AUTOMATIC FUSION AND CLASSIFICATION OF HYPERSPECTRAL AND LIDAR DATA USING RANDOM FORESTS

Andreas Merentitis, Christian Debes, Roel Heremans and Nikolaos Frangiadakis

AGT International
64295 Darmstadt, Germany
E-mail: {cdebes, amerentitis, rheremans, nfrangiadakis}@agtinternational.com

ABSTRACT

In this paper we discuss the use of the random forest algorithm for automatic fusion and classification of hyperspectral and LiDAR data. We demonstrate how relative feature relevance can be used in random forests to perform automatic and unsupervised feature selection. This allows using a large number of features without suffering from the curse of dimensionality. The effectiveness of the proposed approach is demonstrated on two datasets. The first dataset features a combination of hyperspectral and LiDAR data for urban classification whereas the second dataset is the well-known Indian Pines dataset featuring pure hyperspectral imagery. We show that by using the proposed approach classification accuracies can be improved significantly.

1. INTRODUCTION

Recently, the fusion of hyperspectral and LiDAR data has shown to bring significant improvement in classification of airborne imagery [1]. While hyperspectral data traditionally is used to obtain spectral fingerprints that give insight to the materials contained in a scene, LiDAR data provides topological information. Thus, these two sources can be fused in a complementary way to e.g., distinguish between streets and flat commercial buildings that can have a similar hyperspectral profile but differ in height.

In this paper we follow the framework proposed in [1],[2] for automatic fusion and classification of hyperspectral and LiDAR-derived imagery. In the core of this framework a supervised classification stream is implemented using random forests on features extracted from hyperspectral and LiDAR data. Further, in a post-processing step, reassignment of labels is done based on a Markov Random Field-based approach using class uncertainties.

The contribution of this paper is an extension of the random forest algorithm in [1], [2] which includes automatic feature selection. In the remainder of the paper, first a study on hyperspectral and LiDAR-derived features used for unsupervised segmentation is carried out. A large number of possible segmentation features are generated, but using all of them as inputs for the classifier is often not the optimal choice since they can be redundant. Moreover, using too many features can also lead to performance degradation due to the curse of dimensionality. As a mitigation to this challenge, we demonstrate how a relative feature relevance can be used in random forests to perform automatic and unsupervised feature selection. The segmentation features, their relevant ranking, and their effect on the final classification result are investigated in two different and challenging hyperspectral datasets. For both of them application of the proposed approach improves the outcome, in one case considerably, resulting on the state of the art performance for the given training/validation partitioning.

2. HYPERSPECTRAL FEATURES

In order to increase classification performance it is necessary to incorporate spatial information in the classification stream. This can be achieved if properties of the hyperspectral dataset that are effective in segmenting the different regions are extracted and provided as features to the classifier. In this paper we propose to provide this information to the classifier in the form of segmentation maps that can be derived by either the hyperspectral or LiDAR data, following a different approach for each of them.

In order to derive segmentation maps from the hyperspectral data, we adopt and extend an approach similar to the one proposed in [3]. The idea is to identify the most important edges in the image, where the spectral content changes abruptly. This is done by applying watershed segmentation, using one or more synthetic spectral features, which are extracted by standard techniques such as Principal Component Analysis (PCA), Minimum Noise Fraction (MNF), Maximum Autocorrelation Factor (MAF), etc. Either one dimensional or multidimensional features extracted from these techniques can be used, taking into consideration that in the multidimensional case we need either a way to select or combine the segmentation maps.

The watershed transform considers an image as a topographic relief. More specifically, when the transform is used

for segmentation, the value of a pixel is interpreted as its elevation and the watershed lines divide the image into regions (catchment basins), in a way that every region is associated with one minimum in the image. The watershed transformation is typically applied on the image gradient, which defines transitions between regions, so that it is assigned high values on the borders between objects and low values on the homogeneous parts [3]. When applied without additional constraints, the watershed transform often results in over-segmentation of the image, since a region is defined for every local minimum of the gradient. This effect is typically alleviated using filtering or post-processing methods [4] (in this paper we use morphology transformations in order to merge similar regions). An example segmentation map derived based on the first MNF component, using also morphological operations to alleviate the over-segmentation problem is depicted in Fig.1.

