European perspectives in hyperspectral data analysis

Paolo Gamba[1], Antonio J. Plaza[2], Jon A. Benediktsson[3] and Jocelyn Chanussot[4]
[1]Department of Electronics, University of Pavia, Italy
[2] Department of Computer Science, Univ. of Extremadura, Spain
[3] School of Engineering, University of Iceland, Iceland
[4] Electrical Engineering Department, Grenoble Institute of Technology, France
Corresponding author: paolo.gamba@unipv.it

***Abstract*— This paper explains some of the the goals and objectives of the newly started HYPER-I-NET Marie Curie Research and Training Network. In particular, the requirements related to the definition and implementation of an efficient, adequate and sufficiently general data processing chain for hyperspectral data analysis are considered. Some of the research lines that are expected to play a central role in the activities of this network are also presented and briefly discussed.**

I. INTRODUCTION

Hyperspectral imaging generates very high-dimensional imagery through the use of sensor optics with a large number of nearly contiguous spectral bands [1]. Hyperspectral sensors are, and will be available aboard European airborne (ARES) and satellite platforms (MERIS). The high spatial resolution and significantly improved spectral resolution provided by these last-generation instruments has opened ground-breaking perspectives in many remote sensing applications [2], including environmental modeling and assessment, target detection for military and defense/security purposes, urban planning and management studies, risk/hazard prevention and response including wild land fire tracking, biological threat detection, monitoring of oil spills and other types of chemical contamination, etc. As a result, studies based on hyperspectral image data allow a highly interdisciplinary approach, since hyperspectral imaging is an emerging research field open to several new challenges and perspectives.

With the purpose of addressing these issues, the European community recently funded an Hyperspectral Imaging network (in short HYPER-I-NET) for both research and training of young researchers in this exciting and novel field. The stress is especially on remotely sensed hyperspectral imaging, due to the background of the institutions and companies involved. Training will be focused on early stage researchers, and research will be oriented towards scientific goals which are achieved through thorough training, in all the research areas that comprise the entire hyperspectral processing chain, from sensor design and flight operation to data collection, processing, management and interpretation. In fact, even though hyperspectral imaging has been a very active area recently, no specific attention has been given to research activities covering the entire data processing chain.

More specifically, from a European perspective the abilities in this area are fragmented throughout various research teams and companies, a fact that has largely resulted in the lack of data standardization and validation procedures. For instance, the recent European Fleet for Airborne Research (EUFAR) initiative is currently working toward an integrated management of European fleets, including those in hyperspectral remote sensing studies. In this regard, the proposed RTN may provide a timely and unique opportunity to bridge the gap between the operational procedures of hyperspectral imaging (including sensor design) and the development of techniques for efficient data exploitation and management.

II. HYPERSPECTRAL REMOTE SENSING AND DATA PROCESSING

Hyperspectral data processing is a really complex procedure, and it's difficult to define it in a comprehensive and consistent way. HYPER-I-NET leverages the results of a research project funded as support action by the European community, entitled HYRESSA (HYperspectral REmote Sensing in Europe specific Support Actions). This project already addressed the issues about the data processing of hyperspectral data by means of a dedicated expert workshop, hold at DLR premises in July 2006. HYRESSA findings are summarized in the final report of that workshop [3] and suggest the definition of a two stages' chain, whose structure is also graphically presented in fig. 1.

The first part of the chain is devoted to the so called providers side, and highlights the needs for addressing some of the problems that prevent hyperspectral data to be widely used in applications. The relevant processing steps are:

- radiometric calibration, where new algorithms need to be developed, e. g. for thermal sensors;
- geometric correction, where the lack of precise DEMs does not allow to have consistent products;
- atmospheric correction;
- objective evaluation of product accuracy, where the sector suffers from the lack of standards for data format and metadata description.

However, once the data of interest has been pre-processed and corrected, there is a need to extract relevant information from the collected data sets. The second part of the data processing chain (user's side) will therefore deal with four more steps, described in the following paragraphs.

- Data transformations, with the need of defining BRDF and its effects, especially in areas where the viewing

1-4244-1212-9/07/$25.00 ©2007 IEEE.

