SEMANTIC SEGMENTATION OF HYPERSPECTRAL IMAGES WITH THE FUSION OF LIDAR DATA

Hakan AYTAYLAN[1,2] and Seniha Esen YUKSEL[1]

[1]Department of Electrical and Electronics Engineering, Hacettepe University, Ankara, Turkey
[2]ASELSAN Inc., Ankara, Turkey
Contact: haytaylan@aselsan.com.tr, eyuksel@ee.hacettepe.edu.tr

ABSTRACT

Semantic segmentation is an emerging field in the computer vision community where one can segment and label an object all at once. In this paper, we propose a semantic segmentation algorithm that takes into account both the hyperspectral images and the LiDAR data. In our segmentation framework, we propose a new energy function that is composed of two terms: a unary energy term and a pairwise energy term. The unary energy term provides the segmentation maps for the hyperspectral data as well as for the LiDAR data which is explained with Fisher Vectors. The pairwise spatial term uses both the UTM coordinates as well as the LiDAR data. Finally, the system is solved with graph-cuts. We report the effect of the parameters in energy minimization and show that the best results are achieved with an SVM-MRF classifier among the several classifiers.

Index Terms— Semantic Segmentation, Hyperspectral image classification, LiDAR, Graph Cuts, Fisher Vectors.

1. INTRODUCTION

Joint segmentation and classification of an image is called semantic segmentation, and it is an active research topic in the computer vision field [1] [2] [3]. This topic is very much in parallel to the interests of the hyperspectral community, who are interested in spectral-spatial segmentation, target identification and fusion.
Hyperspectral imagery technology, by capturing information from hundreds of frequencies of light, provides valuable information on the material of the subject. The use of this information on remote sensing has been a highlighted topic of research in the recent years. Even though the information on hundreds of spectral bands of a surface can be stored within one pixel [4], factors such as atmospheric effects, changing light conditions, especially in the urban areas may cause misclassification and decrease the overall performance [5]. To overcome these negative effects, light detection and ranging (LiDAR) data is commonly fused together with hyperspectral data to overcome the weaknesses of both. LiDAR data, containing three dimensional spatial information of a scene, has been proven to be an excellent candidate to work together with the hyperspectral data [5].

The goal of our study is the fusion of LiDAR and hyperspectral datasets with the purpose of the semantic segmentation of hyperspectral images. In this study, we propose a novel unary term that includes both spectral attributes and LiDAR information, which is described by Fisher Vectors [6]. Also, we propose a pairwise term that considers both the spatial distance in UTM coordinates and the spectral distance of two pixels. Finally, this energy function is optimized with the Graph Cut algorithm [7], [8].

2. RELATED WORK

There are several studies that make use of graph-cuts methods using hyperspectral datasets alone. In [9], probabilistic SVM is used for spectral classification, and a spatial energy term is used for spatial information. In [10], SVM is used for spectral classification and subspace Multinomial Logistic Regression (MLRsub) [3] is used for spatial classification. Classifier results are fused together within a Markov Random Field (MRF) Model [11]. Since in both cases, probabilistic SVM and MLRSub are proven to be a good candidate for the unary term of the energy function, we also use these two classifiers for the spectral term in our experiments.

The fusion of LiDAR and hyperspectral data is exploited in [5], [16]. In [5], the authors' use morphological attributes [12] of both LiDAR and hyperspectral data and classify the data using MLRsub classifier and MRFs.

Different from these related approaches, instead of performing fusion during feature extraction, we combine the results of energy descriptors and jointly solve the label problem. With this approach, we guide the spectral classification results to be consistent with the spatial properties, including elevation. In doing so, the fusion task is seamlessly integrated into the energy model for semantic segmentation.

3. PROPOSED METHOD

Semantic segmentation task is defined as assigning each pixel in an image to a class from a previously determined class set $C = \{ C_1, C_2, C_3 ... C_n \}$. To achieve this, we use the

Markov Random Field (MRF) assumption, and define the maximum a posteriori (MAP) labeling of the random field as:

$$y^* = \underset{y \in C}{\arg\max}\, P(Y = y \mid x) = \underset{y \in C}{\arg\min}\, E(y)$$

where **x** is the observed data vector $x = \{x_1,\ x_2,\ x_3 \ldots x_k\}$. and y is the assigned class. With Z being a normalization constant, the energy function

$$E(y) = -\ln(P(Y = y \mid X = x)) - \ln(Z)$$

can be written as two terms:

$$E(y) = E_{unary}(y) + E_{pairwise}(y)$$

The unary term of the energy function is the probabilistic classification of every pixel to a certain class. In our experiments, we have experimented both with the probabilistic SVM classifier [13] and also the MLRSub. Other classifiers may also be used for this term, however we chose SVM and MLRSub due to their proven track record [5] [9] for this task.

