A EUROPEAN PROPOSAL FOR TERMS OF REFERENCE IN DATA FUSION

L. Wald
Groupe Télédétection & Modélisation, Ecole des Mines de Paris
BP 207, 06904 Sophia Antipolis cedex, France

Commission VII

KEY WORDS: data fusion, concepts, definitions, remote sensing, EARSeL

ABSTRACT

Data fusion is a subject becoming increasingly relevant as scientists try to extract more and more information from remotely sensed data. Archives are growing, as well as the number of space missions devoted to Earth observation. It is generally correct to assume that improvements in terms of classification error probability, rejection rate, and interpretation robustness, can only be achieved at the expenses of additional independent data delivered by more separate sensors. Sensor data fusion allows to formalise the combination of these measurements, as well as to monitor the quality of information in the course of the fusion process.
A Special Group of Interest ‘data fusion’ has been established jointly within the European Association of the Remote Sensing Laboratories (EARSeL) and the French Society for Electricity and Electronics (SEE). This Group has defined several major tasks to be handled in order to increase our understanding and use of data fusion. One of these tasks is the establishment of terms of reference that are accepted by both the scientific and the industrial communities at least in Europe.
A definition of the data fusion is proposed, which allows to set up a conceptual approach to the fusion of Earth observation data by putting an emphasis on the framework and on the fundamentals in remote sensing underlying data fusion. Several other definitions are given which are useful to describe any problem of data fusion.

RESUMÉ

La fusion de données revêt une importance grandissante au fur et à mesure que les scientifiques tentent d’extraire de plus en plus d’information des mesures et images d’observation de la Terre. La taille des archives croît rapidement, de même que le nombre de missions spatiales dédiées au domaine. Il est généralement observé que les améliorations en termes de qualité, taux de classification, taux de rejet, robustesse d’analyse ou d’interprétation, ..., sont atteintes par l’augmentation des observations independantes effectuées par des instruments indépendants. La fusion de données permet de formaliser la manière dont les observations sont combinées, et également de mieux suivre l’évolution de la qualité au cours des différents processus de traitement.
Un groupe de travail (SIG) ‘fusion de données’ a été créé au sein de l’association européenne des laboratoires de télédétection (EARSeL) et de la société française d’électricité et d’électronique (SEE). Ce groupe a défini plusieurs tâches principales à traiter afin d’augmenter notre compréhension de la fusion de données et son usage. L’une d’entre elles est l’établissement de termes de référence, qui soient acceptés par les communautés scientifiques et industrielles, tout au moins en Europe. Cette communication présente ces termes. La définition proposée pour la fusion de données, met l’accent sur la formalisation de la fusion de données et sur les problèmes fondamentaux sous-jacents en télédétection, et par conséquent, permet de définir une approche conceptuelle de ce domaine. On définit également d’autres termes qui sont utiles dans la description des problèmes de fusion de données.

1 INTRODUCTION

Data fusion is a very recent word. It means an approach to information extraction spontaneously adopted in several domains. An illustration is given by the human system which calls upon its different senses, its memory and its reasoning capabilities to perform deductions from the information it perceives. However the operation by itself is not new in remote sensing: classification procedures are performed for more than twenty years, and are obviously relevant to data fusion (see e.g., Mangolini 1994, Pohl 1996).
The quantity of information available to describe our environment increases rapidly. Archives are growing, as well as the number of space missions devoted to Earth observation. Accordingly, data fusion is a subject becoming increasingly relevant as scientists try to extract more and more information from remotely sensed data using the concept of synergy. Sensor data fusion allows to formalise the combination of these measurements, as well as to monitor the quality of information in the course of the fusion process.

The European Association of Remote Sensing Laboratories (EARSeL) and the French society for Electricity and Electronics (SEE) are fully aware of the importance of data fusion in remote sensing. They jointly organised a conference in 1996, called ‘Fusion of Earth Data - merging point measurements, raster maps and remotely sensed images’, which was the first of a series with a venue every 2 years. During the round table of that meeting, the need for a working group, called special interest group (SIG) in the EARSeL jargon, was expressed, and it was created later this year.

2. THE CREATION OF A SIG WITHIN EARSeL - SEE

The SIG 'data fusion' will contribute to a better understanding and use of data fusion by tackling the fundamentals in remote sensing underlying data fusion. Further the SIG will certainly be helpful in designing and engineering tools and methods for assessing *a priori* or *a posteriori* the quality of a fusion product, that is answering the following questions: is it worth performing a fusion process ? was it worth doing it ?

