A Graph Based Classification Method for Hyperspectral Images

Jun Bai, Shiming Xiang, and Chunhong Pan
Institute of Automation, Chinese Academy of Sciences, Beijing

***Abstract*—The goal of this paper is to apply graph cut (GC) theory to the classification of hyperspectral remote sensing images. The task is formulated as a labeling problem on Markov Random Field (MRF) constructed on the image grid, and graph cut algorithm is employed to solve this task. In general, a large number of user interactive strikes are necessary to obtain satisfactory segmentation results. Due to the spatial variability of spectral signatures, however, hyperspectral remote sensing images often contain many tiny regions. Labeling all these tiny regions usually needs expensive human labor. To overcome this difficulty, a pixel-wise fuzzy classification based on support vector machine (SVM) is first applied. As a result, only pixels with high probabilities are preserved as labeled ones. This generates a pseudo user strike map. This map is then employed for graph cut to evaluate the truthful likelihoods of class labels and propagate them to the MRF. To evaluate the robustness of our method, we have tested our method on small training sets. Additionally, comparisons are made between the results of SVM, SVM with stacking neighboring vectors, SVM with morphological pre-processing and our method. Comparative experimental results demonstrate the validity of our method.**

***Index Terms*—Hyperspecral, classification, graph cut, support vector machine, Markov Random Field.**

I. Introduction

Classification or segmentation of hyperspectral images is a perennial topic in remote sensing imaging and has received more and more research interest in the recent years. The simplest approach is based on the pixel-wise classification, where the whole spectral information is encapsulated as the sole feature. However, this approach manages the data not as an image but as a disarranged list of spectral signals, and the spatial correlation between the image pixels can not be reflected at all. As a result, the objects in the results through this approach mostly lack continuity and often contain many random noises at the boundaries.

The integration of spatial and spectral information is a research focus in hyperspectral classification. One simple and typical measurement is the vector stacking (VS) approach, where the feature vectors are selected as the concatenation of the pixels and their neighbors [1]. However, this concatenation also introduces the extra burden of increasing dimensions, making the original high dimensional classifications more intractable. For this reason, some measures of dimension reduction are necessarily carried out as the pre-processing steps. One advance for the integration of spatial and spectral information in the last years is the method based on mathematical morphology, which can generate good results. In [2] and [3], the morphological algorithm is successfully introduced to remote sensing imaging, proposing an entirely new approach for the classification of remote sensing images. However, in [4], the authors also point out that the spatial information utilized in this methodology is mainly the sizes of structures.

Graph cut is a graph based method related to the MRF approach using the thinking of semi-supervised learning. It is well studied by some researchers in the field of computer vision. This method aims to solve the well known metric labeling (ML) problem via a maxflow/mincut algorithm [5]. The ML problem maps an object set to a label set by the optimization of a minimum energy cost. Although it has a history of more than ten years, it is only when a primal-dual schema is proposed, in [6], [7], that this graph based ML approximation algorithm becomes really practical and robust to use.

This paper presents a methodology in the generalization of the ML graph cut algorithm from natural image processing to hyperspectral classification tasks. Our approach is developed in terms of supervised classification with a very small number of training samples. However, the good performances of graph cut algorithm are obtained with enough user specified strokes on the objects. To overcome this difficulty, we propose a two-step strategy: a pixel-wise fuzzy SVM classifier is first used and the pixels with high probabilities are preserved to simulate the user interactive result (We call it "pseudo map"). This pseudo map is then fed as input to the graph cut algorithm. In the second step, the graph based method can regulate the initial results, making fully use of the spatial relations. Our intention to use the fuzzy pixel-wise classification is just to find the reliable labeled samples. Other pixels are not considered to be reliable so that they are left as unclassified.

The remainder of this paper is organized as follows. Section II describes the fuzzy SVM algorithm. In section III we will give a detailed description of the primal-dual solution to the ML problem based on graph cut. The application of this algorithm will be presented in section IV and experimental results are reported in section V. In the final section we will give some conclusions.

II. The fuzzy SVM and its Probabilistic Output

Consider the two-class case. f is the output of general SVM and $y \in \{-1, 1\}$ is the label. What we are interested in here is the posterior probability $P(y = 1|f)$. To map the SVM outputs to posterior probabilities, an empirical sigmoid model

