Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion*

Véronique CLÉMENT, Gérard GIRAUDON and Stéphane HOUZELLE

INRIA, Sophia Antipolis, BP109, F-06561 Valbonne - Tel.: (33) 93657857
Email: vclement@sophia.inria.fr giraudon@sophia.inria.fr houzelle@sophia.inria.fr

Abstract. This paper presents a scene interpretation system in a context of multi-sensor fusion. We present how the real world and the interpreted scene are modeled; knowledge about sensors and multiple views notion (*shot*) are taken into account. Some results are shown from an application to SAR/SPOT images interpretation.

1 Introduction

An extensive literature has grown since the beginning of the decade on the problem of scene interpretation, especially for aerial and satellite images [NM80,Mat90] [RH89] [RIHR84] [MWAW89] [HN88] [Fua88] [GG90]. One of the main difficulties of these applications is the knowledge representation of objects, of scene, and of interpretation strategy. Previously mentioned systems use various knowledge such as: object geometry, mapping, sensor specifications, spatial relations, etc...

In the other hand, there is a growing interest in the use of multiple sensors to increase both the availability and capabilities of intelligent systems [MWAW89,Mat90] [LK89] [RH89]. However, if the multi-sensor fusion is a way to increase the number of measures on the world by complementary or redundancy sensors, problems of control of the data flow, strategies of object detection, and modeling of objects and sensors are also increased.

This paper presents a scene interpretation system in a context of multi-sensor fusion. We propose to perform fusion at the intermediate level because it is the most adaptive and the most general for different applications of scene analysis. First, we present how the real world and the interpreted scene are modeled; knowledge about sensors and multiple views notion (*shot*) are taken into account. Then we give an overview of the architecture of the system. Finally, some results are shown from an application to SAR/SPOT images interpretation.

2 Modeling

Consistency of information is one of the relevant problems of multi-sensor fusion systems; in fact, various models must be used to express the *a priori* knowledge. This knowledge can be divided into knowledge about the real world and knowledge about the interpretation.

2.1 Real World Modeling

For an interpretation system, *a priori* knowledge on the scene to be observed is necessary: for example, the description of objects which might be present in the scene. Moreover,

* This work is in part supported by AEROSPATIALE, Department E/ETRI, F-78114 Magny-les-hameaux, and by ORASIS contract, PRC/Communication Homme-Machine.

in order to perform multi-sensor fusion at different levels of representation, and to use the various data in an optimal way, characteristics of available sensors have to be taken into account: this allows the selection of the best ones for a given task. In the following, we first develop object modeling, then sensor modeling.

Objects to Detect: Usually, in single-sensor systems, two main descriptions are used: the geometric description, and the radiometric one. These two criteria can be used to detect an object (by allowing the choice of the best-adapted algorithm, for instance), or to validate the presence of an object (by matching the computed sizes with the model sizes, for example).

In a multi-sensor system, the distinction must be made between knowledge which is intrinsic to an object, and knowledge which depends on the observation. Geometric properties can be modeled on the real world, however geometric *aspects* have to be computed depending on the sensor. Concerning radiometric properties, there is no intrinsic description; radiometric descriptions are sensor-dependent. In fact, only the observation of an object can be pale or dark, textured or not. Thus, the notion of material has been introduced in our system to describe an object intrinsically. Materials describe the composition of an object: for example, a bridge is built of metal, cement and/or asphalt. So, radiometric properties of an object can be deduced from its composition: an object mainly made of cement, and another one mainly made of water would not have the same radiometry in an image taken by an infra-red sensor. These criteria (geometry, composition) which are only descriptions of objects can be used in a deterministic way.

Another sensor-independent knowledge very important in human interpretation of images is spatial knowledge, which corresponds to the spatial relationships between objects. Spatial knowledge can link objects of the same kind, as well as objects of different kinds. This heuristic knowledge can be used to facilitate detection, validation, and solving of the conflicts among various hypotheses. For example, as we know that a bridge will be over a road, a river, or a railway, it is not necessary to look for a bridge in the whole image; the search area can be limited to the roads, rivers, and railways previously detected in the scene. In multi-sensor interpretation, we can even detect the river on one image, and look for the bridges on another one.