During the 1996 conference, it has been proposed to restrict the SIG to the so-called radiometric aspects. The so-called geometrical aspects are already tackled by several working groups at national and international levels. The exact meaning of these words (radiometric, geometrical) is still unknown to me but every one can understand this share in problems. In other words, the present SIG will not handle matter related to any pure geometrical use of the data, e.g. image matching for geo-referencing, or geo-referencing, accurate positioning, assessment of digital terrain models, ... unless it appears necessary for the wealth of our studies and if information cannot be obtained from elsewhere.
All applications are foreseen, none is excluded. The already foreseen work tasks are:

- list and better understand the fundamentals in data fusion
- list and better understand tools and methods in data fusion, develop new ones
- develop and provide instruments for the assessment of the quality of the fusion
- prepare application cases exhibiting several processes and levels of fusion. These cases will help in illustrating data fusion and in students training. Several domains (urban domain, meteorology, ...) should be handled.
- prepare sets of data for well-documented sites which may be useful for testing algorithms. This should be performed with the help of the space agencies and other data providers.

3. THE NEED FOR TERMS OF REFERENCE

The concept of data fusion is easy to understand. However its exact meaning varies from one scientist to another. Several words have appeared, such as merging, combination, synergy, integration, ... All of them appeal more or less to the same concept but are however felt differently. There is also a fashion. Several times, the word « fusion » is used while « classification » would be more appropriate, given the contents of the publication.
There is a need for terms of reference in the remote sensing community. It has been strongly expressed in several meetings, including those organised by EARSeL or SEE (see e.g., Van Genderen, Pohl 1994; Wald 1997). The establishment of a lexicon or terms of reference permits to the scientific community to express the same ideas using the same words, and also to disseminate their knowledge towards the industry and 'customers' communities. Moreover it is a *sine qua non* condition to set up clearly the concept of data fusion and the associated formal framework. Such a framework is mandatory for a better understanding of data fusion fundamentals and of its properties. It allows a better description and formalisation of the potentials of synergy between the remote sensing data, and accordingly, a better exploitation of these data.
The present communication aims at providing the basis for this framework. It should be noted that this is not the only attempt to set up definitions in data fusion. The remote sensing community should not establish terms which are also used elsewhere with different meanings. Therefore, whenever possible, definitions were adopted which are already widely used in the broad scientific community, especially that dealing with information. Examples of such terms are image, features, symbols, etc.
Several lexicons have been already set up. They have all been established in the framework of the Defence domain (e.g., U.S. Department of Defence 1991; DSTO 1994). Most of the terms are part of the military jargon. They express needs of the Defence which may be partly similar to those in other domains where crisis occur, such as the management of a power plant. However it is not easy to translate military terms in meaningful words for the scientific community dealing with Earth observation. Using these military lexicons would imply a refinement of the military terms to expand their meaning, with a reference to the time-space scales. It was concluded that using an existing lexicon is not straightforward, and that a new one is required to tackle the specific needs of our community. However we should benefit from these previous works as much as possible, and, whenever possible, we should use either the terms already adopted or global architectures, etc.
The present communication summarises the discussions held within the SIG since the first conference 'Fusion of Earth data' held in Cannes, France, in February 1996. It proposes some terms of reference which have met a consensus during the second conference, held in January 1998.

4. THE EARSeL - SEE PROPOSAL

Data fusion means a very wide domain. It gathers a large number of methods and mathematical tools, ranging from spectral analysis to plausibility theory. Fusion is not specific to a theme or an application. On the contrary the tools used in a fusion process for a specific application may be tailored to that case. It is very difficult to provide a precise definition of data fusion. This large domain cannot be simply defined by restricting it, for example, to specific wavelengths, or specific acquisition means, or specific applications. Fusion process may call upon so many different mathematical tools that it is impossible to define fusion by these tools. For example, both the simple sum of two images acquired by two different sensors, and the more sophisticated encrustation of one image into the other using the multiresolution analysis (Wald, Ranchin 1995), are considered as fusion processes. Both implies at least a preliminary geocoding of the data. A classification technique based upon a sophisticated neural network is also a fusion process.
Several definitions have already been proposed. They have been discussed by e.g., Buchroithner (1998) or Wald (1997, 1998). During the meetings of the SIG as well as in the conferences 'Fusion of Earth Data', it was felt that most of these definitions were focusing too much on methods though paying some attention to quality. Some of them are restricted to sensors and their output signals. As a whole, there is no reference to concept in these definitions while the need for a conceptual framework was clearly expressed in these meetings.
In data fusion, information may be of various nature: it ranges from measurements to verbal reports. Some data cannot be quantified; their accuracy and reliability may be difficult to assess. In Earth observation domain, one may use some features held in a geographical information system to help in classifying multispectral images provided by several sensors. In this particular case, some data are measurements of energy, and others may be symbols.
Accordingly the definition for data fusion should not be restricted to data output from sensors (signal). Opposite to most of the published definitions, it should not be restricted to methods and techniques or architectures of systems, since we aim at setting up a conceptual framework for data fusion. Based upon the work of Wald (1997, 1998), the following definition was adopted: « data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application ».