978-1-4577-1964-6/12/$31.00 ©2012 IEEE

is constructed instead of Bays formulation:

$$P(y=1|f) = \frac{1}{\exp(Af+B)} \tag{1}$$

The inner function $g = Af + B$ is a linear function of f. The unknown parameters are A and B. Given a training set $\{(x_i, y_i)\}$, let us define a new set $\{(f_i, t_i)\}$, where f_i is the output of standard SVM, and t_i is defined as:

$$t_i = \frac{y_i + 1}{2} \tag{2}$$

According to (1), we also define p_i:

$$p_i = \frac{1}{\exp(Af_i + B)} \tag{3}$$

This estimation of parameters finally turns out to be a minimization problem with a so called cross-entropy minimization function:

$$\min \sum_i t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \tag{4}$$

A remaining problem is how to generalize the probabilistic results from the case of two-class to multi-class. In [8], an optimization method is suggested based on the strategy of OAO. For a single sample x, let r_{ij} be the probability that the sample belongs to class i when considering only the two classes i and j, which can be obtained from the above discussion, and p_i is the final probability that we need to solve. It is obviously that we have $r_{ij} = 1 - r_{ji}$, and p_i satisfies the following constraint:

$$\sum_{i=1}^{k} p_i = 1 \tag{5}$$

With this constraint, the target function is:

$$\min_p \frac{1}{2} \sum_{i=1}^{k} \sum_{j:j \neq i} (r_{ji} p_i - r_{ij} p_j)^2 \tag{6}$$

In [9] it is reported that this probabilistic algorithm has similar accuracies when comparing to the traditional SVM. The advantage is that this estimation can predict the probability that a sample belongs to a certain class as well as the assigned class labels.

III. The Primal-dual ML Algorithm Based on Graph Cut

This section mainly discusses the definition and solution to the ML problem based on graph cut, which will be employed in our experiment specially to the classification of hyperspectral images.

A. The Description of the Problem

A general description of the ML problem is: given a set of objects V and a set of labels L, the goal is to find a labeling function $f : V \longrightarrow L$ so that the cost $C(f)$ can be minimized. Granted that E is a collection of pairs of objects that are connected, this cost can be expressed as:

$$C(f) = \sum_{p \in V} C_{p,f(p)} + \sum_{(p,q) \in E} w_{pq} d(f(p), f(q)) \tag{7}$$

The first term in (7) is the loss function of the classification result for single objects, where $C_{p,f(p)}$ is the label cost for each object. The second term, however, is regarding to the pairwise relationships among the objects, where w_{pq} is the weight between edges and $d(f(p), f(q))$ is a metric distance which stands for the separation cost when labeling the pairs of objects p and q with $f(p)$ and $f(q)$.

Before the solution to (7) is given, it is necessary to give some formulations for the terms to fit the labeling model, making fully use of the spatial characteristic of hyperspectral images. To describe these formulations clearly, a distance between the pixels is firstly to be defined.

For a pixel p, $\mathbf{x}_p$ is the data vector, and (p_x, p_y) is its spatial coordinate. To reflect the spatial information related to the pixels, a new vector is introduced by combining the data and the spatial coordinate: $z_p = (\mathbf{x}_p^T, p_x, p_y)^T$. The data vector $\mathbf{x}_p$ and the spatial coordinate are both normalized, with all their components within $[-1, 1]$. Therefore, we have $d(p, q) = d(z_p, z_q)$. When the traditional Euclid distance is used, this distance can be formulated as:

$$d(p,q) = \sqrt{\|\mathbf{x}_p - \mathbf{x}_q\|^2 + (p_x - q_x)^2 + (p_y - q_y)^2} \tag{8}$$

The detailed formulations in (7) can be expressed as follows.

$C_{p,f(p)}$: In this work, we first cluster training samples of each class to a few subsets and then calculate the distances between the pixel and the center of each clustered subset. The cost function is thus defined as the minimum distance. Let $l \in L$ be the label and the training samples with this label are clustered to n_l groups by the algorithm of K-means. The cluster center of the ith group is denoted as $c_{l,i}$, $i = 1, \cdots n_l$. Contrast with (7), the assigned label is $f(p)$. That is $l = f(p)$. Then the first item in (7) can be expressed as:

$$C_{p,f(p)} = \min_{i=1}^{n_{f(p)}} d(p, c_{f(p),i}) \tag{9}$$

$d(f(p), f(q))$ is a metric distance that measures the labeling smoothness cost. For the sake of simplicity, we define it as:

$$d(f(p), f(q)) = \begin{cases} 0, & f(p) = f(q) \\ 1, & f(p) \neq f(q) \end{cases} \tag{10}$$

It is easy to prove that the distance defined by (10) is a metric distance.

w_{pq}: In the graph $G = (V, E)$, this parameter can be interpreted as the weight of edges that connect the vertexes in the edge set E. We use this distance between the two feature vectors to express this similarity. The expression of our formulation is as (11):

$$w_{pq} = \begin{cases} \dfrac{1}{\varepsilon + d(p,q)} & p,q \text{ are neighbors} \\ 0 & \text{others} \end{cases} \tag{11}$$

In (11) a constant value ε is added to the denominator just for the purpose of avoiding a zero divisor.