Sensors: Some sensors are sensitive to object reflectance, other to their position, or to their shape... Radiometric features mainly come from the materials the objects are composed of, and more precisely from *features* of these materials such as *cold, homogeneous, rough, textured, smooth....* The response to each aspect is quite different depending on the sensor.

Therefore sensors are modeled in our system using the sensitivity to aspects of various materials, the sensitivity to geometry of objects, the sensitivity to orientation of objects, the band width described by minimum and maximum wave length, and the type (active or passive). Note that the quality of the detection (*good, medium,* or *bad*) has been dissociated from the aspect in the image (*light, grey, dark*).

Due to their properties, some objects will be well detected by one sensor, and not by another one; other objects will be well detected by various sensors. To be able to detect easily and correctly an object, we have to choose the image(s), i.e. the sensor(s), in which it is best represented. For that, our system uses the sensitivities of the sensors, and the material composition of the objects.

Knowing the position of the sensor, and its resolution is also important to be able to determine whether an object could be well detected. We call *shot* the whole information:

the description of the sensor, the conditions of acquisition including the point of view, and the corresponding image.

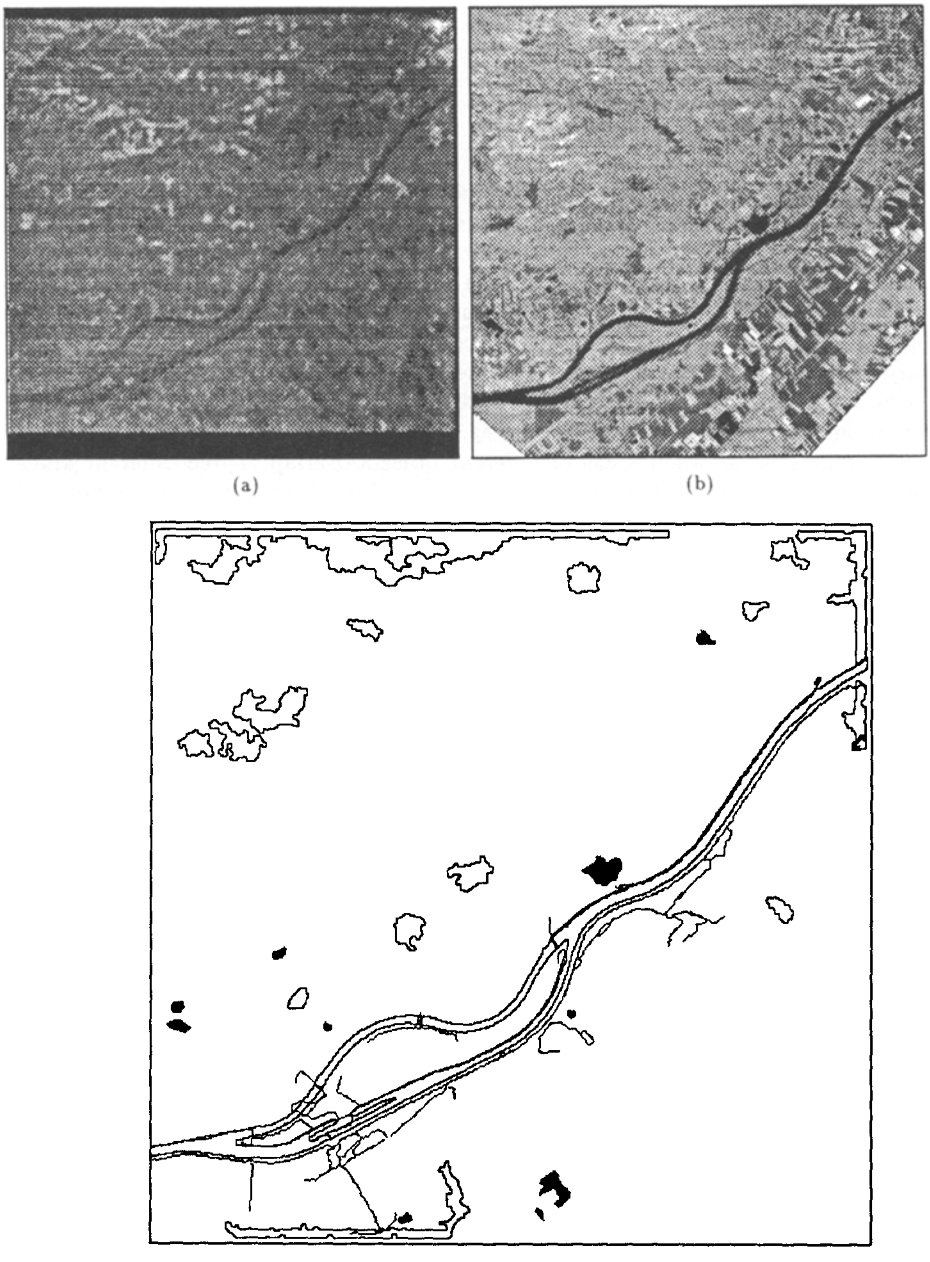
2.2 Interpretation

The main problem is how to represent the scene being interpreted. First of all, we are going to precise what we call an interpreted scene, and which information must be present in an interpretation. Our goal is not to classify each pixel of the image; it is to build a semantic model of the real observed scene. This model must include: the precise location of each detected object, its characteristics (such as shape, color, function...), and its relations with other objects present in the scene. To capture such information, it is necessary to have a spatial representation of the scene; in the 2D-case, this can be done using a location matrix. This representation allows to focus attention on precise areas using location operators such as *surrounded by, near...*, and to detect location conflicts. Location conflicts occur when areas of different objects overlap. Three different kinds of conflicts can be cited: conflicts among superposed objects (in fact, they are not real conflicts: a bridge over a road); conflicts among adjacent objects (some common pixels; such a conflict is due to low level algorithms, digitalization...); conflicts arising because of ambiguous interpretation between different sorts of objects (this kind of conflict can be elucidated only by using relational knowledge).