This definition is clearly putting an emphasis on the framework and on the fundamentals in remote sensing underlying data fusion instead of on the tools and means themselves, as is done usually. The latter have obviously strong importance but they are only means not principles. Secondly it is putting also an emphasis on the quality. This is certainly the aspect missing in most of the literature about data fusion, but one of the most delicate. Here quality has not a very specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the « customer » when performing the fusion process than without it. For example, a better quality may be an increase in accuracy, or in the production of a more relevant information.

In this definition, spectral channels of a same sensor are to be considered as different sources, as well as images taken at different instants.

It then has been suggested to use the terms merging, combination in a much broader sense than fusion, with combination being even broader than merging. These two terms define any process that implies a mathematical operation performed on at least two sets of information. These definitions are very loose intentionally and offer space for various interpretations. Merging or combination are not defined with an opposition to fusion. They are simply more general, also because we often need such terms to describe processes and methods in a general way, without entering details. Integration may play a similar role though it implicitly refers more to concatenation (*i.e.* increasing the state vector) than to the extraction of relevant information.

Another domain pertains to data fusion: data assimilation or optimal control. Data assimilation deals with the inclusion of measured data into numerical models for the forecasting or analysis of the behaviour of a system. A well-known example of a mathematical technique used in data assimilation is the Kalman filtering. Data assimilation is daily used for weather forecasting.

Fusion may be performed at different levels: at measurements level, at attribute level, and at rule or decision level. These terms as well as others related to information are defined in the following. These definitions are those used in information theory and have been found in several publications (e.g., Bijaoui 1981; Lillesand, Kiefer 1994; Kanal, Rosenfeld 1981; Tou, Gonzalez 1974).

Measurements are primarily the outputs of a sensor. It is also called signal, or image in the 2-D case. The elementary support of the measurement is a pixel in the case of an image, and is called a sample in the general case. By extension, measurement denotes the raw information. For example, a verbal report is a piece of raw information, and may be considered as a signal. In remote sensing, in the visible range, the measurements are digital numbers that can be converted into radiances once the calibration operations performed. If corrections for the sun angle are applied, one may get reflectances which are still considered as signal.

An object is defined by its properties, e.g., its colour, its materials, its shapes, its neighbourhood, etc. It can be a field, a building, the edge of a road, a cloud, an oceanic eddy, etc. For example, if a classification has been performed onto a multispectral image, the pixels belonging to the same class can be spatially aggregated. This results into a map of objects having a spatial extension of several pixels. By extension, a pixel may be considered as an object.

An attribute is a property of an object. For example, the classification of a multispectral image allocates a class to each pixel; this class is an attribute of the pixel. The equivalent terms label, category or taxon are also used in classification. Another well-known example is the spatial context of a pixel, computed by local variance, or structure function or any spatial operator. This operation can be extended to time context in the case of time-series of measurements. Equivalent terms are local variability, local fluctuations, spatial or time texture, or pattern. By extension, any information extracted from an image (or mono-dimensional signal) is an attribute for the pixel or the object. The aggregation of measurements made for each of the elements of the object (for example, the pixels or samples constituting the object), such as the mean value, is an attribute. Some authors call mathematical attribute such attribute deriving from statistical operations on measurements. Feature is equivalent to attribute.

The properties of an object constitute the state vector of this object. This state vector describes the object, preferably in an unique way. The state vector is also called feature vector, or attribute vector. The common property of the elements of the state vector is that they all describe the same object. If the object is a pixel (or a sample), the state vector may contain the measurements as well as the attributes extracted from the processing of the measurements.

