = P(L(R_i) = k|R_i), k = 1, ..., K\}$.

    2) Compute the DC between all pairs of neighboring re-

gions, using an eight-connectivity neighborhood. A DC between two regions R_i and R_j $DC(R_i, R_j)$ is calculated using the following algorithm:

- Compute the dissimilarity measure $DC_{spectral}(R_i, R_j)$ between two regions by comparing spectral values of the pixels within these regions. We investigated the use of two dissimilarity measures for this purpose. The Spectral Angle Mapper (SAM) between the region mean vectors $\mathbf{u}_i = (u_{i1}, ..., u_{iB})^T$ and $\mathbf{u}_j = (u_{j1}, ..., u_{jB})^T$ is defined as the angle between them:

$$SAM(\mathbf{u}_i, \mathbf{u}_j) = \arccos\left(\frac{\sum_{b=1}^{B} u_{ib}u_{jb}}{[\sum_{b=1}^{B} u_{ib}^2]^{1/2}[\sum_{b=1}^{B} u_{jb}^2]^{1/2}}\right). \tag{1}$$

The Square root of band sum Mean Squared Error (MSE) measure is based on minimizing the increase of MSE between the region mean vector and the original image data and is computed as

$$MSE(\mathbf{u}_i, \mathbf{u}_j) = \left[\frac{n_i n_j}{(n_i + n_j)}\sum_{b=1}^{B}(u_{ib} - u_{jb})^2\right]^{1/2}, \tag{2}$$

where n_i and n_j is a number of pixels in the regions R_i and R_j, respectively.

- If the regions have equal class labels $L(R_i) = L(R_j)$, the DC between these regions

$$DC(R_i, R_j) = DC_{spectral}(R_i, R_j). \tag{3}$$

- If the regions have different class labels $L(R_i) \neq L(R_j)$, the DC between them is found as:

    a) If $n_i > M$ and $n_j > M$, $DC(R_i, R_j) = \infty$ (the upper *maximum* value of *float*). This means that if a two large regions are classified to different classes, they cannot be merged together. This condition is included for favoring merging small regions. We propose to set $M = 20$.

    b) Otherwise,

$$DC(R_i, R_j) = W \cdot DC_{spectral}(R_i, R_j), \tag{4}$$

where $W > 1$. This means that if two regions have different class labels, the DC between them is penalized by a constant W.

    3) Find the smallest DC value DC_{min}.

    4) Merge all pairs of neighboring regions satisfying $DC = DC_{min}$. For each new region R_{new} created by merging two regions R_i and R_j, recalculate:

- Class probabilities as

$$P_k(R_{new}) = \frac{P_k(R_i)n_i + P_k(R_j)n_j}{n_{new}}, \tag{5}$$

$k = 1, ..., K$, where $n_{new} = n_i + n_j$.

Table 1. Information Classes, Number of Labeled Samples (No. of Samp.) and Classification Accuracies in Percentage for the *Indian Pines* Image.

$DC_{spectral}$	No. of Samp. Train	No. of Samp. Test	SVM	ECHO	SVM MSF	HSEG +MV	CaHO (W = 1.5) SAM	CaHO (W = 1.5) MSE
Overall Accuracy	-	-	78.17	82.64	88.41	**90.86**	88.87	89.15
Average Accuracy	-	-	85.97	83.75	91.57	**93.96**	93.75	93.82
Corn-no till	50	1384	78.18	83.45	90.97	90.46	**95.38**	94.22
Corn-min till	50	784	69.64	75.13	69.52	**83.04**	80.36	79.21
Corn	50	184	91.85	92.39	95.65	95.65	**97.28**	96.20
Soybeans-no till	50	918	82.03	90.10	**98.04**	92.06	97.28	94.99
Soybeans-min till	50	2418	58.95	64.14	81.97	**84.04**	73.53	74.52
Soybeans-clean till	50	564	87.94	89.89	85.99	**95.39**	89.89	94.86
Alfalfa	15	39	74.36	48.72	94.87	92.31	**97.44**	94.87
Grass/pasture	50	447	92.17	94.18	94.63	94.41	93.96	**97.32**
Grass/trees	50	697	91.68	96.27	92.40	97.56	**97.70**	97.56
Grass/pasture-mowed	15	11	**100**	36.36	**100**	**100**	**100**	**100**
Hay-windrowed	50	439	97.72	97.72	**99.77**	99.54	99.54	99.32
Oats	15	5	**100**	**100**	**100**	**100**	**100**	**100**
Wheat	50	162	98.77	98.15	**99.38**	98.15	**99.38**	**99.38**
Woods	50	1244	93.01	94.21	97.59	98.63	98.63	**99.04**
Bldg-Grass-Tree-Drives	50	330	61.52	81.52	68.79	**82.12**	79.70	81.82
Stone-steel towers	50	45	97.78	97.78	95.56	**100**	**100**	97.78

