Visual Information Processing in Wireless Sensor Networks:

Technology, Trends, and Applications

Li-minn Ang
University of Nottingham (Malaysia Campus), Malaysia

Kah Phooi Seng
Sunway University, Malaysia

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Acquisitions Editor: Erika Gallagher
Development Manager: Joel Gamon
Development Editor: Michael Killian
Book Production Manager: Sean Woznicki
Typesetters: Mackenzie Snader and Milan Vracarich
Print Coordinator: Jamie Snavely
Cover Design: Nick Newcomer

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Visual information processing in wireless sensor networks: technology, trends, and applications / Li-minn Ang and Kah Phooi Seng, editors.
p. cm.
Summary: “This book provides a central source of reference on visual information processing in wireless sensor network environments and its technology, application, and society issues”-- Provided by publisher.
Includes bibliographical references.
ISBN 978-1-61350-153-5 (hardcover) -- ISBN 978-1-61350-154-2 (ebook) -- ISBN 978-1-61350-155-9 (print & perpetual access) 1. Wireless sensor networks. 2. Optical detectors. I. Ang, Li-minn, 1971- II. Seng, Kah Phooi, 1974-
TK7872.D48V57 2012
681’.2--dc23

2011026269

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

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

some ontology-based approaches to HLIF, with a particular focus on visual HLIF.

The chapter is organized as follows. First, we overview the nature of Information Fusion and introduce the JDL process model, the canonical model to describe Information Fusion systems. In this section, we also present the challenges of Distributed Information Fusion in sensor networks. Next, we specifically focus on Visual Sensor Networks and describe the tasks that need to be faced in this kind of sensor networks from an Information Fusion perspective, as well as some previous approaches and techniques to solve them. In section 4, we discuss on the four categories of knowledge that can be applied in developing any IF approach. In section 5, we focus on the definition and the role of contextual information in HLIF; more precisely, we depict two frameworks for "a priori" and "a posteriori" exploitation of context. In section 6, we propose an architecture and a knowledge representation model for high-level context-based Information Fusion in Visual Sensor Networks. We also explain the abductive reasoning processes that are performed to achieve scene understanding from perceived and contextual information. These aspects are illustrated with some examples in a surveillance scenario. The chapter concludes with a section on the research issues related to contextual exploitation in sensor network-based IF applications. This section offers some ideas for a research agenda in this area.

INFORMATION FUSION IN SENSOR NETWORKS

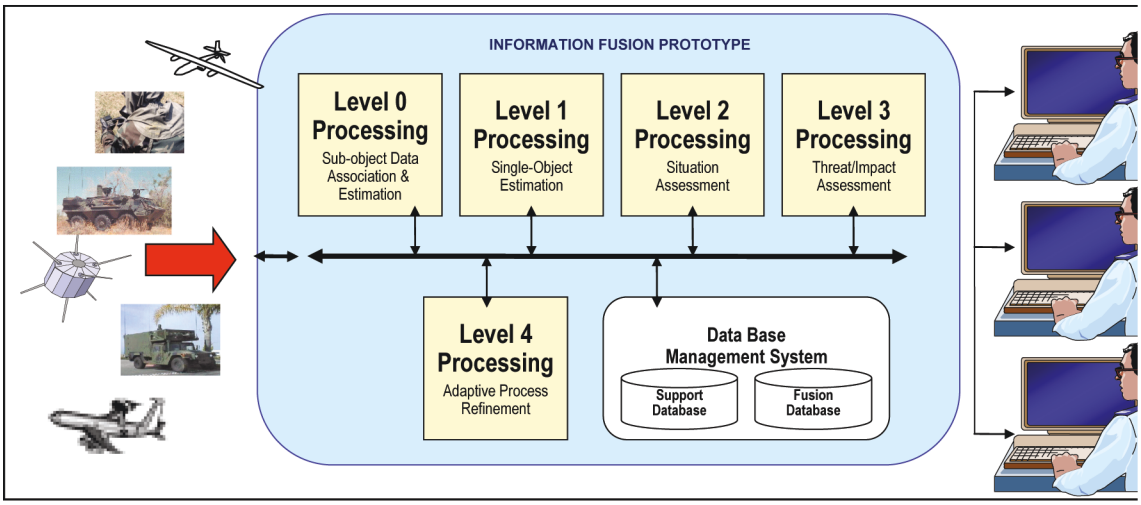
Information Fusion: A Short Overview

Information fusion (IF) techniques combine data from multiple sensors, and other information and knowledge, to achieve improved accuracies and more specific inferences or estimates than could be achieved by the use of a single sensors or information sources alone[1]. It is important to understand thus that no matter the methods and sophistication of the IF process that its output is an imperfect estimate with some degree of uncertainty (but hopefully less than the estimation uncertainty of other possible techniques). The concept of multisensor IF is hardly new. Humans and animals have evolved the capability to use multiple senses to improve their ability to survive. Applications for multisensor IF are widespread. Military applications include: automated target recognition (e.g., for smart weapons), guidance for autonomous vehicles, remote sensing, battlefield surveillance, and automated threat recognition systems, such as identification-friend-foe- neutral (IFFN) systems.