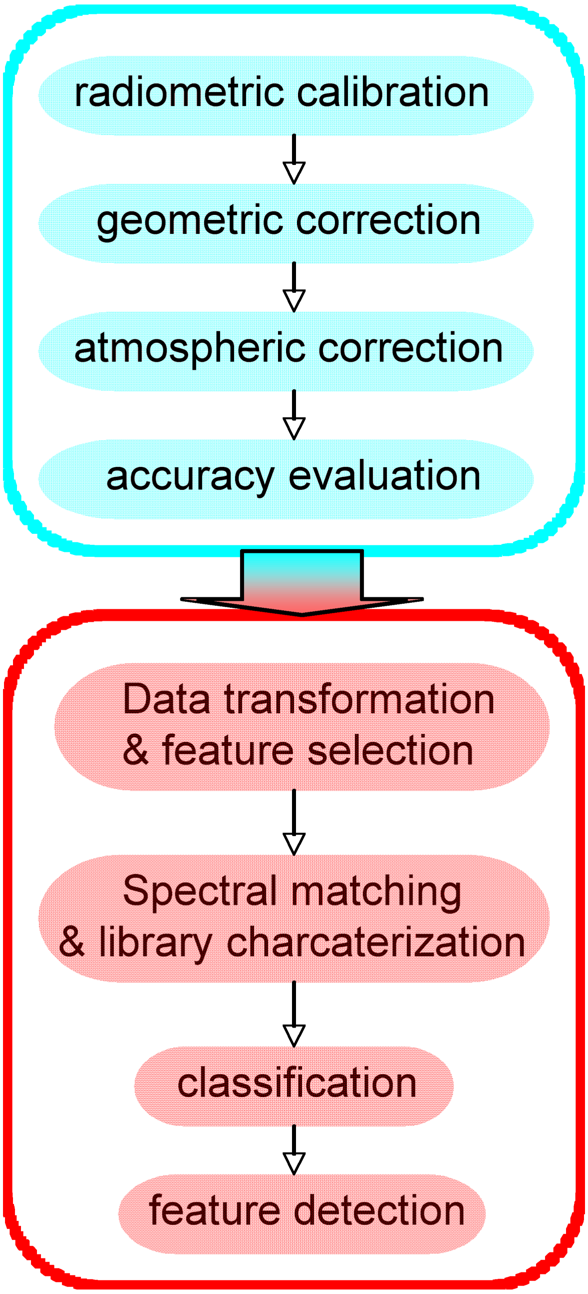

Fig. 1. The most general hyperspectral data processing chain structure.

angle may be extremely diverse, such as urban areas, forests or very steep terrains;

- spectral matching, with the need of having centralized libraries for many different materials;
- classification, with the possibility to take into account not only spectral characteristics but also spatial relations among neighboring pixels/materials;
- Feature detection, with the need to define precisely characterized and accurately validated high level products.

According to this list, the strategy of HYPER-I-NET is based on the basic idea of a collaborative projet. Therefore, each research laboratory or company involved in the project team work towards reducing weaknesses of one or two particular steps of the chain (according to the outputs of the first year), possibly in team with other units. The designed chain will be tried and checked for the first time at the summer school at the end of the second year (2008) on a well-known data set. Following the results of this first check, further optimization will be considered. General issues to be addressed might be for instance computational efficiency and suitability to multiple sensors. In parallel, more specific issues might instead be new or improved processing algorithms for some of the steps, as well as corrected methodologies based on wrong assumptions amended after the first check.

The final year of the project will be devoted to provide guidelines for HW/SW implementation of the data processing chain algorithmic steps, built over the experience of the previous three years and addressing both the requirements for SW developers with respect to interfaces, data formats, and handling procedures, as well as parallel processing issues. Moreover, validation sites will be selected among the already available data sets and those obtained during the summer schools and related acquisition campaigns. They will designed not only to validate the chain (as in previous slides), but also to serve as test sites to European companies and research laboratories, i.e. for benchmarking purposes.

A. *Open lines of research*

According to this general scheme, while several data analysis techniques have been proposed to implement all of the above mentioned steps in the technical literature (and won't be summarized here for space constraints), many possibilities still remain open. In particular, the network chose to focus on the two topic areas listed in the following paragraphs.

- **Multidimensional approaches**. A big problem in hyperspectral data analyses is the well-known curse of dimensionality or Hughes effect [4]. This problem resides on the fact that it is extremely difficult to extract useful information from very high-dimensional spaces. As a result, there is a need for unsupervised feature extraction methods able to reduce the dimensionality of the data to the right subspace without losing the original information that allows for the separation of classes. While the Hughes effect has a strong impact on supervised classification procedures, due to the need for large volumes of training data which are generally very costly to be obtained in remote sensing studies, there are also other applications that need all spectral bands to analyze the full spectral shape, e.g., in radiative transfer-based approaches [5]. Further, it is worth noting that most available hyperspectral analysis techniques do not incorporate information on the spatially adjacent data, i.e., the data is generally not treated as an image, but as an unordered listing of spectral measurements that can be shuffled arbitrarily. In order to exploit hyperspectral imagery in applications requiring high spatial resolution, e.g., urban land-cover mapping, it is not only necessary to use the available spectral information, but to incorporate the spatial information as well [6], [7].