- Class label as

$$L(R_{new}) = \arg\max_{k}\{P_k(R_{new})\}. \qquad (6)$$

5) Stop if each image pixel has been involved at least once in the region merging procedure. Otherwise, recalculate the DC values for the new regions and all regions spatially adjacent to them, and return to step 3.

The proposed convergence criterion assumes that the image does not contain one-pixel regions of interest. If such regions may exist, the algorithm must be converged earlier. The convergence criterion in this case can for instance compare class probabilities of next candidates for merging, and stop the procedure when these candidates belong to different classes with probabilities higher than the defined threshold. Another, simpler criterion consists in stopping the algorithm when $[(1-P)n]$ pixels get involved in region merging, where P ($0 < P < 1$) is a probability of occurrence of one-pixel regions in the considered image. Since the images used for our experiments do not contain one-pixel regions of interest, we use the convergence criterion proposed in step 5.

3. EXPERIMENTAL RESULTS AND DISCUSSION

We applied the proposed *CaHO* method to two hyperspectral airborne images described in the following:

1) The *Indian Pines* image was recorded by the AVIRIS sensor over the vegetation area. It is of 145 by 145 pixels, with a spatial resolution of 20 m/pixel and 200 spectral channels. Sixteen information classes are considered, which are detailed in Table 1, with the number of training and test samples for each class. Training samples were randomly selected

Table 2. *CaHO* Overall and Average Classification Accuracies (OA and AA, respectively) for the *Indian Pines* Image for Different Values of the Parameter W.

DC	W	1.0	1.25	1.5	1.75	2.0	3.0
SAM	OA	87.34	88.61	**88.87**	88.42	88.44	86.80
SAM	AA	91.74	93.72	**93.75**	93.04	93.49	92.81
MSE	OA	88.93	88.34	**89.15**	87.81	88.08	87.26
MSE	AA	87.48	92.65	**93.82**	93.23	93.32	93.27

Table 3. Classification Accuracies in Percentage for the *Center of Pavia* Image.

$DC_{spectral}$	SVM	SVM MSF	HSEG +MV	CaHO (W = 1.5) SAM	CaHO (W = 1.5) MSE
Overall Acc.	94.96	91.31	**96.67**	96.58	96.51
Average Acc.	92.56	92.64	95.41	**95.61**	95.60

from the reference data. The remaining samples composed the test set.

2) The *Center of Pavia* image was acquired by the ROSIS sensor over the urban area of Pavia, Italy. The image is of 785 by 300 pixels, with a spatial resolution of 1.3 m/pixel, 102 spectral channels and nine classes of interest. Thirty samples for each class were randomly chosen from the reference data as training samples. The remaining samples were used as the test set. More information about the image, with the used training-test set can be found in [8].