The most fundamental characterization of IF involves a layered transformation between the observed energy or parameters related to an observable entity of some type (this energy provided by multiple sensors as input), and an estimate or inference (produced by fusion estimation and/or inference processes) regarding the location, characteristics, and identity of an entity, and an interpretation of the estimated entity in the context of a surrounding environment and relationships to other entities (Hall & Llinas, 1997). The definition of what constitutes an entity depends upon the specific application under consideration (e.g., an enemy aircraft for a tactical air-defense application, or the location and characteristics of a tumor in a medical diagnosis application).

Despite the benefits of improved system operation and application effectiveness by using multiple sensors and fusion processes, actual implementation of effective IF systems is far from simple. In many or most cases, the fusion process is a value-adding process that is often added to or integrated into some overarching system concept; it is often said that "we don't build fusion systems; we add fusion processes into systems". This implies that there may be existing, legacy functions and components, existing developmental

Figure 1. The JDL Functional Model of the Information Fusion Process. (Adapted with permission fromLiggins, Hall, & Llinas, 2009)



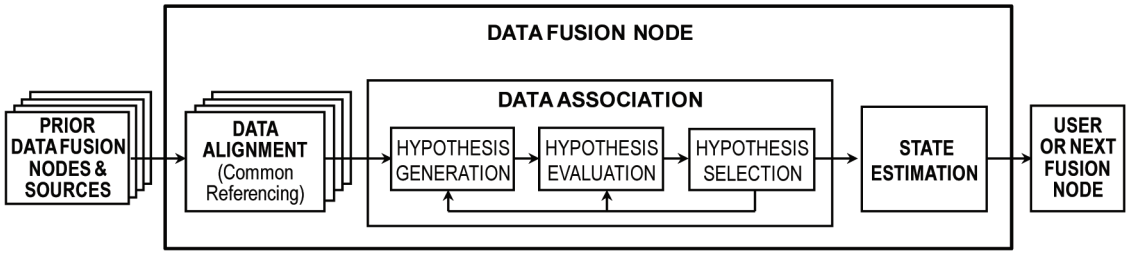
standards, design standards, etc. that the fusion process must adapt to. It also imposes a requirement to assess the added benefits of implementing a fusion process via some type of comparative analysis against some baseline system that may not incorporate fusion, or incorporates a "lesser" degree of fusion (e.g. fusion with fewer sensors), to assess the cost-benefit.

One of the descriptive, functional models of the information fusion process is the JDL[2] process model that is intended to be very general and useful across multiple application areas (Liggins, Hall, & Llinas, 2009); this process is shown in Figure 1.

While the boundaries of the IF process are fuzzy and case-dependent, generally speaking the input boundary is usually at the post-detection, extracted parameter level of signal or observational processing. The output of the IF process is (ideally) a minimally ambiguous identification and characterization (viz., location and attributes) of individual entities, as well as a higher level interpretation of those entities in the context of the application environment. As noted previously, a good way to think about the process is as an estimation process that needs to address the stochastic aspects of the various error-generating factors of any observational process or of other amplifying data, to include errors and inaccuracies in contextual information. As shown, the process conceptually involves several "Levels" of processing; these Levels were conceptualized simply to distinguish different types of complexity and abstraction that may exist in any problem, and (an important point) that any complete fusion process will involve an integrated set of functioning solution techniques. It is important to understand that Figure 1 is not an architecture but a pedagogical diagram to convey the basic concepts of IF processing. The basic Levels involve:

- **Level 0 Processing (Sub-object Refinement).** This process involves those methods to estimate the existence and features of structures of interest that may be discernible before the declaration of a named entity can be realized (e.g., segmentation in imagery)
- **Level 1 Processing (Object Refinement).** This process combines locational, parametric, and identity information to achieve

Figure 2. Example of Fusion Node



refined representations of individual objects or entities (e.g., time histories (tracks) of the kinematic properties of an object, i.e., location, velocity, etc.)

- **Level 2 Processing (Situation Refinement (similar to Scene Understanding)).** Level 2 processing develops a description of current relationships among objects and events in the context of their environment (e.g., a cyber-security related situation such as "coordinated virus attack on sub-network X")
- **Level 3 Processing (Threat or Impact Refinement).** Level 3 processing focuses on what could be called special situations that relate to some type of threatening, critical, or otherwise special world states (situations, scenes, etc.) that are of interest for awareness or for decision and action-taking. Another way to think about Level 3 is developing a fused estimate of the Impact of that special situation to the application of interest.
- **Level 4 Processing (Process Refinement).** Level 4 processing may be considered a meta-process, i.e., a process concerned about other processes; it can be thought of as the "control law" of the overall fusion process, concerned with invoking any feasible capability needed to improve the process (one example is to control the sensing or observational process in ways that improve the fused estimate that is the output of the process).

Another way to understand the IF process is by consideration of a typical "Fusion Node". An example is shown in Figure 2.

