Abstract. The paper outlines the principles and methods of data fusion in Earth observations. Here, observations of a particular region of the Earth’s surface at a certain instant, obtained at different wavelengths or using different sensors, are combined in order to derive more useful information than would be obtained from a single source.

Keywords: data fusion, definitions, terms of reference, architecture

1. Definitions

1.1. Introduction

The quantity of information available to describe our environment is increasing rapidly. Archives are growing, as are the number of space missions devoted to Earth observation. Many observation systems are presently available, including space-borne, imaging or not, optical or radar, which provide various measurements, partly redundant, partly complementary. Data fusion is a subject becoming increasingly relevant as scientists try to extract more and more information from these measurements. Indeed, 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 expense of additional independent data delivered by sensors. Data fusion allows us to formalise the combination of these measurements, as well as to monitor the quality of information in the course of the fusion process.

Data fusion is a recent concept; it means an approach to information extraction spontaneously adopted in several domains. However, the operation by itself is not new in remote sensing: classification procedures have been performed for a long time and are obviously relevant to data fusion. Data fusion gathers together 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 data fusion process for a specific application may be tailored to that specific case.

A formal framework is mandatory for a better understanding of the fundamentals and properties of data fusion to allow a better description and formalisation of
the potential synergy between the remote sensing data, and accordingly, a better exploitation of these data.

1.2. DEFINITION OF DATA FUSION

Data fusion is a formal framework in which are expressed the means and tools for the bringing together 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.

Data fusion is exploited by a large number of biological systems. An illustration is given by the human system which calls upon its different senses to perceive its environment. Acquired information is fused within the brain, which will use its memory, its experience, a priori knowledge and its reasoning capabilities to perform deductions, and to produce a representation of the environment.

Data fusion is, for example, used to improve results from a classification, or control laws and their robustness. It is applied in various domains, ranging from image processing in medicine to management and control of industrial processes. In data fusion, the information may be of various kinds, ranging from measurements to verbal reports. This statement illustrates the difficulties encountered in data fusion. Some data cannot be quantified; their accuracy and reliability may be difficult to assess. In Earth observations, we may use some features held in a Geographic Information System (external knowledge) to help in classifying multispectral images provided by several sensors. In this particular case, some data are measurements of electromagnetic energy, and others may be symbols.

In the definition given above, quality does not have a very specific meaning. It is a generic word denoting that the resulting information is more satisfactory for the “customer” than that available without the fusion process. For example, better quality may be an increase in accuracy of a geophysical parameter or of a classification. It may also be related to the production of more relevant information of increased utility, or to the robustness in operational procedures. Greater quality may also mean a better coverage of the area of interest, or a better use of financial or human resources allotted to a project.

This paper mostly deals with the fusion of data from sensors, also called sensor fusion. In this case, the information to be fused are acquired by sensors that can be described precisely. Image fusion is a sub-class of sensor fusion.

1.3. OTHER DEFINITIONS

According to this definition, the different spectral channels of one sensor are to be considered as different sources, as well as images taken at different times. Hence, any processing of the data acquired by the same sensor is relevant to the data fusion domain. Examples in Earth observation are the classification of multispectral imagery, computation of the NDVI (normalised difference vegetation index), or atmospheric correction of spectral bands using other bands of the same sensor.