For both images, the probabilistic one-versus-one SVM classification with the Gaussian Radial Basis Function (RBF) kernel was performed. The optimal parameters C (penalty during the SVM optimization) and γ (spread of the RBF ker-
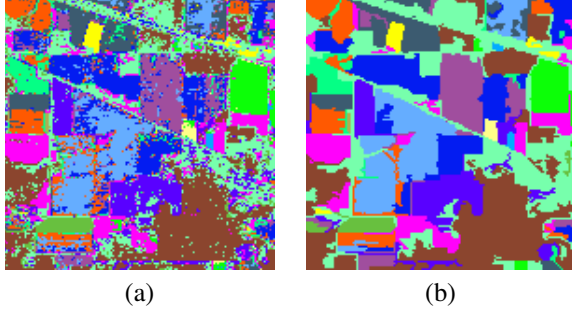
(a)   (b)

Fig. 2. *Indian Pines* image. (a) SVM classification map. (b) *CaHO* classification map (MSE $DC_{spectral}$, $W = 1.5$).

nel) were selected by fivefold cross validation. Then, the proposed hierarchical optimization was applied using the SAM and the MSE spectral dissimilarity measures (the algorithm was implemented using the Hierarchical Segmentation software [9]). Table 2 gathers overall and average (i.e., average over the classes) accuracies of the *CaHO* method for the *Indian Pines* image for different values of the parameter W. It can be seen from the table that the method is robust to the choice of W, and quite a wide range of values of W leads to high classification accuracies for both SAM and MSE dissimilarity measures. The best accuracies are achieved with $W = 1.5$.

Table 1 summarizes global [4] and class-specific accuracies of the pixelwise SVM classification and the proposed *CaHO* technique with $W = 1.5$ for the *Indian Pines* image. In order to compare the results of the proposed method with other advanced techniques, we have included results of the ECHO classification [10], a classification using the construction of a minimum spanning forest from the SVM-derived markers (*SVMMSF*) [4] and a classification by majority voting within neighborhoods defined by HSEG segmentation (*HSEG+MV*, with $S_{wght} = 0.0$, which is equivalent to HSWO, and the SAM DC) [5]. Table 3 gives global accuracies of the SVM, *SVMMSF*, *HSEG+MV* and *CaHO* classification methods for the *Center of Pavia* image. As can be seen from the tables, the *HSEG+MV* and the *CaHO* methods yield the best global and most of the class-specific accuracies (the average accuracies of these approaches are non-significantly different). However, in the *HSEG+MV* method a segmentation map was chosen interactively from the segmentation hierarchy, while the *CaHO* method is automatic. Fig. 2(b) shows the *CaHO* classification map (with the MSE $DC_{spectral}$, $W = 1.5$), which is less noisy when compared to the SVM map (see Fig. 2(a)).

4. CONCLUSIONS

In this paper, a new *CaHO* method for spectral-spatial classification of hyperspectral images is proposed. The method consists in performing a probabilistic pixelwise classification, followed by a hierarchical optimization, where at each step two "closest" neighboring regions are merged, and a classification map is recomputed. Experimental results demonstrate that the proposed method improves classification accuracies, when compared to previously proposed classification schemes, and is sufficiently robust for classifying different kinds of images.

5. REFERENCES

[1] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geos. and Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, June 2005.

[2] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*, Wiley-Interscience, 2007.

[3] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral RO-SIS images over Pavia city, northern Italy," *Int. Journal of Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, 2009.

[4] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Systems, Man, and Cybernetics: Part B*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[5] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[6] J.-M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: a stepwise optimization approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 2, pp. 150–163, Feb 1989.

[7] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[8] Y. Tarabalka, J. C. Tilton, J. A. Benediktsson, and J. Chanussot, "Marker-based hierarchical segmentation and classification approach for hyperspectral imagery," in *Proc. of ICASSP'11*, Prague, Czech Republic, 2011.

[9] J. C. Tilton, "RHSEG users manual: Including HSWO, HSEG, HSEGExtract, HSEGReader and HSEGViewer, version 1.50," *available via email request to James.C.Tilton@nasa.gov*, 2010.

[10] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, Inc., 2003.