B. Solution

(7) has been considered as a NP-hard problem and several works have explored the solution to this problem. Specifically to the purpose of our application, we adopt the method proposed by Komodakis in [6], which is based on graph cut as well, and gives a primal-duel solution. In this paper we only give a brief description of this algorithm. Readers may read [6] for the details.

Like other work, this ML problem is first formulated as an integer linear one. However, the novel contribution of this work is to change the linear programm to its dual form by relaxing the constraints, and the dual variables are called balance variables. An iterative primal-dual scheme is then carried out based on the principal of relaxed complementary slackness. The iteration steps consist of two layers of iterations called the inner iteration and the outer iteration. In each inner iteration, a label c is fixed and only the balance variables of the c labels are modified. This is called a c-iteration and the primal-dual pairs of solutions are updated in this step. All the c-iterations make up the outer iteration and after this iteration, the approximate optimal solution can be reached. Our experiment has testified the convergence is very fast. It only takes a few outer iteration steps and quite a little computing time.

IV. Applications in Hyperspectral Classification

In this work, we introduce a two-step strategy for classification. With the limited training set, this image is first processed by a pixel-wise fuzzy SVM classifier. In this fuzzy output, only a sub set of the resultant pixels with the probabilistic value above a certain threshold are retained and others are deserved as unknowns. We call this pseudo map because the classified pixels in it are treated as if they were labeled manually and act as the training samples for the next step. As is discussed in section IV, the graph cut based classification is carried out, using this pseudo map as the training and the original image.

A remaining problem is how to decide the probability threshold. In our experience, this threshold is controlled to allow about $1/2 \sim 2/3$ of the pixels can pass this selection. These reserved samples can then be used as training for the purpose of further classification based on graph cut.

V. Experimental Results

The data set, Pavia center, in our experiment is provided by the HySens project, operated by the Deutschen Zentrum fur Luft-und Raumfahrt (DLR, the German Aerospace Agency). These data are from the ROSIS-3 optical sensor. The spatial resolution of about $1.3m$ per pixel is acquired. The number of bands is 102. This data set has been atmospherically corrected.

TABLE I
Information about the data set of Pavia Center

Label	Name	Training	Testing	Color
1	Water	824	65971	0,0,255
2	Asphalt	816	9248	192,192,192
3	Trees	820	7598	0,128,0
4	Shadow	476	2863	255,255,0
5	Meadows	824	3090	0,255,0
6	Bare Soil	820	6584	184,92,0
7	Tiles	1260	42826	255,102,0
8	Bricks	808	2685	255,0,0
9	Bitumen	808	7287	0,255,255

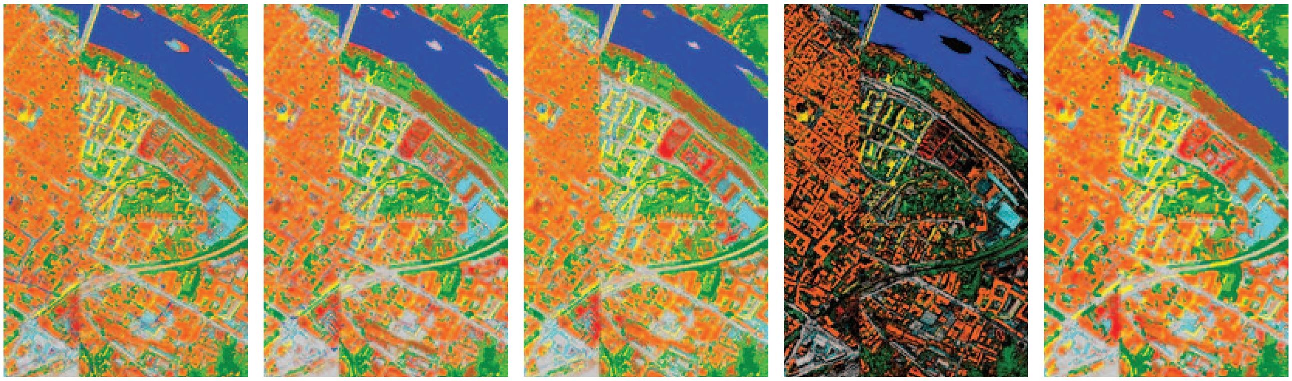
This data set contains 1096×1096 pixels, with a 381 wide-pixel black strip. Similar with the experiments in [4], we remove this strip and get an image with 1096×715 pixels. This image consists of nine classes. Table I presents the list of the nine classes including the training and testing samples, and the labels and their corresponding colors used to illustrate our classification results.