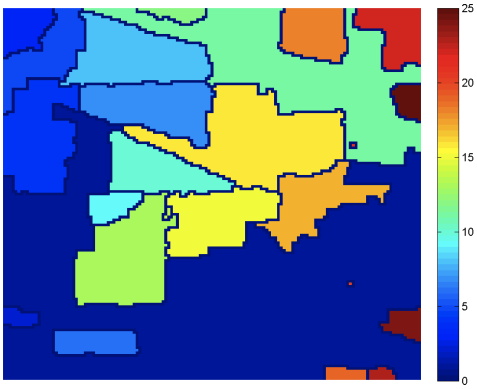

Fig. 1. *Segmentation map derived based on the first MNF component (each region is displayed in a certain colorcode).*

It is clear from the previous analysis that a number of alternative segmentation maps can be derived using the spectral information, based on the type of feature extraction (e.g., MNF or PCA), the number of components selected (e.g., only the first versus several), as well as the values used in the morphological operations during the post-processing step. Typically, it is not easy to identify intuitively which of these segmentation maps are relevant features in a given hyperspectral imagery and using all of them can result in degraded classification performance due to the curse of dimensionality. A systematic approach for selecting relevant segmentation maps is discussed in Section 4.

3. LIDAR-DERIVED FEATURES

Based on LiDAR data, several features can be derived including elevation, absolute value of the gradient, direction of the gradient or flatness, as depicted in Fig.2. Similarly to the hyperspectral segmentation maps, a multitude of more complex synthetic LiDAR features can be composed, for example by applying different thresholds on the gradient to capture flatness properties. By doing so, they can be treated similarly to the hyperspectral segmentation maps described in Section 2, i.e., unsupervised feature ranking by the random forest can be used to select the most informative of them.

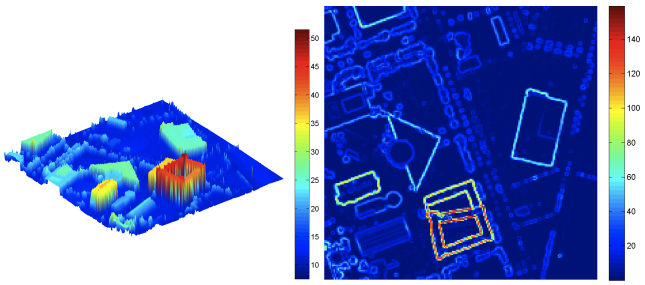

Fig. 2. *LiDAR elevation map (left) and gradient magnitude (right).*

4. RELATIVE FEATURE RELEVANCE USING RANDOM FORESTS

As indicated in previous sections, a multitude of different segmentation maps can be derived based on the hyperspectral and LiDAR data, using different feature extraction techniques, different post processing thresholds, etc. However, feeding all these segmentation maps to the classification scheme may not result in optimal performance due to the curse of dimensionality effect. On the other hand, selecting the optimal values for the different thresholds for every hyperspectral dataset is not an easy process. Supervised feature selection schemes can offer a solution, but these typically either rely on a heuristic (such as forward/backward selection) which is not always optimal or can result in very long (even intractable) computation times as the initial feature set increases. In this paper we propose to apply a standard relative feature ranking step that is an integral part of the classification process in random forest classifiers. The outcome of this step is the relative importance (ranking) of the features, derived already in an unsupervised fashion. This ranking can be used to select the most relevant segmentation maps in each dataset and use them as a source of spatial information in the classification framework.

The random forest algorithm, introduced by Breiman [5] utilizes both bagging and random attribute subset selection for achieving diversity between weak learners. As indicated by various empirical studies (e.g., [5], [6], [7], [8], the random forest algorithm has emerged as serious contester to state-of-the-art methods such as boosting [9] and Support Vector Machines [10]. Recently, it has also been used with great success in the classification of hyperspectral images [1], [2]. Here we provide only a very short description of the algorithm, to explain how feature ranking is done as part of the classification process. For more details we refer to

the survey by [11] which is an excellent introduction to the method, offering also useful practical guidelines.