In the unary term, we consider two separate unary probabilistic distributions, namely $P(C_i \mid x_i)$ for spectral and $P(C_i \mid s_i)$ for LiDAR data. Here x_i are the hyperspectral data; and s_i are the intensity and elevation information, extracted from LiDAR data, which is expanded using Fisher Vectors. Therefore, our proposed unary probability distribution energy is as follows:

$$E_{unary} = -k\left[\alpha \ln(P(C_i \mid x_i)) + (1-\alpha)\ln(P(C_i \mid s_i))\right] \quad (1)$$

where k is a constant term for emphasizing the weight of the unary term with respect to the pairwise term, and α is a constant that is used to adjust the weights between the two probability distributions. In the rest of this paper, we will refer to the unary energy term as *dSVM* or *dMLRsub* depending on which function we use as the classifier.

The pairwise term is concerned about the spatial relation of the pixel with its neighboring pixels. The standard energy expression for this term is:

$$E_{pairwise}(x_i, x_j) = \sum_{x_j \in N_i} \beta(1 - \delta(C_i, C_j))$$

where, N_i is the set of neighboring pixels, δ is the Kronecker delta function and β is a positive constant which determines the weight of contribution of the pairwise data to the overall energy.

The pairwise term acts as a regularization term. As it is usually accepted that neighboring pixels should have higher probability of having the same label, this term penalizes two neighboring pixels that have different class labels. We assume that neighboring pixels having different elevations should have lower probability of being assigned to the same label. For example, asphalt on a rooftop or asphalt on the ground should be assigned to different segmentations. LiDAR elevation information might help classify correctly the pixels that have different elevation. Hence, we have modified the standard pairwise energy function to penalize neighboring pixel pairs having their difference of elevations. Our proposed energy function is as follows:

$$E_{pairwise}(x_i, x_j) = \sum_{x_j \in N_i}(1 - \delta(C_i, C_j))\begin{pmatrix} \beta\exp(-d(x_i, x_j)) \\ + \eta\exp(-d(h_i, h_j)) \end{pmatrix} \quad (2)$$

Here, $d(h_i, h_j)$ provides the 3D spatial distance of two neighbors, where h_i are 3 band instances containing UTM coordinates and elevation of the pixel. In implementation, absolute UTM coordinates are not actually necessary, so we simply use the resolution of the pixels and the height. Hence for each pixel, the following rule applies for its coordinate in 3D space:

$$U(p) = [\,\epsilon r_p \ \epsilon c_p \ h_p\,]$$

where p is the pixel in focus, r_p is the row, c_p is the column of the pixel p in the image, h_p is the height and ϵ is the pixel resolution of the sensor. In our case, $\epsilon = 2.5$. $d(x_i, x_j)$ on the other hand, is the spectral distance. Tarabalka and Rana [6] consider 3 different alternatives for this distance; such as the SAM, SID and L2-norm. Also, η is a user defined constant which adjusts the weight between these two distance functions.

4. DATASET AND EXPERIMENTS

We are using the University of Houston dataset provided by the 2013 GRSS Data Fusion Contest [14]. The dataset has both LiDAR and hyperspectral data, contains 15 classes and around 200 training and 1000 test pixels for each class. There are 144 hyperspectral bands in the data.

Our workflow is as follows: first, LiDAR intensity and elevation information is extracted and simply concatenated, generating a 4 band data which consists of first, average, last returns of the LiDAR intensity together with the elevation. We then obtain local descriptors from this data using Fisher Vectors. Both Lidar and Hyperspectral datasets are then processed with the HySime algorithm [15]. The training and validation data is used to train probabilistic SVM and MLRSub classifiers separately. We used the method proposed in [3] for the implementation of MLRSub classifier. The impact of α in (1) on overall performance can be viewed in Figure 1.