3 Implementation

Our goal was to develop a general framework to interpret various kinds of images such as aerial images, or satellite images. It has been designed as a shell to develop interpretation systems. Two main knowledge representations are used: frames and production rules. The system has been implemented using the SMECI expert system generator [II91], and the NMS multi-specialist shell [CAN90]; it is based on the blackboard and specialist concepts [HR83]. This approach has been widely used in computer vision [HR87,Mat90], and in multi-sensor fusion [SST86]. We have simplified the blackboard structure presented by Hayes-Roth, and we have build a centralized architecture with three types of specialists: **the generic specialists** (application-independent), **the semantic object specialists** (application-dependent), and **the low level specialists** (dependent on image processing, and feature description). They work at different levels of representation, are independent, and work only on a strategy level request; so the system is generic and incremental. The detection strategy is based on the fundamental notion of spatial context linking the objects in the scene, and the notion of salient object.

To demonstrate the reliability of our approach, we have implemented an application for the interpretation of SAR images registered with SPOT images, a set of sensors which are complementary. Five sensors (the SIR-B Synthetic Aperture Radar, and the panchromatic, XS1 [blue], XS2 [green], XS3 [near infra-red] SPOT sensors), ten materials (water, metal, asphalt, cement, vegetation, soil, sand, rock, snow, and marsh), and five kinds of semantic objects (rivers, lakes, roads, urban areas, and bridges) are modeled in this application. We present Figure 1 an example of result (fig 1.c) we obtained using three images: SAR (fig 1.a), SPOT XS1, and SPOT XS3 (fig 1.b). Closed contours point out urban areas. Filled regions indicate lakes. Thin lines represent bridges, roads, and the river. More details about this application can be found in [CGH92], while low-level algorithms are described in [HG91].