Note that the Fusion Node (FN) nominally accepts either sensor type input from some input source or an estimate (fused or otherwise formed) from some prior FN or processing node. Other variations could be possible such as control inputs or inputs from a human operator, etc. but for this discussion we assert that the inputs are either sensor data (measurements) or fused estimates from a prior FN; the data are batched according to some batching logic (not shown here). In this characterization, the FN processing operations involve three basic functions:

1. **Data alignment (also known as Common Referencing):** Here, a variety of what could be called normalization operations are performed, ranging from the simple such as coordinate or units transformations or more complex such as uncertainty transformations; the goal of such operations is to establish common representational and semantic forms for the composite input.
2. **Data association:** Here, the multiple inputs of either estimates or measurements are examined in order to determine which (hypothetical) entity that the system believes to exist they are associated to or come from. By gauging this, the association process culminates in assigning (partitioning) the data to the on-going fusion estimation processes for each such entity. Data association

comprises three sub-functions as shown; the term "hypothesis" shown here should be understood as an association hypothesis, that is one feasible association of the measurements or estimates to the entities for which state estimates are being formed.

a. **Hypothesis Generation:** This is a step that nominates the possible causes of the measurements or the estimates coming in; an example for a measurement would be a False Alarm, i.e., establishing that in the IF designer's view, it is plausible that a measurement is not valid (from a real entity) but is an artifact. Note that it is only the nominated hypotheses to which associations can be made.
b. **Hypothesis Evaluation:** This is a step where the notion of "closeness", of say a measurement to its expected or predicted value, is judged by the computation of some association score or metric. This operation has to take into account both any measurement errors as well as any prediction errors, and so it is typical that these metrics are of a statistical nature.
c. **Hypothesis Selection:** in this final step, the information at hand can be conceptualized as a matrix of say labeled measurements and labeled association hypotheses, with the matrix populated by the "association scores" just computed in the Hypothesis Evaluation step. The problem is to determine the optimal allocation of the measurements to the hypothesized entities. In operations research, such optimization problems are called "assignment" problems since the challenge is to find an optimal assignment of the measurements. (The fact that scores have been computed for each measurement to entity pair does not necessarily lead to an obvious solution since there may be other measurements competing for the same assignment; this is a problem in combinatorial optimization, to find the best assignment that yields the best composite or measurement-set score.)

What really happens in effect at the end of Data Association is that the measurements are assigned to fusion-capable estimation algorithms that can then "digest" the multiple inputs to form an updated, fused, improved estimate of whatever is being estimated for that entity. Consider multiple frames of imagery that are being used to generate a fused estimate of an object A's shape. In the above operations, it would have been determined that certain segments of multiple images are best associated to object A, and those composite segments would be given to a shape-estimating algorithm to generate the fused shape estimate.

3. **State estimation:** This stage is just that, the step where whatever type of estimate—often about entity attributes at fusion Level 1 (e.g., kinematic properties such as location, velocity, and classification attributes such as color, etc. or a direct estimate of identity)—or about the existence of events or of certain behavioural activity of the estimates are of a Level 2 type.

Sensor Networks and Information Fusion

In addressing any topic incorporating the phrase "sensor network", one is faced with the challenge of specifying the type(s) of network being addressed. While literature search seems to reveal no standard taxonomy of sensor-network types, there are these variations mentioned:

- **Mobile Ad Hoc Networks:** A mobile ad hoc network (MANET) is a kind of wireless ad hoc network, and is a self-

configuring network of mobile routers (and associated hosts) connected by wireless links—the union of which forms an arbitrary topology. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Alternatively, a MANET may be considered as a mobile version of an ad hoc network.

- **Wireless Sensor Network:** A wireless sensor network (WSN) is a wireless computer network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants, at different locations.
- **Mobile Wireless Sensor Network:** A wireless sensor network in which the nodes are mobile.
- **Ad Hoc Wireless Sensor Network:** A wireless sensor network in which the nodes self-organize.
- **P2P Network:** A network without the notion of clients or servers, but only equal peer nodes that simultaneously function as both clients and servers.

In each flavor of network there are different system parameters that can affect the strategy for and viable realizations of different HLIF fusion processes. The topological dynamics of MANETs can possibly influence what is observable in the world, the bandwidth limitations of WSNs can affect what is communicable between nodes and thus the nature of what is being fused, the mobility and self-organizing factors of Mobile and Ad Hoc nets interplay with observability issues and the nature of what can be aggregated at any node, and the insertion of fusion processes impacts the fundamental homogeneity of the P2P network concept.

