

A DATA INTERPRETION CHAIN FOR HYPERSPECTRAL REMOTE SENSING DATA AIMED AT BASIC VEGETATION MAPPING APPLICATIONS

Karoly Bakos, Paolo Gamba, Senior Member, IEEE

Department of Electronics, University of Pavia
Via Ferrata 1. Pavia 27100 Italy
{name.surname}@unipv.it

I. INTRODUCTION, MATERIALS AND METHODOLOGY

Hyperspectral data is more and more affordable for users and recently many applications address the problems of it applications [1-3]. These applications are promising but in the same time most of the studies are strongly specific to the particular application. As an outcome of HyRESSA workshop [4] it became clear that unfortunately there is no existing methodology for producing generic mapping products by means of hyperspectral images and the missing standards and framework makes it difficult for end users the efficient exploitation of these datasets. Although many users are interested in mapping applications of hyperspectral datasets, their computational resources, ability and willingness to build the algorithms proposed for each different application is a very limiting factor in promoting hyperspectral image processing.

In this paper we present the methodology used and the results obtained while selecting and testing suitable data processing chains for vegetation mapping by means of hyperspectral data. In the study we used a generic concept of the term processing chain introduced in [5], that starting from a radiometrically and geometrically correct data consist of a data volume reduction and transformation (Feature selection / Feature extraction) , a classification and optionally a post processing stage [6]. During the testing procedure several methodologies were applied at different stages of data processing and the results were compared. We applied the overall accuracy as a measure of performance of a data processing chain while after processing the data an in depth analysis of processing steps was carried out to understand the differences among the different processing chains. In the study we also propose a novel, Decision Tree [7] like, multi-stage data processing approach that is able to fulfil the needs of data classification while it is flexible enough to allow users to obtain a high quality mapping product using their available data and resources. The optimization procedure of the multistage classifier [8] is also presented in which only a limited number of classifications carried out on a subset of the data enables to carry out simulated classification by converting the confusion matrices obtained before. With this methodology an optional number of algorithms and inputs can be used from any source and also ancillary datasets can be included.

The aim of the study was to firstly limit the selection of suitable data processing chains for generic vegetation mapping for the most suitable ones and secondly to develop a multi stage flexible processing approach for the mapping problem that is able to aggregate the advantages of the best processing chains. The former aim was addressed by carrying out data processing using different chains (see Figure 1), analysing the results and also trying to understand why the differences arisen. The latter aim was addressed by constructing a decision tree classifier that uses propagated accuracy level of the whole decision tree structure for the optimization of the configuration and also handles ancillary datasets.

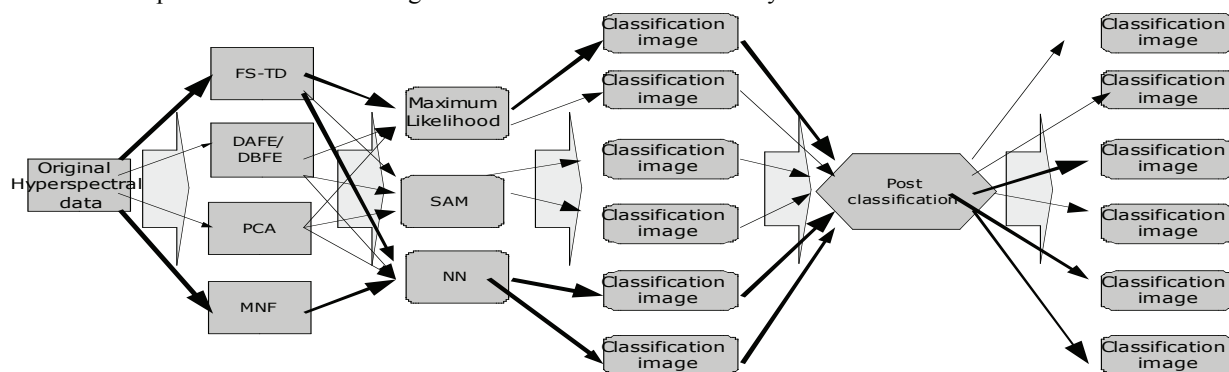


Figure 1. Graphical representation of the approach used in designing the different processing chains. Paths with highlighted arrows are identified to be more suitable than others. (FS-TD=Feature Selection with Transformed Divergence; MNF=Minimum Noise Fraction rotation; PCA=Principal Components Analysis; SAM=Spectral Angle Mapper; NN=Neural Net)