(a) (b) (c)

Fig. 1. Top : Sensor images used for scene interpretation : (a) SIR-B Radar image; (b) near infra-red SPOT XS3 image. Bottom : (c) Objets detected in the scene after interpretation. Closed contours point out urban areas. Filled regions indicate lakes. Thin lines represent bridges, roads, and the river.

4 Conclusion

We have proposed a way to model real world and interpreted scene in the context of multi-sensor fusion. *A priori* knowledge description includes the characteristics of the sensors, and a semantic object description independent of the sensor characteristics. This architecture meets our requirements of highly modular structure allowing easy incorporation of new knowledge, and new specialists. A remote sensing application with SAR/SPOT sensors aiming at detecting bridges, roads, lakes, rivers, and urban areas demonstrates the efficiency of our approach.

Acknowledgments: The authors would like to thank O.Corby for providing useful suggestions during this study, and J.M.Pelissou and F.Sandakly for their contribution in the implementation of the system.

References

[CAN90] O. Corby, F. Allez, and B. Neveu. A multi-expert system for pavement diagnosis and rehabilitation. *Transportation Research Journal*, 24A(1), 1990.

[CGH92] V. Clément, G. Giraudon, and S. Houzelle. A knowledge-based interpretation system for fusion of sar and spot images. In *Proc. of IGARSS*, Houston, Texas, May 1992.

[Fua88] P. Fua. Extracting features from aerial imagery using model-based objective functions. *PAMI*, 1988.

[GG90] P. Garnesson and G. Giraudon. An image analysis system, application for aerial imagery interpretation. In *Proc. of ICPR*, Atlantic City, June 1990.

[HG91] S. Houzelle and G. Giraudon. Automatic feature extraction using data fusion in remote sensing. In *SPIE Proceedings, Vol 1611, Sensor Fusion IV: Control Paradigms and data structures*, Boston, November 1991.

[HN88] A. Huertas and R. Nevatia. Detecting building in aerial images. *ICGCV*, 41.2:131–152, February 1988.

[HR83] B. Hayes-Roth. The blackboard architecture : A general framework for problem solving? *Stanford University, Report HPP-83-30*, 1983.

[HR87] A. Hanson and E. Riseman. The visions image-understanding system. In C. M. Brown, editor, *Advances in Computer Vision*, pages 1–114. Erlbaum Assoc, 1987.

[II91] Ilog and INRIA. *Smeci 1.54 : Le manuel de référence.* Gentilly, 1991.

[LK89] R. Luo and M. Kay. Multisensor integration and fusion in intelligent systems. *IEEE Trans on Sys. Man and Cyber.*, 19(5):901–931, October 1989.

[Mat90] T. Matsuyama. *SIGMA, a Knowledge-Based Aerial Image Understanding System.* Advances in Computer Vision and Machine Intelligence. Plenum, New York, 1990.

[MWAW89] D.M. McKeown, Jr. Wilson, W. A.Harvey, and L. E. Wixson. Automating knowledge acquisition for aerial image interpretation. *Comp. Vision Graphics and Image Proc*, 46:37–81, 1989.

[NM80] M. Nagao and T. Matsuyama. *A Structural Analysis of Complex Aerial Photographs.* Plenum, New York, 1980.

[RH89] E. M. Riseman and A. R. Hanson. Computer vision research at the university of massachusetts, themes and progress. *Special Issue of Int. Journal of Computer Vision*, 2:199–207, 1989.

[RIHR84] G. Reynolds, N. Irwin, A. Hanson, and E. Riseman. Hierachical knowledge-directed object extraction using a combined region and line representation. In *Proc. of Work. on Comp. Vision Repres. and Cont.*, pages 238–247. Silver Spring, 1984.

[SST86] S. Shafer, A. Stentz, and C. Thorpe. An architecture for sensor fusion in a mobile robot. In *Int. Conf. on Robotics and Automation*, pages 2202–2011, San Francisco, June 1986.