To at least give some sense of the nature of IF operations in sensor networks without addressing all of the net-specific factors in these network variations, we take the point of view regarding Distributed IF (DIF). In the most fundamental perspective, consider any Node in some DIF/Network architecture; this Node may only fuse two types of information—that which it "owns" (e.g., the data from local, sometimes called "organic" sensors)—and that which comes to it somehow from the network. The "somehow" is important; a Node can possibly receive information it has requested, or information sent by prior agreement from another Node (in a "Publish/Subscribe" protocol), or according to any "Information-Sharing Strategy (ISS)" that has been developed for the network. There is thus a critical interdependency between whatever the ISS protocol is and what DIF processes at various Nodes can achieve. Note that in the above network variations the feasible ISS protocols may for example be based in part on optimal energy (Node battery) power management concerns; rates and degree of intermodal communications may be energy-bound. Further, ISS´s may also involve non-technical aspects such as imputed by Nodal priorities of some type, or command authority in military applications, etc. As a result, the constraints of any given ISS may result in some degree of compromise in what can be achieved in the corresponding DIF processing operations.

DIF can also be susceptible to what is variously called the "double-counting" problem or "data incest" problem resulting from network topologies that have cycles in them, allowing redundant intermodal messages to flow into any DIF fusion Node, i.e., situations where duplicative evidential data arrives at a DIF fusion Node unbeknownst to that Node (Julier, 2009). Such issues impute requirements for appropriate meta-data tagging so that data "pedigree" can be tracked in the network (Arambel, 2008). Similar effects can be encountered because many IF algorithms are purposely built to be recursive, to deal with the dynamic

world of interest. This means the algorithm outputs have a built-in memory of the effects of prior data. In the same way as just mentioned then, if IF algorithm outputs are shared across the nodes, similar double-counting effects can corrupt the correctness of the fused results at the DIF Node. Depending on the specifics of the network, and in particular what each Node knows about the Nodes from which it can receive information, there may be yet other effects on DIF process design. In large networks for example, it is unreasonable for any DIF Node to have full information about Nodes at far distance from it; in fact many network design strategies have Nodes only knowing Nodal details about their one-hop-away neighbors. This can impute a layered DIF process at any DIF Node, in which there may be at least two IF layers, in essence of different quality, to fuse data from Nodes that the DIF Node knows the details of (thus a notional high-quality fusion process) and to fuse data from Nodes about which it has less information (such as not possibly knowing certain sensor data details).

Another general problem that can happen in virtually any IF system but is more likely in DIF applications is what is sometimes called the "Out of Sequence Measurement" or OOSM problem (Shi, Wan, & Ge, 2009). This is fundamentally the problem for system/process design of how to handle the problem of retrospective corrections to fused estimates when just-arriving data or information reveals new information related to a prior time period. Often, to allow such retrospective corrections to IF-based inferencing, one must choose the largest time window over which such adjustments will be allowed, and then the system will have to store all data across such intervals to allow the current inference to be undone backward in time in accordance with the new information from that prior time point. Such problems are more likely in sensor networks due to intermodal communication delays and other factors involved in ISS management.

So it can be appreciated that what IF can be done in any sensor network can be impacted by a variety of factors that reflect the peculiarities of that network but that in addition there are some generic DIF concerns that apply for all sensor networks.

INFORMATION FUSION IN VISUAL SENSOR NETWORKS

Modern Visual Sensor Networks (VSNs) involve the deployment of a large number of cameras and the management of geographically-spread monitoring points covering a wide area. Although the existence of multiple cameras inevitably increases the complexity of the systems, it also improves the results, since multiple viewpoints can be combined. Additionally, distribution increases system robustness and fault tolerance, since the same information may be captured and replicated at different points of the network. In this state of affairs, IF is not only a convenience but a requirement; local data acquired by distributed video cameras must be combined to obtain a global understanding of the current scenario. Essentially, DIF in VSNs is concerned with the following tasks (Castanedo, García, Patricio, & Molina, 2009):

- Camera calibration, or common referencing, i.e., the process of transforming from the local coordinates of each camera to a global coordinates space. This translation may include a reconstruction step to obtain a 3D representation of the 2D image. The most used methods for camera calibration are those proposed by Tsai (1987), Heikkila (1997), and Zhang (2000).
- Object detection, i.e., to identify interesting objects in the sequence of images provided by the camera. There are various usual approaches to the detection of moving objects, e.g.: temporal differencing

–based on calculating the pixel-by-pixel difference of various consecutive frames (Lipton, Fujiyoshi, & Patil, 1998); background subtraction –based on subtracting the current snapshot pixel values with a predefined background image (Piccardi, 2004); statistical methods –a variation of basic background subtraction method based on the differences of additional statistical measures (Wang, Hu, & Tan, 2003); or optical flow –based on the computation of the flow vectors of moving objects over time (Barron, Fleet, & Beauchemin, 1994).