Initially, the training data are sampled (random sampling with replacement) to create T different subsets of the data (where T is the selected number of trees in the forest), each of size N, with N approximately 66% of the complete training set. Each of these sets will be used in training one tree. At the first node of each tree a subset of $mtry$ of the predictor variables is selected at random from the set of all the predictor variables. The predictor variable that provides the best split in this subset, according to the selected objective function (usually based on an impurity measure, such as the information gain or the Gini gain) is used to do a binary split on that node. The algorithm continues with the next node, where a subset of $mtry$ of the predictor variables is selected at random and the process is repeated. After the training is done and the algorithm operates in classification mode, when a new input is entered it is run down all of the trees. If the range of valid predictions is $\mathcal{C} = \{1, ..., C\}$ where C is the total number of classes, then the estimated probability of predicting class $y \in \mathcal{C}$ for a given point $\mathbf{x_o}$ is:

$$p(y|\mathbf{x_o}) = \frac{1}{T}\sum_{t=1}^{T} p_t(y|\mathbf{x_o})$$

with $p_t(y|\mathbf{x_o})$ being the estimated density of class labels on the leaf of the t^{th} tree [5].

It is clear from this description that the algorithm provides inherently a good indication of feature importance, the ranking that each feature has based on the number of times it was used to do the split across all the nodes of the random forest. This ranking (often called relative feature importance or relative feature relevance) is derived as part of the classification stream in a fast and unsupervised way. Moreover, in contradistinction with most unsupervised selection schemes the algorithm can use the Out of Bag Error (which has been shown to be an unbiased error estimation for the algorithm performance) in order to validate that the classification is useful (feature selection is only meaningful if the classifier performs well [5], a validation criterion that most unsupervised schemes do not meet). The ranking derived in this way can be used to select a subset of the most informative segmentation maps and use only these as the source of spatial information for a given dataset.

5. EXPERIMENTAL RESULTS

In order to assess the effectiviness of the proposed classification framework and especially the performance improvement offered over [1], [2] by generating and evaluating a multitude of segmentation maps we use two hyperspectral datasets with completely different proporties. Specifically, our first dataset is the one featured in the IEEE GRSS 2013 Data Fusion Contest (University of Houston campus) [12]; this dataset captures a challenging and varied semi-urban environment with 15 classes. It also contains LiDAR information that is very well suited for deriving segmentation features. The second dataset we consider is Indian Pines. Despite being quite old, it is also a very challenging dataset of an agricultural area with 16 different classes of crops. It does not contain LiDAR information and segmentation of the different areas is both challenging and critical for performance. The experimental setup for each dataset is the same as in [1], [2]. Specifically, for the University of Houston we used the same training set as in the contest, while for Indian Pines the training set consists of 50 samples for each class that have been randomly chosen from the reference data, except for classes alfalfa, grass/pasture-mowed, and oats that have very few members; thus, only 15 samples for each of these classes were randomly picked to be used as training samples. All other samples composed the test set.

For both datasets we generate a multitude of segmentation maps, using the first 5 MNF and PCA components, applying watershed segmentation and then using morphological operators in order to alleviate the over-segmentation problem. For the University of Houston dataset we also generate the segmentation features based on LiDAR that were proposed in [1], as well as additional flatness features by applying different thresholds on the gradient as described in Section 3. The feature ranking generated by the random forest algorithm for the University of Houston is presented in Fig. 3. We can easily see that from the total of approximately 80 features, more than half are not improving the classification outcome and can be rejected (as expected a lot of the information on the different segmentation maps is redundant). By selecting only those that have ended up ranked positively, we can get a modest improvement on our classification result of [1]. Specifically, the final overall accuracy (forth column in Table III of [1]) increases from 94.4% to 94.6%. It is interesting to note that only 2 additional segmentation features derived from hyperspectral data are included in the best result, meaning that the LiDAR features that were used in [1] are already relatively sufficient in capturing the topological information.

