BEST MERGE REGION GROWING WITH INTEGRATED PROBABILISTIC CLASSIFICATION FOR HYPERSPECTRAL IMAGERY

Yuliya Tarabalka and James C. Tilton

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

ABSTRACT

A new method for spectral-spatial classification of hyperspectral images is proposed. The method is based on the integration of probabilistic classification within the hierarchical best merge region growing algorithm. For this purpose, preliminary probabilistic support vector machines classification is performed. Then, hierarchical step-wise optimization algorithm is applied, by iteratively merging regions with the smallest Dissimilarity Criterion (DC). The main novelty of this method consists in defining a DC between regions as a function of region statistical and geometrical features along with classification probabilities. Experimental results are presented on a 200-band AVIRIS image of the North-western Indiana's vegetation area and compared with those obtained by recently proposed spectral-spatial classification techniques. The proposed method improves classification accuracies when compared to other classification approaches.

Index Terms— Hyperspectral images, classification, segmentation, region growing, support vector machines.

1. INTRODUCTION

Hyperspectral imaging [1], which acquires hundreds of spectral channels for each pixel, opens new perspectives in classification of remote sensing images. While pixelwise classification methods process each pixel independently without considering the correlations between spatially adjacent pixels [2, 3], recent studies have shown the advantage of including information from a spatial neighborhood for accurate classification, i.e., performing spectral-spatial classification [4, 5].

In previous works, we have proposed to perform a segmentation of the input image and then use the identified segments as pixels' neighborhoods [6]. However, automatic segmentation of hyperspectral images is a challenging task, since its performance strongly depends on the chosen measure of region homogeneity. Hierarchical multilevel definition of regions (achieved by performing iterative best merge region growing) mitigates this dependence [7], but this approach is not completely automated (the user is required to set the parameter defining the multilevel segmentation) and

computationally expensive. An alternative spectral-spatial classification method consists in performing probabilistic classification for selecting the most reliably classified pixels as markers for segmentation [5]. In this case, the choice of markers strongly depends on the performance of the selected classifier.

For both approaches, segmentation and classification techniques were applied one after another, and their outputs were combined in some way. In this work, we aim to interlace these procedures and develop a new methodology, where segmentation and classification are performed concurrently for obtaining an accurate thematic map. For this purpose, a new *Hierarchical Segmentation with integrated Classification (HSwC) method* for hyperspectral data is proposed. First, preliminary probabilistic classification of each pixel is performed. Then, best merge region growing is applied, where at each iteration two "closest regions" are merged, and classification probabilities for a new region are recomputed. The main idea behind this new method consists in defining a Dissimilarity Criterion (DC) between regions as a function of region statistical features, classification probabilities and geometrical features. The algorithm is converged when all the pixels get a definite classification label. The output of the proposed *HSwC* method is a thematic map, where every region is assigned to one of the classes of interest.

The paper is organized as follows. In the next section, a new *HSwC* approach is presented. Experimental results are discussed in Section 3. Finally, conclusions are drawn in Section 4.

2. PROPOSED CLASSIFICATION APPROACH

The flowchart of the proposed *HSwC* classification method is shown in Fig. 1. An input B-band hyperspectral image 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 information classes (i.e., has a class label L_j). The following procedure is proposed for this purpose:
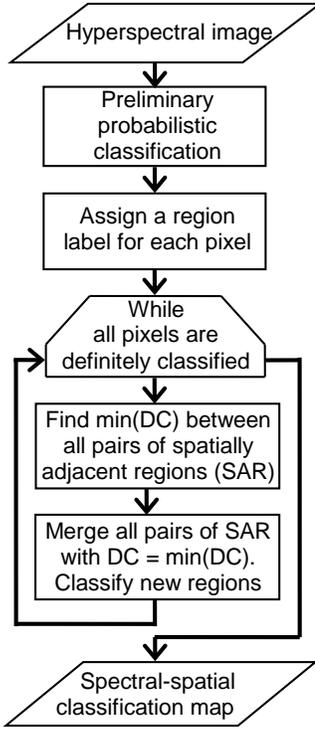
Fig. 1. Flowchart of the proposed *HSwC* classification method. "DC" means Dissimilarity Criterion.

2.1. Preliminary probabilistic classification

The first step consists in performing a preliminary probabilistic classification of the hyperspectral image. We propose to use a Support Vector Machines (SVM) classifier for this purpose which has shown a good performance for classifying hyperspectral data [3]. The standard SVM classifier does not provide probability estimates for the individual classes. In order to estimate these estimates, pairwise coupling of binary probabilistic estimates is applied [5]. This step results in a classification map $\mathbf{L}' = \{L'_j, j = 1, 2, ..., n\}$, where each pixel has a unique class label, and n vectors of class probabilities for each pixel $\{P(L_j = k|\mathbf{x}_j), k = 1, ..., K\}, j = 1, 2, ..., n$.