To apply our method, a fuzzy SVM classification with RBF kernel is first performed with the training set in this image. For more challenges, we randomly selected 15 training samples for each class from the training set in Table I. In this experiment, we use the spectral vector as the feature without any dimension reduction. According to our method, we select 0.55 as the truncation threshold. 404233 out of 783640 pixels are left unclassified. Then this pseudo map is used as the training for further processing via the algorithm based on graph cut. For more details, the number of clusters for each class is set to 20. Our experiments have also testified that there exist almost no differences in the case of more cluster.

For comparisons, three methods are selected as the baselines besides ours (SVM/GC). One is SVM with the spectral features (SVM). Another is the vector stacking approach, followed by an SVM classifier (VS/SVM). In the third one, the feature is obtained with EMP pre-processing and then also fed to a SVM classifier (EMP/SVM). All the SVM parameters are automatically tuned to achieve best results via cross-validation. For reference, the middle result, which we call pseudo map, is also presented in the figures.

Besides the accuracies for each class, an overall accuracy (OA) is also listed, calculated by the number of correctly classified samples divided by the number of test samples. For more comparisons, a kappa coefficient is presented as well [10].

By the comparisons in Fig. 1, the positive effect of spatial/spectral integration can easily be seen. The result of SVM lacks continuity for many objects. There exist many discrete pixels, which is contradictory to common human sense. This is mainly because in the pixel-wise classification, the spatial correlations between the pixels are not considered at all. In the



SVM VS/SVM EMP/SVM Pseudo map SVM/GC

Fig. 1. Experimental results of Pavia center with the reduced training set, 15 samples for each class. From left to right: (a) SVM; (b) VS/SVM; (c) EMP/SVM; (d) Pseudo map with T=0.55; and (e) Our SVM/GC.

TABLE II
ACCURACY OF THE PAVIA CENTER IN PERCENTAGE WITH REDUCED TRAINING SAMPLES, 15 SAMPLES FOR EACH CLASS

Class 1	96.57	98.31	99.17	**100**
Class 2	93.09	**97.85**	97.76	95.80
Class 3	81.93	79.21	83.43	**85.14**
Class 4	**99.83**	99.51	99.62	97.07
Class 5	88.35	88.77	**94.47**	92.91
Class 6	88.68	86.85	**99.04**	93.33
Class 7	97.39	98.14	98.28	**98.57**
Class 8	60.97	77.73	**88.19**	81.97
Class 9	78.02	88.76	82.83	**91.27**
OA	93.82	95.73	96.92	**97.31**
κ	91.31	93.99	95.65	**96.19**

results of VS/SVM, EMP/SVM and our SVM/GC method, more continuities can be seen for most of the structures. This benefits greatly from to the integration of spatial and spectral information. Another advantage of our method is that, from the figures, the sharp corners of the objects can be seen clearly, benefiting from the idea of MRF with minimal spatial smoothing effect.

The accuracies reported in Table II can validate our method. The SVM results exhibit relative low accuracies. This is mainly due to the absence of spatial information. VS/SVM has considered the neighboring information of pixels. However, the simple stacking vectors also reduce the separability between the classes, resulting in the reported accuracies even lower than the spectral SVM approach. The EMP approach performs very well because the structural characteristic of this data set is rather dominant.

VI. CONCLUSION

The goal of this paper is to investigate the application of a graph based method in hyperspectral classification. A two-step strategy for classification is proposed, including a fuzzy SVM classifier and the graph cut based classification. The main advantage of our algorithm is to make fully use of the spatial information to achieve fine results. The adoption of SVM classifier also enables the robustness for the small training sets. Experiments have demonstrated the satisfactory results of our method.

ACKNOWLEDGMENT

This work is funded by the Sustainable Urban Metabolism for Europe project (SUME, Grant Agreement No: 212034). The authors would like to thank Mr. J. A. Benediktsson for his contribution and the University of Pavia and the HySenS project for providing the data.

REFERENCES

[1] C. Yangchi, M. M. Crawford, and J. Ghosh, "Knowledge based stacking of hyperspectral data for land cover classification," in *IEEE Symposium on Computational Intelligence and Data Mining, 2007. CIDM 2007.*, 2007, pp. 316–322.

[2] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 309–320, 2001.

[3] J. A. Benediktsson, K. Arnason, and M. Pesaresi, "The use of morphological profiles in classification of data from urban areas," in *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 2001, pp. 30–34.

[4] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[5] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[6] N. Komodakis and G. Tziritas, "A new framework for approximate labeling via graph cuts," in *Tenth IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 2005, pp. 1018–1025.

[7] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, 2007.

[8] C. Chang and C. Lin, *LIBSVM: a library for support vector machines*, 2001.

[9] C. Chih-Chung and L. Chih-Jen, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[10] J. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 2005.