Chapter 10
High-Level Information Fusion in Visual Sensor Networks

Juan Gómez-Romero
University Carlos III of Madrid, Spain

Jesús García
University Carlos III of Madrid, Spain

Miguel A. Patricio
University Carlos III of Madrid, Spain

José M. Molina
University Carlos III of Madrid, Spain

James Llinas
University at Buffalo, USA

ABSTRACT

Information fusion techniques combine data from multiple sensors, along with additional information and knowledge, to obtain better estimates of the observed scenario than could be achieved by the use of single sensors or information sources alone. According to the JDL fusion process model, high-level information fusion is concerned with the computation of a scene representation in terms of abstract entities such as activities and threats, as well as estimating the relationships among these entities. Recent experiences confirm that context knowledge plays a key role in the new-generation high-level fusion systems, especially in those involving complex scenarios that cause the failure of classical statistical techniques—as it happens in visual sensor networks. In this chapter, we study the architectural and functional issues of applying context information to improve high-level fusion procedures, with a particular focus on visual data applications. The use of formal knowledge representations (e.g. ontologies) is a promising advance in this direction, but there are still some unresolved questions that must be more extensively researched.

DOI: 10.4018/978-1-61350-153-5.ch010
INTRODUCTION

This chapter provides an overview of the nature of Information Fusion (IF) as a process, issues regarding the design and implementation of IF systems, and the special functions and algorithmic methods typically employed in IF processes. In addition, we focus on what is called “High Level Information Fusion” (HLIF), meaning those inferences developed by IF systems that are at a higher level of abstraction, from which the terminology derives.

We argue that there are four categories of information that can be applied to any (IF) problem: observational data, a priori knowledge models, inductively learned knowledge, and contextual information. For a broad class of applications, many IF processes and systems have been designed to work largely on the first two types of information; these are the class of systems built on a deductive-model foundation and that largely employ real-time observational data (obtained from various sources, usually sensors or instrumentation) in a scheme to more or less match the data against the models. These approaches can work well for what could be called well-behaved and well-studied problem domains but cannot be expected to work in problems where the “world-behavior” is very complex and unpredictable. Context can be defined as the set of circumstances surrounding a situation of interest that are potentially of relevance to its completion (Henricksen, 2003). In some applications, these contextual influences are not only important, but they may even be critical to understanding and interpretation, and they need to be considered.

One such type of applications are those involving video surveillance and monitoring applications—where both complex and unpredictable behavior can be expected, and where the contextual physical environment can be a prime driver or constraint to such behavior and to the system observational capability. Video applications often occur within a networked system framework, employing several cameras or other visual devices that are optimally emplaced and connected by data and control links to form a sensor network. Thus, IF techniques are required to collect, fuse, and interpret visual data. Hence, we comment briefly on the general issues in defining, designing, and implementing IF in general sensor networks, and particularly, in Visual Sensor Networks (VSNs).

In addition to describing the nature of the IF process, we also elaborate on the four informational and knowledge elements typically used to develop the IF approach, as were mentioned above. It is important we feel to understand the opportunities and constraints related to the employment of each of the components. Because of our assertion that Contextual Information is a critical informational component for video and vision applications, we further elaborate on our interpretation of the usage modes for such information.

We then focus on a detailed description of a Computer Vision application and describe issues in defining an architectural framework, strategies for dealing with contextual information, and reasoning methods for scene understanding. It is “Scene Understanding” in these applications that is the result of the HLIF process that we describe. The notion of Scene Understanding is application-dependent and must be defined in each case, but such notions usually imply a complex, high-dimensional state of the world that is to be estimated. In the most abstract definition, it can be described as a set of Entities in a set of Relationships. In the application spaces of usual interest, the Entities can be physical objects, events, and behaviors, and the relations can be of wide description. Thus, the notion of what is meant by a “Scene” is in fact a concept at a high level of abstraction, and assembling the component estimates into a Scene state estimate is a complex process. Accordingly, we propose the use of ontologies as a proper formalism to represent and reason with scene Entities and Relationships in the evolution from low-level acquired data to high-level scene descriptions. We also review