Works in pattern recognition have drawn an analogy with the syntax of a language. Terms of higher semantic content have been defined, such as rules and decisions. Rules, like the syntax rules in language, define relationships between objects and their state vectors, and also between attributes of a same state vector. Rules may be state equations, or mathematical operations, or methods (that is a suite of operations, *i.e.* of elementary rules). They are often expressed in elaborated language. Known examples of such rules are those used in artificial intelligence and expert-systems. Decisions result from the application of rules on a set of rules, objects and state vectors. Fusion may also be performed on decisions.

A fusion system can be a very complicated system. It is composed of sources of information, of means of acquisition of this information, of communications for the exchange of information, of intelligence to process the information and to issue information of higher content. The issues involved may be separated in topological and processing issues. Despite the interconnection between both issues in an integrated fusion system design, they can be decoupled from each other in order to facilitate the development of a systematic methodology of analysis and synthesis of a fusion system (Thomopoulos 1990, 1991).

The topological issues address the problem of spatial distribution of sensors, the communication network and issues for the exchange of information, the availability and reliability of information at the time of the fusion. The cost of acquiring the information may also be relevant to the topological issues. In remote sensing, these issues are partly addressed by the space agencies and by the image vendors. It is also partly addressed by the customer, given its objectives and constraints, including the financial budget.

The processing issues address the question of how to fuse the data, *i.e.* select the proper measurements, determine the relevance of the data to the objectives, select the fusion methods and architectures, once the data are available.

5. CONCLUSION

Needs expressed by the remote sensing community in Europe have led to the creation of a SIG on data fusion. This SIG has tackled the problems of terms of reference. A new definition of the data fusion is now proposed which emphasises the concepts and the fundamentals in remote sensing.

Several other terms are also proposed which for most of them are already widely used in the scientific community, especially that dealing with information. These terms of reference will be published on the Web site (www-datafusion.cma.fr) of the SIG.

Besides ensuring the communication between its members and the dissemination of information, the SIG is now undertaking an inventory of methods and tools, and is also thinking about instruments for the assessment of the quality in data fusion.

6. REFERENCES

Bijaoui, A., 1981. *Image et segmentation. Introduction au traitement numérique des images*. Masson, Paris, 242 pp.

Buchroithner M., 1998. Geodata interrelations: inventory and structuring attempt of taxonomic diversity. In Proceedings of the 2nd conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, p. 11-15.

DSTO (Defence Science and Technology Organization) Data Fusion Special Interest Group, Data fusion lexicon. Department of Defence, Australia, 7 p., 21 September 1994.

Kanal, L. N., and Rosenfeld, A., 1981. *Progress in Pattern Recognition*. North-Holland Publ., 391 pp.

Lillesand, T. M., and Kiefer, R. W., 1994. *Remote Sensing and Image Interpretation*. Third edition. John Wiley & Sons, 750 pp.

Mangolini M., 1995. Apport de la fusion d'images satellitaires multicapteurs au niveau pixel en télédétection et photo-interprétation. Thèse de Doctorat, Université Nice - Sophia Antipolis, France, 174 p.

Pohl C., 1996. Geometric aspects of multisensor image fusion for topographic map updating in the humid tropics. Doctorate dissertation, Pub. # 39, International Institute for Aerospace Survey and Earth Sciences, Enschede, Netherlands, 168 p., 2 maps.

Thomopoulos, S. C. A., 1990. Sensor integration and data fusion. *Journal of Robotic Systems*, 7, 337-372.

Thomopoulos S. C. A., 1991. Decision and evidence fusion in sensor integration. *Advances in Control and Dynamic Systems*, Ed. C. T. Leondes, vol. 49, part 5, 339-412, Academic Press.

Tou, J. T., and Gonzalez, R. C., 1974. *Pattern Recognition Principles*. Addison-Wesley Publ., 377 pp.

U.S. Department of Defence, 1991. Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C^3, Data fusion lexicon.

Van Genderen, J., and Pohl, C., 1994. Image fusion: issues, techniques and applications. In Proceedings of the EARSeL workshop on intelligent image fusion, Genderen J. van and Cappellini V. Eds.

Wald, L., 1997. An overview of concepts in fusion of Earth data. *In* Proceedings of the EARSeL General Assembly, Lyngby, June 1997.

Wald L., 1998. Data fusion: a conceptual approach for an efficient exploitation of remote sensing images. In Proceedings of the 2nd conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images", Thierry Ranchin and Lucien Wald Editors, published by SEE/URISCA, Nice, France, p. 17-23.

Wald, L., and Ranchin, T., 1995. Fusion of images and raster-maps of different spatial resolutions by encrustation: an improved approach. *Computers, Environment and Urban Systems*, 19(2), 77-87.