

Artificial Neural Network Approach for Land Cover Classification of Fused Hyperspectral and Lidar Data

Paris Giampouras^{1,2}, Eleni Charou¹, and Anastasios Kesidis³

¹ Computational Intelligence Laboratory,
Institute of Informatics and Telecommunications,
National Center for Scientific Research "Demokritos",
Athens, Greece

² Department of Informatics, University of Athens, Greece
{parisg,exarou}@iit.demokritos.gr

³ Department of Surveying Engineering,
Technological Educational Institute of Athens, Greece
akesidis@teiath.gr

Abstract. Hyperspectral remote sensing images are consisted of several hundreds of contiguous spectral bands that can provide very rich information and has the potential to differentiate land cover classes with similar spectral characteristics. LIDAR data gives detailed height information and thus can be used complementary with Hyperspectral data. In this work, a hyperspectral image is combined with LIDAR data and used for land cover classification. A Principal Component Analysis (PCA) is applied on the Hyperspectral image to perform feature extraction and dimension reduction. The first 4 PCA components along with the LIDAR image were used as inputs to a supervised feedforward neural network. The neural network was trained in a small part of the dataset (less than 0.4%) and a validation set, using the Bayesian regularization backpropagation algorithm. The experimental results demonstrate efficiency of the method for hyperspectral and LIDAR land cover classification.

Keywords: Hyperspectral images, LIDAR, land cover classification, neural networks, principal component analysis.

1 Introduction

Operational monitoring of land cover/ land use changes using remote sensing is of great importance for environmental Remote sensing due to its repetitive nature and large coverage is a very useful technology to perform such kind of study. This technology, if properly integrated with automatic processing techniques, allows the analysis of large areas in a fast and accurate way. Several studies have been carried out in this field, analyzing the potentialities of different remote sensing sensors such us Passive Multispectral (3-10 spectral bands) or Hyperspectral(several 100s spectral bands). The abundance of spectral information in the hyperspectral image has the potential to differentiate land cover classes with

similar spectral characteristics that cannot be distinguished with multispectral sensors. Active LIDAR remote sensing sensors are increasingly used in the context of classification and constitute a valuable source of information related to the altitude. Thus hyperspectral and LIDAR data could be used complementary, as they contain very different information: hyperspectral images provide a detailed description of the spectral signatures of classes but no information on the height of ground covers, whereas LIDAR data give detailed information about the height but no information on the spectral signatures[9]. In this study LIDAR and hyperspectral images are integrated for the land cover classification. To show the usefulness and complementarity of LIDAR and hyperspectral image, a land cover classification was performed using hyperspectral image alone. This paper is organized into 4 sections. Section 2 presents the methodology used, the main preprocessing techniques adopted and describes the data set used in our analysis. Section 3 describes and discusses the experimental results obtained. Finally, Section 4 draws the conclusion of this paper.

2 Methodology and Data

Hyperspectral data undoubtedly possess a rich amount of information. Nevertheless, redundancy in information among the bands opens an area for research to explore the optimal selection of bands for analysis. Theoretically, using images with more bands should increase automatic classification accuracy. However, this is not always the case. As the dimensionality of the feature space increases subject to the number of bands, the number of training samples needed for image classification has to increase too. If training samples are insufficient then parameter estimation becomes inaccurate. The classification accuracy first grows and then declines as the number of spectral bands increases, which is often referred to as the Hughes phenomenon [1]. In the proposed method, hyperspectral image data are fused with LIDAR data in order to perform land cover classification. Principal Component Analysis is applied on the Hyperspectral image to perform feature extraction and dimension reduction while preserving information. The first four principal components along with the LIDAR image are used for classification by training a supervised feed-forward neural network. The following sections describe in detail the several steps of the method.