Performance Metric	**Result of [1]**	**Proposed**
Overall accuracy (%)	94.4	94.6
Class average accuracy (%)	93.9	94.1
k statistic	0.94	0.94

Table 1. *Houston University final result from [1] and corresponding result with updated segmentation feature extraction and selection.*

The situation is quite different in Indian Pines. Here no LiDAR features to help with segmentation are available, therefore 7 hyperspectral-based segmentation features are selected in the best result. In addition, the performance im-

provement over [2] in this case is quite significant, from an accuracy of 93.0% to 96.3% (Table 2).

Performance Metric	Result of [2]	Proposed
Overall accuracy (%)	93.0	96.3
Class average accuracy (%)	87.1	89.8
k statistic	0.93	0.96

Table 2. *Indian Pines final result from [2] and corresponding result with updated segmentation feature extraction and selection.*

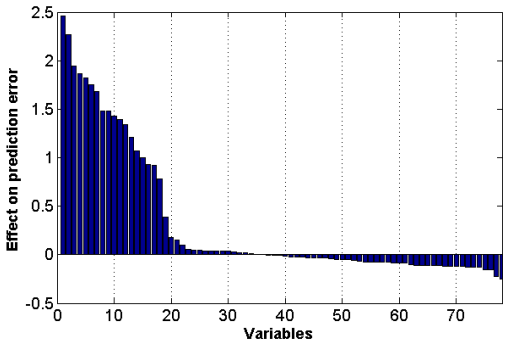

Fig. 3. *Feature ranking generated by the Random Forest algorithm for the University of Houston.*

The results for both University of Houston and Indian Pines are, to the best of our knowledge, the state of the art results in these datasets for the given training set setup.

6. CONCLUSION

In this paper we examined the problem of including topological information in the classification. A multitude of different segmentation maps can be generated using either the LiDAR or the hyperspectral data and applying a set of operations with different thresholds (e.g., thresholds on the gradient, or morphological operations on watershed segmentation maps). However, some of the generated maps are redundant while on the other hand using all of them is usually resulting in performance degradation due to the curse of dimensionality. We have shown how relative feature relevance can be used in random forests to select the most informative features so as to capture the morphological information without suffering from the curse of dimensionality. We have demonstrated the effectiveness of our method on two datasets, the first featuring a combination of hyperspectral and LiDAR data for urban classification, whereas the second being the well-known Indian Pines dataset featuring pure hyperspectral imagery.

7. ACKNOWLEDGEMENT

The authors would like to thank the Hyperspectral Image Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the data sets used in this study, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest.

8. REFERENCES

[1] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and LiDAR data fusion outcome of the 2013 GRSS Data Fusion Contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, to appear.

[2] A. Merentitis, C. Debes, and R. Heremans, "Application of ensemble learning in hyperspectral image classification: Towards selecting favorable spots in the bias-variance plane," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 1089 – 1102, April 2014.

[3] Y. Tarabalka, J. Chanussot, and J.A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognition*, vol. 43, no. 7, pp. 2367 – 2379, 2010.

[4] S. van der Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert, "Classifying segmented hyperspectral data from a heterogeneous urban environment using support vector machines," *Journal of Applied Remote Sensing*, vol. 1, 2007.

[5] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[6] V. Svetnik et al., "Random forest: A classification and regression tool for compound classification and QSAR modeling," *Journal of Chemical Information and Computer Sciences*, vol. 43, pp. 1947–1958, 2003.

[7] R. Diaz-Uriarte and S. Alvarez de Andres, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7 (3), pp. 1–13, 2006.

[8] J.-M. Poggi R. Genuer and C. Tuleau-Malot, "Variable selection using random forests.," *Pattern Recognition Letters*, vol. 31, pp. 2225–2236, 2010.

[9] Y. Freund and R. Shapire, "Experiments with a new boosting algorithm," in *Proceedings of the 13th International Conference on Machine Learning*, In L. Saitta, Ed., San Francisco, 1996, pp. 14–156.

[10] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press,, 2004.

[11] J.-M. Poggi R. Genuer and C. Tuleau, "Random forests: Some methodological insights," *arXiv:0811.3619, SSN 0249-6399*, 2008.

[12] 2013 IEEE GRSS Data Fusion Contest, "Online: http://www.grss-ieee.org/community/technical-committees/data-fusion/," .