For the pairwise term; our experiments showed that SAM proved to be the best distance calculation method for both MLRSub and SVM trained unary distributions. The other methods were also not too far behind but since their comparison is not the scope of this study, we will not be demonstrating their differences. The β and η weight terms for the pairwise data is experimentally detected and their impact

on overall performance is also recorded. Unfortunately, in our proposed method, β and η in (2) did not show a remarkable performance improvement. There is always a larger β which leads to same accuracies when η = 0. This can be observed in Figure 2.
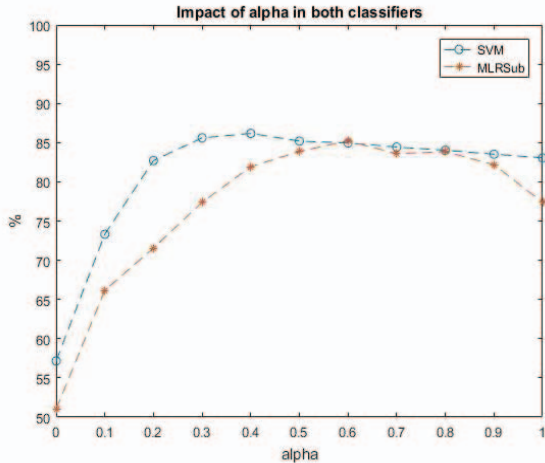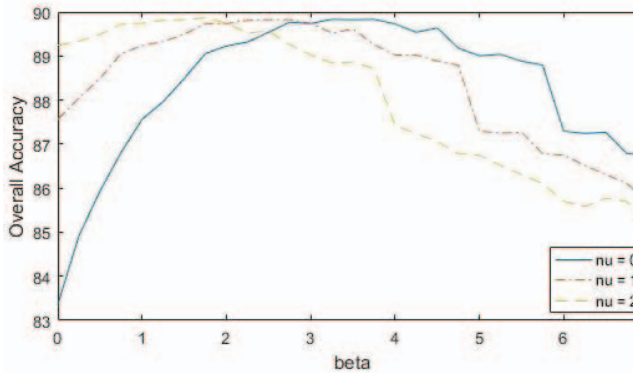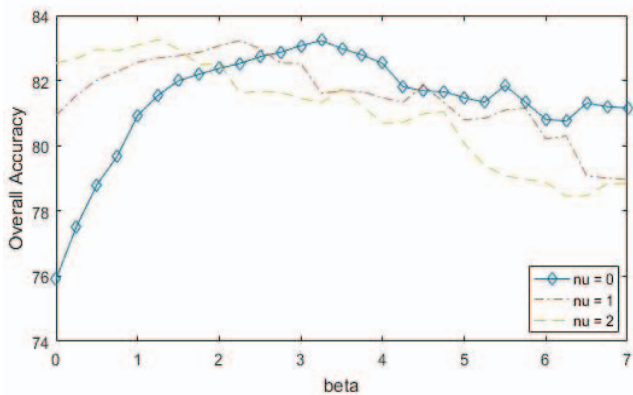

Figure 1: Impact of alpha. The best value of alpha found for SVM is α = 0.4 and for MLRSub, α = 0.6

(a)

(b)

Figure 2:. The Impact of β and η (a) SVM, (b) MLRSub

We also compared our results to the classical SVM-MRF and MLRSub-MRF methods in Figure 3. We observed that, dSVM-MRF outperformed other methods. The dMLRSub-MRF method however showed the most drastic improvement from its unary stage. Note that there is a cloud that covers most of the commercial, railway and highway class test pixels. The dSVM and dMLRSub performed much better in classifying these against other methods. Fisher Vectors improved the LiDAR unary classification performance in MLRSub classifier but did not perform better in SVM classifier in our experiments; however, we believe in our future studies, we can optimize for higher accuracies.

5. CONCLUSION

In this paper, we introduced a novel semantic segmentation approach, where we can both segment and label pixels at the same time. We introduced an MRF based approach that combines the spectral information of hyperspectral image and the UTM coordinate, elevation information and intensity information of the LiDAR data during energy minimization. Also, we used Fisher Vectors in order to improve the unary performance of the classified LiDAR data. Although the optimization in pairwise term did not lead much better performance than classical MRF methods, we concluded that fusion of two unary terms improve the overall accuracy. In our future studies, we wish to improve the performance of this term as well as the pairwise term.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] K. K. Evans and A. Treisman, "Perception of Objects in Natural Scenes: Is It Really Attention Free?," *Journal of Experimental Psychology: Human Perception and Performance,* pp. 1476 - 1492, 2005.

[2] P. Arbelaez, B. Hariharan, G. Chunhui, S. Gupta, L. Bourdev and J. Malik, "Semantic Segmentation Using Regions and Parts," in *IEEE Conf. on CVPR*, 2012.

[3] J. Li, J. M. Bioucas-Dias and A. Plaza, "Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields," *IEEE Trans. on Geosc. and Remote Sens.,* vol. 50, no. 3, pp. 809 - 823, March 2012.

[4] J. A. Richards and Jia Xiuping, Remote Sensing Digital Image Analysis, Berlin: Springer, 2006.