- Object tracking, i.e., to estimate the number of objects in a continuous scene, together with their instantaneous location, kinematic states, and other characteristics. Object tracking, which is a particular case of state estimation, has been traditionally tackled by applying statistical prediction and inference methods; for instance, widely-used tracking methods in general DIF are distributed MHT (Chong, Mori, & Chang, 1990), distributed JPDA (Chang, Chong, & Bar-Shalom, 1986), Covariance Intersection / Covariance Union (Julier & Uhlmann, General decentralized data fusion with Covariance Intersection, 2001), and distributed Kalman filter (Olfati-Saber, 2007). As we explain below, these methods may be insufficient in VSNs and require the incorporation of additional information and knowledge in the DIF process.
- Object and activity recognition, i.e., to determine the type of an object (e.g., car, human, aircraft) or the type of an activity (e.g., approaching, walking, maneuvering). Recognition can be seen as a probabilistic reasoning problem, in which case it is tackled through probabilistic models (Markov models, Bayesian networks, etc.), as in (Hongeng, Nevatia, & Bremond, 2004); or as a classification problem, in which case it is tackled through pattern recognition techniques (neural networks, self-organizing maps, etc.), as in (Hu, Xie, Tan, & Maybank, 2004).
- Process enhancement, or active fusion, i.e., to implement suitable mechanisms that use the more comprehensive interpretation of the current situation obtained after fusing data to improve the development of the previous tasks. Generally speaking, process enhancement consists in improving a fusion procedure by using feedback generated at a more abstract level; for instance, the behavior of a tracking algorithm can be changed once a general interpretation of the scene has been inferred –if we recognize that an object is moving out of the camera range through a door a object, the tracking procedure can be informed to be ready to delete this track in the near future.

These tasks can be largely classified according to the JDL model depicted in section 2.1. Camera calibration is a necessary task that must be performed before any other procedure, and hence it can be considered as a Level 0 –or even pre-Level 0– process. Depending on the nature of the object detection process, it can be seen as a Level 1 or, more likely, as a Level 0 procedure, since it aims at estimating the existence of an object, but not its properties. Object tracking and recognition are typical Level 1 procedures, given that it is concerned with the estimation of properties at object level. Some authors have discussed the convenience of taking special consideration to the fusion information and procedures between Level 1 and Level 2 processes defining a Level 1-1/2 (Das, 2008), given that the estimation of object properties frequently has a close dependence to the relations with other objects, an idea that partially overlaps with the notion of Level 2 Situation Refinement. This may be the case of object tracking, so this distinction must be noticed. Clearly, activity recognition, in its basic form, is a Level 2 procedure, because it aims at inferring

the relations between objects in terms of a scene interpretation. If this process implies determining if the situation is a threat or not, it can be classified as a Level 3 task. Level 2 and Level 3 fusion are usually used to delimit the boundaries of HLIF. Last but not least, process enhancement is a Level 4 procedure, because it is concerned with providing all the possible means to improve the overall fusion procedure, from data sampling to activity recognition.

System communication and scalability are non-functional requirements of DIF in VSNs. Naturally, distributed nodes of the VSN must be properly communicated, but in several scenarios, communication simply understood as information transmission may not be enough, and coordination and/or cooperation between nodes may be demanded in the form of an ISS protocol. For example, when adjacent regions are watched by different cameras, they can take advantage of information handover when an object moves from one area to another, in which case coordinated tracking plans might be implemented (Castanedo, García, Patricio, & Molina, 2010). Scalability of the global system must be guaranteed with independence of the specific sensors and their configuration, which can be difficult when new and heterogeneous cameras are incorporated to build a large and scattered circuit. The cornerstone of effective ISS protocols in such distributed architectures is the use of common and well-defined communication vocabularies. The use of a common language facilitates component interoperability, thus allowing systems to decouple the internal implementation of sensor and their communication features. Ontologies (see section 4.2) can be successfully used to define de content language of the messages of the VSN, since their sharing and reusing features make them especially suitable to develop flexible and extensible solutions.

The accomplishment of DIF tasks in VSNs is strongly affected by the great variety of situations in which these technologies are being applied. Visual sensors are different from traditional sensors (radar, microphones, etc.), since they provide much more information while need more computational resources. As mentioned, classical statistical methods are insufficient when dealing with complex scenarios where illumination changes, occlusions from other objects or persons, and shadows and reflections appear. These methods are based on very strong statistical assumptions about the video data (e.g., Gaussian linear dynamic models of targets and noise), which unfortunately do not hold in most application domains. In addition, most of them rely on expensive computational costs that may be unaffordable in real time video applications.

The solution to these problems is to incorporate knowledge about the observed entities not directly obtained by the cameras to the image-processing procedures, i.e., to exploit additional information to achieve successful DIF –mainly, context knowledge. This *complementary* knowledge can be conveniently managed with formal representations, such as ontologies, which allow representing and reasoning with well-defined context descriptions. In the next section, we discuss on the types and the nature of the information and knowledge sources that can be incorporated to IF processes, whereas in section 5 we study in detail the role of context knowledge in HLIF. Particularly, we will show in section 6 an approach to an architecture and a framework for HLIF in VSNs based on the use of context knowledge represented with formal ontologies.

EXPLOITATION OF INFORMATION AND KNOWLEDGE IN INFORMATION FUSION

As was commented in the Introduction, there are only four types of information or knowledge that can be applied in developing any IF approach: observational data, a priori knowledge models, inductively learned knowledge, and contextual information. Each of these categories provides a

distinctive type of input or support to the design of an IF process.