2.1 Principal Component Analysis

Classification performance depends on four factors: class separability, training sample size, dimensionality, and classifier type [2]. Dimensionality reduction can be achieved in two different ways. The first approach is to select a small subset of features which could contribute to class separability or classification criteria. This dimensionality reduction process is referred to as feature selection or band selection. The other approach is to use all the data from original feature space and map[10] the effective features and useful information to a lower-dimensional subspace. This method is referred to as feature extraction. The Principal Component Analysis PCA is a well known feature extraction method in which original

data is transformed into a new set of data which may better capture the essential information. When variables are highly correlated the information contained in one variable is largely a duplication of the information contained in another variable. Instead of throwing away the redundant data principal components analysis condenses the information in intercorrelated variables into a few variables, called principal components. PCA analysis have been successfully used in Hyperspectral data classification[11].

2.2 Neural Networks

The use of Artificial Neural Networks(ANNs) for complex classification tasks of high dimensional data sets, such as hyperspectral images, has been widely spread in the last years. Several studies dealing with pattern recognition/classification problems of remote sensing data, [3,4] have shown that ANNs, in most cases, achieve better results in comparison to conventional classifiers.

Their non-linear properties and the fact that they make no assumption about the distribution of the data, are among the key factors of the NNs classification power. Neurons with nonlinear activation functions are arranged in layers and act like a set of piece-wise nonlinear simulators [5,7]. Neural networks are able to learn from existing examples adaptively, thus, the classification is made objective[8].

2.3 Data Description

The data used in this study is obtained from the data fusion contest 2013, organized by the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society (GRSS). The current contest involves two sets of data a hyperspectral image (Fig. 1) and a LIDAR derived Digital Surface Model (DSM) (Fig. 2), co-registered and both at the same spatial resolution (2.5 m). The size of the bands of the hyperspectral image as well as the LIDAR is 349x1905 pixels. They were acquired over the University of Houston campus and the neighboring urban area. For the current data set, a total of 144 spectral bands were acquired in the 380 nm to 1050 nm region of the electromagnetic spectrum. The goal is to distinguish among the 15 pre-defined classes. The labels as well as a training set for the classes are also provided (Table 1).

3 Results and Discussion

For the purpose of this study the MATLAB software is used. The Principal Component Analysis is applied on the hyperspectral image. The first four PCA components i.e PCA1,PCA2,PCA3,PCA4 which convey the 99.81% of the information along with the LIDAR opted as our dataset for the processing that follows. A three layer feed-forward ANN is implemented. The input layer of the ANN consists of 5 neurons : 4 neurons for PCA1, PCA2, PCA3, PCA4 and one neuron for the LIDAR image. The hidden layer is composed of 120 neurons and