The test set used in this research was a DAIS 7915 dataset acquired by DLR and kindly provided by the University of Warsaw. In the study a case example of vegetation mapping of high mountainous ecosystems was used in the Tatra National Park in Poland. This was identified to be a good scenario as vegetation mapping by hyperspectral data were identified to be particularly suitable for these ecosystems because of the high complexity of land cover types and the inaccessibility of the

area [9]. Also the fragmented landscape and the rapidly changing elevation is highly limiting the application of other type of remote sensing data application e.g. multispectral data. The study area encompasses only a part of the Tatra Mountains – Gasienicowa Valley with its surroundings. The landscape is defined by one of three main geobotanical units of the Western Carpathians as follows: upper mountain belt (1200 - 1550 m) - spruce forests up to the timberline, with a cool climate; Subalpine belt (1550 - 1800 m) - dwarf pine, very cool climate; alpine belt (1800 - 2250 m) - alpine tundra with low high – mountain-grasslands dominance and moderately cold climate; and subnival belt (above 2250 m). The zones of lichens, mosses, liverworts and loose, low grasslands characterize it, with a low number of species, dispersed amidst the bare rocks. For standard processing chain certain methodologies were tested as shown in Figure 1. The different chains were constructed and the classifications were done. Using the confusion matrices gained during this analysis phase a novel method was implemented to optimize a multistage classifier. Using the matrices the combination of the best input image and best classification method was identified on a per class basis and a Decision Tree like structure was constructed.

II. EXPERIMENTAL RESULTS

The obtained results show that some processing chains significantly outperform other ones in specific input scenarios. In this study the highest mapping accuracy level could be obtained by the application of a Minimum Noise Fraction forward rotation followed by the application of a fuzzy neural network for data classification. Application of post processing showed suitability for further increasing accuracy of interpretation. The results of a few processing chains are listed in the table below. For dimensionality reduction MNF was found to be the most suitable and in the analysis phase it was seen that hyperspectral data contains significant noise at certain bands. In overall, NN outperformed all the other classification algorithms and inclusion of the information on spatial contexts also raised the accuracy level, although not dramatically.

Table 1. Overall accuracy values obtained using different processing chain configurations.

<i>Input data source</i>	<i>Dimensionality reduction</i>	<i>Classification</i>	<i>Post processing</i>	<i>Accuracy</i>
MNF	15 band	NN	+	87,96%
Original data	FS TD	ML	+	87,92%
Original data	FS TD	SAM	-	57,77%
Original data	DAFE	ML	-	72,08%
	Multi stage DTC approach			84,37%

The application of the multi stage approach gained similar results (84.37% measured with independent samples) than the best single input processing chain but was able to detect more accurately classes that were particularly difficult to separate by using a single stage classification approach. This shows high potential for vegetation mapping as in most of the cases the biggest problems occur when similar vegetation types are to be separated [10].

The presented methodologies and test results show that hyperspectral airborne imagery can be successfully used to derived high quality mapping products. The selection of methodology to process the data is particularly important and highly affects the quality of the derived land cover map. The study also shows that the analysis of the data and underlying conditions are highly important and could led to a proper methodology selection for data interpretation. The application of a multistage classification system enables more flexibility and can improve the performance of the mapping algorithm.

III. REFERENCES

- [1] V.L. Kamallesh, A. Luckman, W.M.F. Grey, and E.J. Milton, "Estimation of leaf area index from PROBA/CHRIS hyperspectral multi-angular data", Proc. of the Remote Sensing and Photogram. Soc. Conference 2008, Exeter, 15-17 September 2008.
- [2] E. Belluco, M. Camuffo, S. Ferrari, L. Modenese, S. Silvestri, A. Marani, M. Marani, "Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing", *Remote Sensing of Environment*, Vol. 105, No. 1, pp. 54-67, Nov.2006.
- [3] J. Ham, Y. Chen, M.M. Crawford and J. Ghosh, Investigation of the Random Forest framework for classification of hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing* 43 (2005), pp. 492–501.
- [4] HYRESSA SWOT and User Needs workshop report, accessed on line at <http://www.hyressa.net/documents/>, April 2007
- [5] P. Gamba, A. Plaza, J. Benediktsson, J. Chanussot,, "European perspectives in hyperspectral data analysis", 2007 IEEE Geoscience and Remote Sensing Symposium, 23-28 July 2007pp. 4794-4797.
- [6] J. A. Richards X. Jia *Remote sensing digital image analysis* 3rd Edition, pp. 196 Springer, 2005.
- [7] C. R. P. Hartmann, P. K. Varshney, K. G. Mehrotra, and C. L. Gerberich, "Application of information theory to the construction of efficient decision trees," *IEEE Trans. Inform. Theory*, vol. IT-28, No.4, pp. 565-577, 1982.
- [8] A. V. Kulkarni and L. N. Kanal, "An optimization approach to hierarchical classifier design," Proc. 3rd Int. Joint Conf. on Pattern Recognition, San Diego, CA, 1976.
- [9] B. Zagajewski, A. Kozłowska, M. Krowczyńska, M. Sobczak and M. Wrzesien, "Mapping High Mountain vegetation using hyperspectral data," *EARSel eProceedings* 4, 1/2005.
- [10] J. Franklin Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients Janet Franklin. *Progress in Physical Geography*, Vol. 19, No. 4, 474-499 (1995) DOI: 10.1177/030913339501900403