Observational Data

Observational data are the fundamental data from the dynamic, changing world of interest as collected from some observational capability, often an electro-mechanical sensor of some type[3]. At the basic level of fundamental physics, this is about either energy emanating from an entity (collected, observed by a passive sensor) or energy reflected from an entity that has been illuminated by an active, emitting sensor. Here we define that these data are about the observable entities in the world that are of interest.

In developing an (automated) multisensor IF system, it is clear that IF process designers need to know sensor characteristics very well, as the observational data are one of the important forms of input to the IF software process (and so, a "garbage-in, garbage-out" system). By and large, electro-mechanical sensors provide attribute and feature data about entities (such as color and shape in vision systems, and kinematic properties from radar), but do not provide *prima facia* evidence about inter-entity relationships. (In fact the philosophical literature argues that relationships are only discerned by a comparative-type processes that examine relative attributes from the entities of interest (Brower, 2001).) This limitation is one factor that makes the high-level fusion inferencing difficult due in part to the combinatorics of feasible relationships among entities and their attributes, comparing the extensive combinatorial set of all entity-entity attribute sets.

There are clearly many sensor parameters and operating characteristics that have to be understood and factored into the way any IF system would exploit such data, such as: resolution, sampling rate, detection performance, space-time pointing capability, etc. Observational conditions (a contextual factor) also have to be known. Critically important in the analysis and understanding of these inputs are the error characteristics of each sensor. This type of information is central to the formation of inter-sensor data association and the computation of the association scores is discussed in section 2.1.

A Priori Knowledge Models and Ontologies

We take the assumption that the world of interest is a dynamically-changing one and that "staring" problems involving the accumulation of observational data over a static world are not of interest. If we are trying to infer/estimate the existence of a "scene" or "situation", a HLIF product, we must understand what temporal resolution we need, since the scenes and situations are always changing. It can be argued in fact that if behaviors are components of scenes/situations, and if it is agreed that behaviors can be defined as actions over time, then placing "a" behavior as part of a scene imputes that scene development consumes time, so that the notion of a "snapshot" of a situation is inconsistent. It should be clear that a layered knowledge/model system will be required to support the layered fusion to the "Level 2" capability of figure 1, since the observational data by and large only provide discrete entity attributes; all of the relational constructs of interest need to be modeled a priori—in terms of their temporal evolution— so that the situational estimate can be formed by, in effect, hypothesis testing of these models against the observational (and other) data. Most IF systems are built on this type of deductive framework although certain environments demand more than just deductive models can provide.

Closely related to but different than models that model dynamical behaviors of interest in a world scene is the development of an ontology for the expected world of interest. Even the notions and definitions of the term ontology have become controversial in the last years in part because there is conflict regarding the degree of formalization required, which can range from an engineering-

based to a philosophically-based viewpoint. We will not engage in this discussion but agree that an appropriate ontology is very often a helpful a priori component of the knowledge upon which an IF system can be developed. Ontologies aid in normalization of semantic meaning which aids in interoperability and specified ontologies can readily be made machine-readable today so that they can be exploited in the IF system if the opportunity arises. The engineering-type definition is "An explicit formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them" (Staab & Studer, 2009), whereas the philosophical definition is "An ontology is a formal explicit specification of a shared conceptualization for a domain of interest" (Gruber, 1993), with all formal philosophical rigor to follow. From the logician's perspective, ontologies are Description Logics (DLs) knowledge bases, i.e., knowledge representations expressed in one of these well-defined subsets of first order logics (Baader, Calvanese, McGuiness, Nardi, & Patel-Schneider, 2003). But as HLIF is concerned with entities and relationships, it should be clear that ontologies have a role to play in HLIF development; much more will be said on this for the vision applications in section 6.

Learned Knowledge

In those cases where the required a priori knowledge for deductively-based HLIF process development cannot be formed—this can be due to a variety of factors—one alternative is to try and excise the knowledge through online machine learning processes operating on the observational and other data. These are procedural and algorithmic methods for discovering relationships among and behaviors of entities of interest. A wide variety of alternative machine learning strategies exists (e.g., see (Bishop, 2004)). Of course, it is likely that some a priori deductive knowledge will be applicable, and it can always be argued that the real-world problem domain may travel out of the bounds of a priori knowledge in any case, so for many applications the best approach, if workable, will be some combination of a priori and runtime-learned knowledge working in cooperation. This of course generates another problem involving dynamic knowledge management (which could be considered a "Level 4" process refinement requirement), as well as possible conflict resolution techniques, among possible other effects. As IF applications increase in sophistication and complexity, another system design question revolves about the role of humans in IF process operations. Clearly there is a cost-benefit tradeoff involved in trying to develop fully-automated algorithmic IF processes for complex problems where the insertion of human intelligence at some point in the process may be a much more judicious choice.