- **Computing**. Another major issue in hyperspectral imaging is the need for fast, parallel data processing techniques able to transform the massive amount of information currently produced by hyperspectral sensors into scientific understanding quickly enough for practical use. Systems for near real-time information extraction and mining from very large data archives are highly required (it is estimated that hundreds of Gigabytes of hyperspectral data are collected in Europe every day). Also, real-time systems for onboard data analysis and compression are only partially available, and need to be fully incorporated to remote sensing missions. Current sensor design practices could greatly benefit from the inclusion of data pre- and post-processing modules such as digital signal processors (DSPs) or field-programmable gate arrays (FPGAs), which can be mounted or embedded in the sensor. Although the development of such systems can greatly benefit from the integration of sensor operation and data processing knowledge, the true fact is that current systems do not incorporate any information feedback between sensor design and data analysis experience, which is a major shortcoming.

The need for research in these areas has been distributed and made more "experimental" within the HYPER-I-NET research and training program by means of the proposals for Ph.D. student funds to be activated jointly by the network partners.

- **Efficient endmember extraction and spectral unmixing algorithms**. The research effort is oriented towards the development and implementation of innovative endmember extraction and spectral unmixing algorithms for hyperspectral image analysis.
  Hyperspectral unmixing is a source separation problem. Compared with the canonical source separation scenario, the sources in hyperspectral unmixing (i.e., the materials in the scene) are statistically dependent and may combine in a nonlinear fashion. These characteristics, together with the high dimensionality of hyperspectral vectors, place the unmixing of hyperspectral mixtures beyond the reach of most source separation algorithms, thus fostering active research in this theme.
  Different strategies (spectral versus hybrid approaches) need to be explored for endmember extraction, while linear and nonlinear spectral mixture models have also to be compared and assessed. Given the extremely high dimensionality of hyperspectral imagery, particular attention is necessary to the efficient implementation of developed techniques.
- **Spectro-spatial classification in hyperspectral imagery**. the research effort is oriented towards the definition and implementation of advanced algorithms for the classification of hyperspectral data. The major goal is the innovation of algorithms incorporating the spectral information, which characterizes the physical nature of the materials, but also the spatial information about the objects (shape description or texture). Improved classification performances are expected from this joint use. Key scientific issues include feature extraction, dimension reduction and data fusion. This work follows previous successful studies on hyperspectral data from urban areas.
- **Vegetation monitoring using hyperspectral data and advanced processing techniques**. The research effort is oriented towards vegetation monitoring using hyperspectral airborne or spaceborne data in both mountainous and urban areas. The research will require the definition and the implementation of a full processing chain able to start with radiometrically calibrated data and ending with maps of vegetated species and tree status in various environments. Testing of known algorithms and development of innovative and advanced ones for each step of the defined chain will be required, with particular emphasis on geometric and atmospheric corrections and classification procedures based on known spectral of vegetated species. Joint spectral and spatial analyses will be also considered to implement the final mapping chain. Mandatory part of the Ph.D. will be also validation and accuracy assessment of the outputs of the processing chain, by means of both in situ spectral measurements and visual inspection.
- **Source separation in hyperspectral imagery**. The research effort is oriented towards the definition and implementation of advanced algorithms for source separation in hyperspectral imagery. The goal is to perform endmember extraction (i.e. estimate the respective proportions of different materials included within one pixel) using statistical methods such as the Independent Component Analysis (ICA) or Bayesian approaches. Physical considerations modeling the mixture of elements at the pixel resolution should be incorporated for a better separation (linear or non-linear model?). Different strategies (spectral vs spatial approaches) will be investigated.

III. Spatio-spectral data analysis

Just to give a more detailed example of ongoing researches involving all the four partners of the HYPER-I-NET team mostly involved in data processing, we will discuss here briefly about the problems and the perspectives connected to joint spectral and spatial analysis of hyperspectral data. the problem is really interesting because the addition of spatial features to the (already huge) number of spectral feature of an hyperspectral data sets makes the task of data analysis extremely more complex that in the simple per-pixel case.