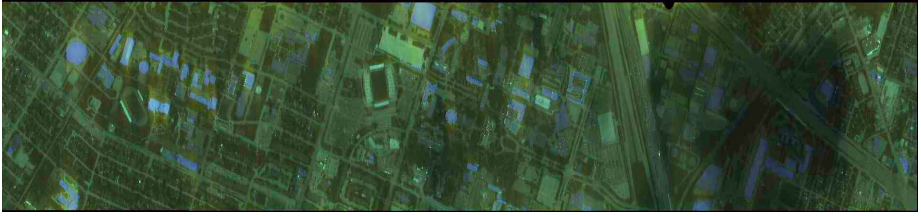


Fig. 1. RGB combination of hyperspectral image of the study area

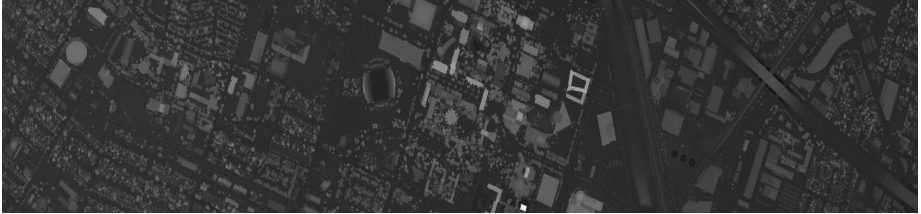


Fig. 2. LIDAR image of the study area

Table 1. Distribution of training sample among investigated classes

| Class Index | Class Name | Training Samples Number |
|-------------|-----------------|-------------------------|
| 1 | Grass Healthy | 198 |
| 2 | Grass Stressed | 190 |
| 3 | Grass Synthetic | 192 |
| 4 | Tree | 188 |
| 5 | Soil | 186 |
| 6 | Water | 182 |
| 7 | Residential | 196 |
| 8 | Commercial | 191 |
| 9 | Road | 193 |
| 10 | Highway | 191 |
| 11 | Railway | 191 |
| 12 | Parking lot 1 | 192 |
| 13 | Parking lot 2 | 184 |
| 14 | Tennis Court | 181 |
| 15 | Running Track | 187 |

the tan-sigmoid is selected as the layer's transfer function. The output layer comprised of 15 nodes equal to the total number of the classes. The linear function is opted as the output's layer transfer function. The ANN is trained by the set of 2832 samples of the Table 1. The training set is divided randomly into three sets, namely, the training, validation and test set, respectively. About 70% of the total samples is used as the training set of the ANN. Another 15% is used for validation while the rest 15% is dedicated to testing purposes. The ANN is trained

using the Bayesian Regularization back-propagation algorithm [6]. In this case, the weights and bias values are updated according to Levenberg-Marquardt optimization. This minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well. After 552 epochs the training procedure is stopped. The test set success is 94.6% , as it can be seen in detail in the confusion matrix of (Fig. 3). The performance achieved, is higher than 95% in the majority of the classes. Specifically, in 7 out of 15 classes (grass healthy, grass stressed, grass synthetic, tree, soil, water, running track), a 100% success is achieved. For 5 classes (commercial, highway, railway, parking lot 1, tennis court) the performance achieved lies between 95% and 99.5%. The classes "residential" and "road" present 85%-90% rate of success. The lowest value (61.5%) is noticed for class parking lot 2 mostly due to the high correlation and the low separability between this class and the parking lot 1 class.

| Output Class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |
|--------------|--------------|--------------|--------------|--------------|--------------|----------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|--------------|---------------|----------------|
| 1 | 30 7.1% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 2 | 0 0.0% | 27 6.4% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 3 | 0 0.0% | 0 0.0% | 19 4.5% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 4 | 0 0.0% | 0 0.0% | 0 0.0% | 33 7.8% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 5 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 14 3.3% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 6 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 27 6.4% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 7 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 32 7.5% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 0 0.0% | 97.0% 3.0% |
| 8 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 20 4.7% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 4 0.9% | 0 0.0% | 0 0.0% | 83.3% 16.7% |
| 9 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 0 0.0% | 26 6.1% | 1 0.2% | 1 0.2% | 0 0.0% | 1 0.2% | 0 0.0% | 0 0.0% | 96.7% 13.3% |
| 10 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 0 0.0% | 0 0.0% | 33 7.8% | 0 0.0% | 0 0.0% | 2 0.5% | 0 0.0% | 0 0.0% | 91.7% 8.3% |
| 11 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 25 5.9% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 100% 0.0% |
| 12 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 2 0.5% | 0 0.0% | 1 0.2% | 0 0.0% | 0 0.0% | 0 0.0% | 34 8.0% | 3 0.7% | 0 0.0% | 0 0.0% | 85.0% 15.0% |
| 13 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 1 0.2% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 16 3.8% | 0 0.0% | 0 0.0% | 84.2% 15.8% |
| 14 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 1 0.2% | 0 0.0% | 1 0.2% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 31 7.3% | 0 0.0% | 93.9% 6.1% |
| 15 | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 0 0.0% | 34 8.0% | 100% 0.0% |
| | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 100% 0.0% | 86.5% 13.5% | 95.2% 4.8% | 89.7% 10.3% | 97.1% 2.9% | 96.2% 3.8% | 97.1% 2.9% | 61.5% 38.5% | 96.9% 3.1% | 100% 0.0% | 94.6% 5.4% | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | |

Fig. 3. Test Set Confusion Matrix

By applying the ANN on the whole dataset, we get the resulting image that can be seen in Fig. 4, where each class is depicted with a different color. A quantitative assessment on the whole image, is not possible since no ground truth data are provided. However, a qualitative assessment on the whole classified image reveals promising results. It can be noticed that most pixels which belong to classes where high success rates are achieved, are correctly classified (compared to the RGB image of Fig. 2). In contrast, poorer results are achieved for pixels in classes characterized by lower classification rates, such as "parking lot 2".

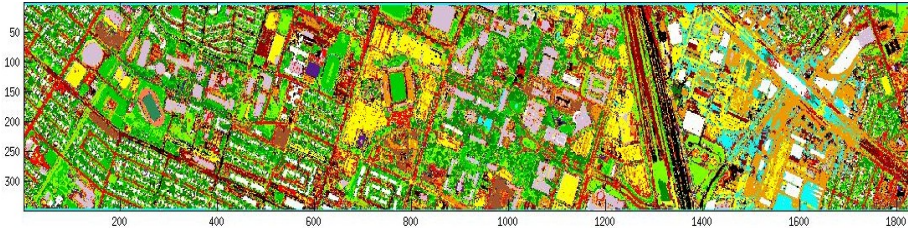


Fig. 4. Resulting Image

4 Conclusion

A fully automated process based on hyperspectral and LIDAR data for efficient land cover classification is presented. The fusion of hyperspectral image and LIDAR data is shown that can be used complementary, for effective land cover classification based on ANNs. Even in highly correlated classes such as grass healthy, grass stressed (unhealthy) and grass synthetic (ANN's) performance is very satisfactory. The methodology presented can also be used in other classification tasks.

Acknowledgements. This research has been co-funded by the European Union (European Social Fund) and Greek national resources under the framework of the Archimedes III: Funding of Research Groups in T.E.I. of Athens project of the Education & Lifelong Learning Operational Programme. The authors would also like to acknowledge the Data FusionContest 2013 Technical Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) for providing the dataset.

References

1. Hughes, G.F.: On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inform. Theory* 14, 55–63 (2003)
2. Hsien, P.F., Landgrebe, D.: *Classification of High Dimensional Data*. Ph. D. Dissertation, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana (1998)

3. Mernyi, Erzsébet, et al.: Classification of hyperspectral imagery with neural networks: Comparison to conventional tools. *Photogrammetric Eng. Remote Sens.* (2007)
4. Mernyi, E., et al.: Quantitative comparison of neural network and conventional classifiers for hyperspectral imagery. In: *Summaries of the Sixth Annual JPL Airborne Earth Science Workshop, Pasadena, CA, vol. 1* (1996)
5. Hornik, K., Stong, M., White, H.: Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366 (1989)
6. <http://www.mathworks.com/help/nnet/ref/trainbr.html> (last access: June 22, 2013, GMT 9:58:12 a.m)
7. Haykin, S.: *Neural Networks: A comprehensive foundation*. Macmillan, New York (1994)
8. Beale, M., Hagan, M.T., Demuth, H.B.: *Neural network toolbox*. Neural Network Toolbox, The Math Works, 5–25 (1992)
9. Dalponte, B.L., Gianelle, D.: Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Transactions on Geoscience and Remote Sensing* 46(5), 1416–1437 (2008)
10. Petridis, S., Charou, E., Perantonis, S.J.: Non redundant feature selection of multiband remotely sensed images for land cover classification. In: *Tyrrhenian International Workshop on remote sensing Elba, Italy*, pp. 657–666 (2003)
11. Rodarmel, C., Shan, J.: Principal Component Analysis for Hyperspectral Image Classification. *Surveying and Land Information Systems* 62(2), 115–122 (2002)