	SVM-Spectral	MLRsb-Spectral	SVM-LiDAR	MLRSb-LiDAR	dSVM (α = 0.4)	dMLRSb (α = 0.6)	SVM-MRF	MLRsb-MRF	dSVM-MRF	dMLRsb-MRF
healthy gras	82.431	78.822	54.131	75.594	82.716	78.822	**83.096**	79.012	**83.096**	83.001
stressd grass	83.271	75.752	36.466	51.41	85.432	74.342	84.117	84.962	**91.353**	81.861
synth grass	98.416	98.218	98.02	88.713	99.208	97.822	99.208	99.406	**99.802**	99.01
tree	92.424	91.667	68.087	86.837	94.413	98.295	93.277	93.277	97.254	**99.716**
soil	98.485	97.159	61.269	11.364	97.633	96.307	99.905	99.811	***100***	99.053
water	88.112	79.021	21.678	1.3986	80.42	82.517	88.811	**91.608**	90.909	**91.608**
residential	80.877	81.25	55.877	42.351	80.224	79.571	84.981	85.821	**90.951**	87.407
commercial	54.131	36.752	65.812	6.4577	77.303	59.544	45.394	24.881	**78.158**	69.136
road	74.882	65.534	14.731	1.983	79.037	59.773	88.48	85.175	**94.618**	85.552
highway	60.811	56.371	23.938	3.2819	57.625	65.637	65.927	41.313	66.216	**74.324**
railway	90.987	64.421	52.941	57.97	83.681	65.655	82.543	76.565	**93.548**	57.685
parking lot 1	79.155	34.006	22.959	0.0961	82.229	61.095	92.315	62.248	**93.66**	76.081
parking lot 2	65.614	47.368	63.509	63.86	74.386	61.404	75.439	69.825	**87.719**	79.649
tennis court	97.571	95.951	82.591	67.611	97.976	95.142	98.381	97.976	**98.381**	**98.381**
running trck	95.56	95.56	72.727	10.994	93.658	93.023	99.366	97.252	96.617	**99.577**
OA	81.299	70.772	49.783	36.238	83.357	75.912	83.676	76.051	**89.866**	83.275
κ	0.798	0.684	0.460	0.312	0.820	0.739	0.824	0.740	**0.890**	0.819

Figure 3 - Table of Overall Accuracies and Kappa Values. Best results are bolded.

[5] M. Khodadadzadeh, J. Li, S. Prasad and A. Plaza, "Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning," *IEEE Journal of Sel. Topics in Appl. Earth Obs. And Remote Sens.,* vol. 8, no. 6, pp. 2971 - 2982, 2015.

[6] F. Perronnin and C. Dance, "Fisher Kernels on Visual Vocabularies for Image Categorization," in *IEEE Conf. on CVPR*, 2007.

[7] Y. Boykov, O. Veksler and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts".

[8] Y. Tarabalka and A. Rana, "Graph-Cut-Based Model for Spectral-Spatial Classification of Hyperspectral Images," in *IEEE IGARSS*, Quebec, Jul 2014.

[9] Y. Tarabalka, M. Fauvel, J. Chanussot and J. A. Benediktsson, "SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images," *IEEE Geosc. and Remote Sens. Letters,* vol. 7, no. 4, pp. 736-740, 2010.

[10] M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian and J. M. Bioucas-Dias, "Spectral-Spatial Classification for Hyperspectral Data Using SVM and Subspace MLR," *IEEE Trans. Geosci. Remote Sens.,* vol. 48, no. 10, pp. 809-823, Oct 2010.

[11] G. R. Cross and A. K. Jain, "Markov Random Field Texture Models," *IEEE Trans. on Pattern Analysis and Machine Intelligence, v*ol. 5, no. 1, pp. 25-39, 1983.

[12] M. M. Dalla, J. A. Benediktsson, B. Waske and L. Bruzzone, "Morphological Attribute Profiles for the analysis of Very High Resolution Images," *IEEE Tran. on Geosc. and Rem. Sens.,* vol. 48, no. 10, pp. 3747-3762, 2010.

[13] C. C. Chang and C. -J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology,* 2011.

[14] "2013 IEEE GRSS Data Fusion Contest," [Online]. Available: http://www.grss-ieee.org/community/technical-committees/data-fusion/.

[15] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral Subspace Identification," *IEEE Tran. on Geosc. and Rem. Sens.,* vol. 46, no. 8, pp. 2435 - 2445, 2008.

[16] C. Debes et al., "Hyperspectral and LiDAR Data Fusion: Outcome of the 2013 GRSS Data Fusion Contest," in *IEEE Jour. of Sel. Top. in Appl. Earth Obs. and Rem. Sens.,* vol. 7, no. 6, pp. 2405-2418, 2014.