2.2. Hierarchical segmentation with integrated classification

At the next step, Hierarchical Step-Wise Optimization segmentation (HSWO) [8] with integrated classification is performed as follows:

1) Initialize the algorithm by assigning a new region label for each pixel. Each one-pixel region R_i gets a preliminary class label $L(R_i)$ and a vector of class probabilities $\{P_k(R_i) = P(L(R_i) = k|R_i), k = 1, ..., K\}$.
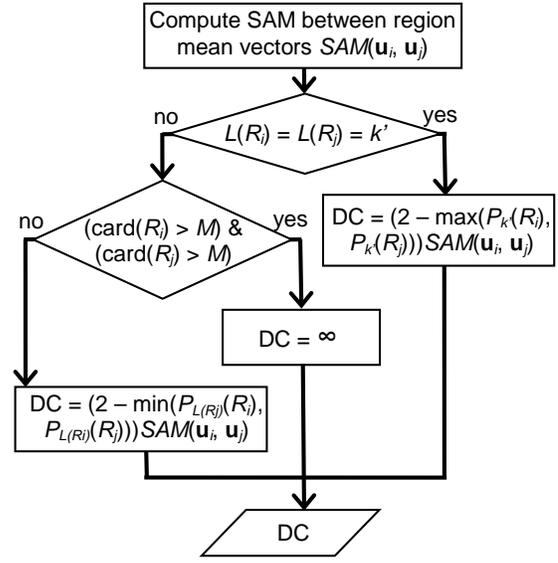
Fig. 2. Flowchart of computing a dissimilarity criterion between two regions R_i and R_j.

2) Calculate the DC between all pairs of spatially adjacent regions. We propose to compute a DC between two regions R_i and R_j as a function of region statistical (mean vector) and geometrical (number of pixels) features along with classification probabilities (see Fig. 2):

- Compute the Spectral Angle Mapper (SAM) measure between the region mean vectors $\mathbf{u}_i = (u_{i1}, ..., u_{iB})^T$ and $\mathbf{u}_j = (u_{j1}, ..., u_{jB})^T$ as

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

- If the regions have equal class labels $L(R_i) = L(R_j) = k'$,

$$DC(R_i, R_j) = (2 - \max(P_{k'}(R_i), P_{k'}(R_j)))SAM(\mathbf{u}_i, \mathbf{u}_j). \tag{2}$$

- Otherwise (i.e., if $L(R_i) \neq L(R_j)$), if a number of pixels in each region is larger than M, $DC(R_i, R_j) = \infty$ (in practice, the upper maximum value of *float*). This condition is included for favoring merging small regions. We propose to set $M = 20$.

- Otherwise,

$$DC(R_i, R_j) =$$
$$(2 - \min(P_{L(R_j)}(R_i), P_{L(R_i)}(R_j)))SAM(\mathbf{u}_i, \mathbf{u}_j). \tag{3}$$

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

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

- A vector of new class probabilities as

$$P_k(R_{new}) = \frac{P_k(R_i)card(R_i) + P_k(R_j)card(R_j)}{card(R_{new})}, \quad (4)$$

$k = 1, ..., K$, where $card(R)$ is a number of pixels in the region R, $card(R_{new}) = card(R_i) + card(R_j)$.

- Class label as

$$L(R_{new}) = \arg \max_k \{P_k(R_{new})\}. \quad (5)$$

All the pixels in the new region R_{new} get a definite class label (i.e., if R_i and/or R_j were composed of one pixel with a preliminary classification label, the corresponding pixels get a definite label of the class).

5) Stop if all the pixels (n pixels) get a definite classification label. If the considered image may have one-pixel regions of interest, the algorithm is converged when $P \cdot n$ pixels $(0 < P < 1)$ are definitely classified. If the algorithm is not converged, recalculate the DC values for the new regions and all regions spatially adjacent to them, and return to step 3.

3. EXPERIMENTAL RESULTS AND DISCUSSION

We applied the proposed method to the *Indian Pines* image recorded by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor over the vegetation area in Northwestern Indiana. The image is of 145 by 145 pixels, with a spatial resolution of 20 m/pixel. Twenty water absorption bands have been removed, and a 200-band image was used for experiments. Sixteen classes of interest are considered, which are detailed in Table 1, with a number of training and test samples for each class. Training samples have been randomly selected from the reference data. More information about the image and reference data can be found in [6].

A probabilistic one-versus-one SVM classification with the Gaussian Radial Basis Function (RBF) kernel was performed. The optimal parameters were chosen by fivefold cross validation: penalty during the SVM optimization $C = 128$, spread of the RBF kernel $\gamma = 2^{-6}$. Classification map is shown in Fig. 3(a). Then, the proposed *HSwC* method was applied (the algorithm was implemented using the Hierarchical SEGmentation (HSeg) software [9]). The algorithm has been converged when: a) 99% of pixels were classified ($P = 0.99$); b) all pixels were classified ($P = 1$).