Contextual Information

Context and the elements of what could be called Contextual Information are both concepts that have had a lot written about them –e.g., see (Zimmermann, Lorenz, & Oppermann, 2007) and (Dey, 2001). Definitions abound but we choose one from (Henricksen, 2003), *"The context of a task is the set of circumstances surrounding it that are potentially of relevance to its completion.",* because of its task-relevance; we think here of a fusion or estimating/inferencing task whose "completion" we take to mean the development of a best-possible estimate. We also choose to add the notion of context as background, i.e., not the specific entity, event, or behavior of prime interest but that information which is influential to the formation of a best estimate of these items. So re-phrased we may say this definition as: "The context of an IF estimation task is the set of background circumstances (the circumstances perhaps equal to but minimally derivable from appropriate contextual information) surrounding it that are of potential relevance to its optimal formation". Context can aid inference

development in at least two ways: (1) by providing an independent (from the core IF system data) basis for consistency-checking, i.e., checking that the asserted inference is consistent with the relevant context, and (2) by providing a basis for the amplification of an inference by adding explanatory aspects when the inference is consistent with the context. Context and contextual information are broad notions; the challenge here is to address the question of what (of all that may be available) is relevant to the clarification or improvements of an IF estimate built from the core observational and other core data.

THE ROLE OF CONTEXTUAL INFORMATION IN HIGH-LEVEL INFORMATION FUSION: UNDERSTANDING AND USING CONTEXT

As remarked in the Introduction, we elaborate here on the use of Contextual Information largely because of its criticality in vision/video applications; recent experience has shown that Contextual Information is a critical factor in modern defense applications involving counter-terrorism and counter-insurgency applications. Following the definitions stated in the previous section, contextual information in HLIF is that information that can be said to "surround" a situation of interest in the world. It is information that aids in understanding the (estimated) situation and also aids in reacting to the situation, if a reaction is required. It can be seen as a set of constraints to a reasoning process about a situation; Kandefer and Shapiro (2008) define it this way: "the structured set of variable, external constraints to some (natural or artificial) cognitive process that influences the behavior of that process in the agent(s) under consideration." Contextual Information can be relatively or fully static or can be dynamic, possibly changing along the same timeline as the situation. It is also likely that the full characterization and specification of Contextual Information may not be able to be known at system/algorithm design time, except in very closed worlds.

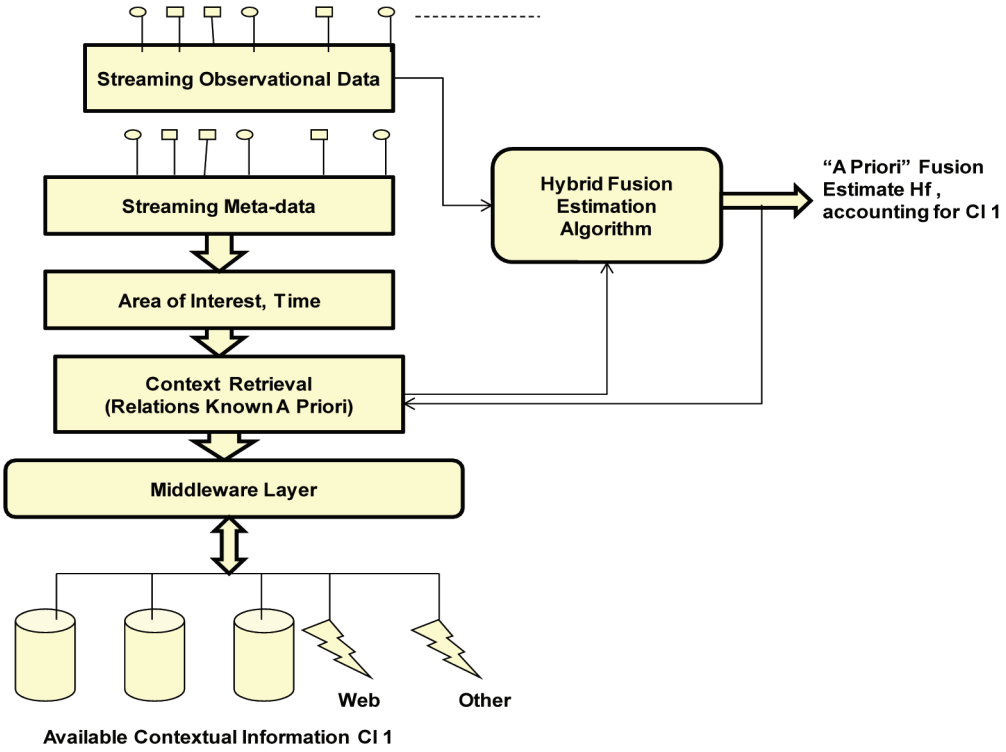
Thus, we envision an "a priori" framework of exploitation of Contextual Information that attempts to account for the effects on situational estimation of that Contextual Information (CI henceforth) that is known at design time; there is a question of the ease or difficulty involved in integrating CI effects into a fusion system design or into any algorithm designs. This issue is influenced in part by the nature of the CI and the manner of its native representation, e.g., as numeric or symbolic, and the nature of the corresponding algorithm. Strategies for a priori exploitation of CI may thus require the invention of new hybrid methods that incorporate whatever information an algorithm normally employs in estimation (usually observational data) with an adjunct CI exploitation process. Note too that CI may, like observational data, have errors and inconsistencies itself, and accommodation of such errors is a consideration for hybrid algorithm design. In this case then, we have a notional processing operation as shown in Figure 3.