More precisely, the past activities of the groups represented in this work has been devoted to texture analysis [8] and morphological processing [9], [10] of multispectral and hyperspectral data over urban areas. The key point of these works is that multi-scale analysis is required in order to consider all the diverse spatial relationships among objects in an urban environment. Adaptive processing is required, or multiple scales should be contemporarily used to gather enough information to process correctly the data. To make the whole procedure as much data-driven as possible, automatic methodologies to extract the scale of an image are required. This is one interesting line of research [11], actively pursued but very difficult to work at very high spatial resolution given the enormous number of possible combinations and the abrupt scale changes that may occur moving from one part to another even of the same town.

One of the major points related to this research is therefore an efficient and safe feature selection strategy, able to maintain the selected set into a reasonable dimensionality, but also to adapt to different problems. The most reasonable choice is to rely on a problem-driven, as opposite to data-driven, approach. In particular, supervised feature selection has been implemented, where applicable, but other methodologies could be useful, too. The point here is that no comparison has been provided so far on the complexity of the algorithms, the number of selected features, the robustness of all these approaches to the feature selection errors induce by a wrong training set.

Just to give an idea of the problems involved, we considered a common test site, available through the network to everybody and part of the urban data set available at the University of Pavia [12]. In particular, we refer here to the data of the DAIS 7915 airborne imaging spectrometer, operated by DLR. The sensor acquired a data set over the center of Pvaia in July 2002. This data set, together with the one recorded during the same flight by the ROSIS sensor, has served as test bench to a variety of algorithms. Some of them are based on textural feature computation and selection using the HDI index, while other ones use morphological profiles [4], [7], sometimes after a PCI transformation. The numbers available in these works as for the dimensions of the most useful subset of original or transformed bands is highly variable. However, in spite of a very limited variability of the overall accuracy values, there is a large variability of the number of features used. This, in turn, means that the important information in the data is captured in different ways by the different procedures proposed. This requires at least some kind of harmonization, in order to provide a more consistent guess of the complexity of this step in the whole hyperspectral data processing chain. An accurate and precise selection strategy is therefore one of the goals of the research within HYPER-I-NET.

IV. Conclusions

The present work presents the goals of the HYPER-I-NET project with respect to definition and implementation of a data processing chain for hyperspectral data. In so doing, this work has also highlighted some of the open question or required optimizations of this chain that need to be addressed during the time of this project.

Acknowledgments

The authors want to thank DLR for providing the DAIS urban hyperspectral data set used as an example in this work.

References

[1] R. O. Green et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, pp. 227248, 1998.

[2] J. A. Richards, "Analysis of remotely sensed data: the formative decades and the future," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 3, pp. 422432, 2005.

[3] HYRESSA SWOT and User Needs workshop report, accessed on line at http://www.hyressa.net/documents/, April 2007.

[4] J. A. Benediktsson, J. A. Palmason and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, pp. 480491, 2005.

[5] G. Strub, M. E. Schaepman, Y. Knyazikhin and K. I. Itten, "Evaluation of spectrodirectional alfalfa canopy data acquired during Daisex99," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, pp. 10341042, 2003.

[6] P. Gamba, F. DellAcqua, A. Ferrari, J. A. Palmason, J. A. Benediktsson and K. Arnasson, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sensing Letters*, vol. 1, no. 4, pp. 322326, 2004.

[7] A. Plaza, P. Martínez, J. Plaza and R. Pérez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 3, pp. 466479, 2005.

[8] A. Plaza, P. Gamba, G. Trianni, J. Plaza, "Efficient multi-band texture analysis for remotely sensed data interpretation in urban areas", Proc. of *2007 Joint Remote Sensing Event*, Paris, France, 11-13 April 2007, unformatted CD-ROM.

[9] J.A. Benediktsson, J.A. Palmason, J.R. Sveinsson, "Morphological preprocessing for classification of hyperspectral data from urban areas", Proc. of *2003 IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, 27-28 Oct. 2003, pp. 290-297.

[10] J.A. Palmason, J.A. Benediktsson, J.R.Sveinsson, J. Chanussot, J"Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis", Proc. of *20'5 IEEE Geoscience and Remote Sensing Symposium*, Seoul, Korea, 25-29 July 2005, Vol. 1.

[11] P. Gamba, F. DellAcqua, G. Trianni: "Semi-automatic choice of scale-dependent features for satellite SAR image classification", *Pattern Recognition Letters*, Vol. 27, n. 4, pp. 244-251, Mar. 2006.

[12] P. Gamba: "A collection of data for urban area characterization", Proc. of IGARSS'04, Anchorage (USA), Sept. 2004, Vol. I, pp. 69-72.