Table 1 summarizes global (overall, average accuracies and kappa coefficient [6]) and class-specific accuracies of the SVM classification and the proposed *HSwC* technique (with $P = 0.99$ and $P = 1$). In order to compare the efficiency of the proposed method with previously proposed techniques, we have included results of the ECHO classification [10], a classification using the construction of a minimum
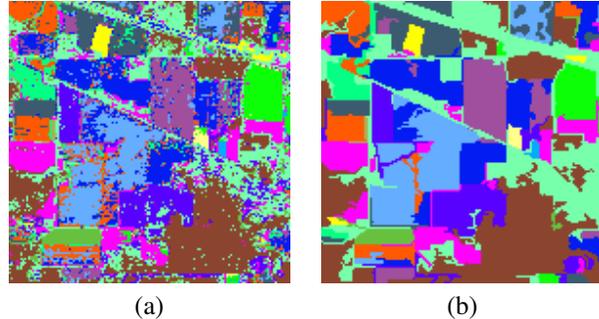
Fig. 3. *Indian Pines* image. (a) SVM classification map. (b) *HSwC* classification map ($P = 1$).

spanning forest from the probabilistic SVM-derived markers (*SVMMSF*) [5] and a classification by majority voting within neighborhoods defined by HSeg segmentation (*HSeg+MV*, with the parameter $S_{wght} = 0.0$, which is equivalent to HSWO, and the SAM DC) [6].

As can be seen from Table 1, accuracies are significantly improved when including spatial information into the classification. The proposed *HSwC* method, with $P = 1$, yields the best average accuracy (which is improved by 8.2 percentage points when compared to the SVM classification) and most of the class-specific accuracies. Fig. 3(b) shows the corresponding classification map, which is much less noisy than a pixel-wise classification map. The *HSwC* algorithm with $P = 0.99$ yields lower classification accuracies. Therefore, unless there are one-pixel regions of interest in the image under consideration, we recommend to run the algorithm until all the pixels get a definite classification label (i.e., $P = 1$).

The best overall accuracy is obtained using the *HSeg+MV* technique. However, in this method a segmentation map at an appropriate level of segmentation detail was chosen by quantitatively evaluating classification accuracies at several hierarchical levels versus the test data and retaining the best results. The test data are not available in most cases for classifying a new data set. Furthermore, the *HSwC* approach proposed in this paper requires less parameters to tune when compared to the *ECHO* and *SVMMSF* techniques. The only parameters to be tuned are C and γ for the SVM classification, a parameter M for favoring merging small regions (which is intuitive to tune), and a parameter P if required.

4. CONCLUSIONS

Hyperspectral imagery is a well-suited technology for accurate image classification, which is an important task in many application sectors (such as precision agriculture, monitoring and management of the environment, etc.). A new *HSwC* method for spectral-spatial classification of hyperspectral data was presented in this paper. In this method, a preliminary probabilistic classification is first applied, using an SVM clas-

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

	No. of Samp.		SVM	ECHO	SVM MSF	HSeg +MV	HSwC	
	Train	Test					$P = 0.99$	$P = 1$
Overall Accuracy	-	-	78.17	82.64	88.41	**90.86**	87.95	89.24
Average Accuracy	-	-	85.97	83.75	91.57	93.96	92.91	**94.18**
Kappa Coefficient κ	-	-	75.33	80.38	86.71	**89.56**	86.31	87.76
Corn-no till	50	1384	78.18	83.45	90.97	90.46	92.20	**93.06**
Corn-min till	50	784	69.64	75.13	69.52	**83.04**	81.63	82.53
Corn	50	184	91.85	92.39	95.65	95.65	95.65	**97.28**
Soybeans-no till	50	918	82.03	90.10	**98.04**	92.06	94.44	95.10
Soybeans-min till	50	2418	58.95	64.14	81.97	**84.04**	72.54	74.36
Soybeans-clean till	50	564	87.94	89.89	85.99	95.39	93.62	**96.10**
Alfalfa	15	39	74.36	48.72	94.87	92.31	92.31	**97.44**
Grass/pasture	50	447	92.17	94.18	**94.63**	94.41	91.28	93.96
Grass/trees	50	697	91.68	96.27	92.40	97.56	**97.85**	**97.85**
Grass/pasture-mowed	15	11	**100**	36.36	**100**	**100**	**100**	**100**
Hay-windrowed	50	439	97.72	97.72	**99.77**	99.54	98.86	98.86
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	98.39	**99.52**
Bldg-Grass-Tree-Drives	50	330	61.52	81.52	68.79	**82.12**	78.48	81.52
Stone-steel towers	50	45	97.78	97.78	95.56	**100**	**100**	**100**

sifier. Then, iterative best merge region growing with integrated classification is performed. The main contribution consists in defining a DC between adjacent regions as a function of region statistical and geometrical features and classification probabilities. Experimental results did show that the proposed approach improves classification accuracies, when compared to previously proposed classification techniques, and provides classification maps with more homogeneous regions. Similar results are obtained for other datasets acquired by the ROSIS (Reflective Optics System Imaging Spectrometer) sensor.

In conclusion, the proposed approach succeeded in taking advantage of the spatial and spectral information simultaneously for accurate hyperspectral data classification.

5. REFERENCES

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

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

[3] 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.

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

[5] 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.

[6] 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.

[7] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosc. and Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sept. 2006.

[8] 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.

[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] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geoscience Electronics*, vol. 14, no. 1, pp. 19–26, Jan. 1976.