Similarly, we envision the need for an "a posteriori" CI exploitation process, due to at least two factors:

1. That all relevant CI may not be able to be known at system/algorithm design time, and may have to be searched for and discovered at runtime, as a function of the current situation estimate.
2. That such CI may not be of a type that was integrated into the system/algorithm designs at design time and so may not be able to be easily integrated into the situation estimation process.

In this case then we envision that at least part of the job of posteriori CI exploitation would be of a type that checks the consistency of a current situational hypothesis with the newly-discovered

Figure 3. Notional Process Flow for "A Priori" CI Exploitation

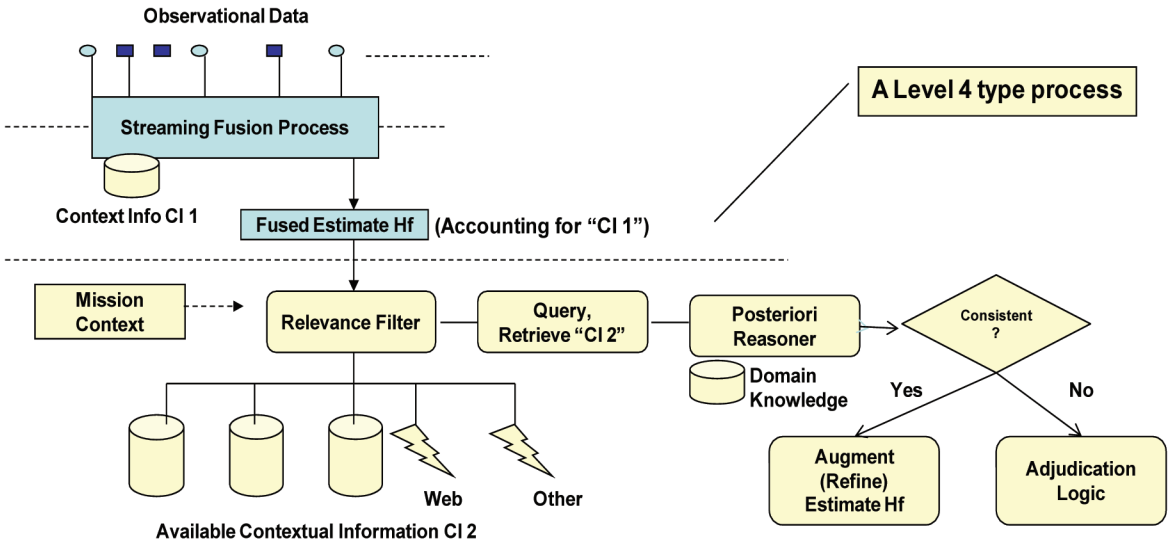


(and situationally-relevant) CI, but also – if the hypothesis is in fact consistent – adds some explanatory aspects to the declared hypothesis.

There are yet other system engineering issues. The first is the question of accessibility; CI must of course be accessible in order to use it, but accessibility may not be a straightforward matter in all cases. One question is whether the most-current CI is available; another may be that some CI is controlled or secure and may have limited availability. The other question is one of representational form. CI data can be expected to be of a type that has been created by "native" users – for example weather data, important in many fusion applications as CI, is generated by meteorologists, for meteorologists (not for fusion system designers) – thus, even if these data are available, there is likely to be a need for a "middleware" layer that incorporates some logic and algorithms to both sample these data and shape them into a form suitable for use in fusion processes of various type. In even simpler cases, this middleware may be required to reformat the data from some native form to a useable form.

In spite of some a priori mapping of how CI influences or constrains the way in which situational inferences or estimates can be developed, which may serve certain environments, the defense and security type applications, with their various dynamic and uncertain types of CI, demand a more adaptive approach. Given a nominated situational hypothesis H_f from a fusion process or "engine" (that might already have accounted for some CI, as explained above), the first question is: what CI type information is relevant to this hypothesis? As cited by Kandefer and Shapiro (2008), "The relevancy problem is defined by Ekbia and Maguitman (2001) as 'the problem of identifying and using properly [only] the information that should exert an influence on our beliefs, goals,

Figure 4. Notional Process Flow for "A Posteriori" CI Exploitation



or plans'." Said otherwise, relevant CI is only that information that influences our interpretation or understanding of H_f. Presuming a "relevancy filter" can be crafted, a search function would explore the available or retrievable CI and make this CI available to a "posteriori" reasoning engine. That reasoning engine would then use a CI-guided subset of Domain Knowledge, and the retrieved CI to reason over H_f to first determine consistency of H_f with the relevant CI. If it is inconsistent, then some type of adjudication logic will need to be applied to reconcile an inconsistency between the fusion process that produced H_f and the posteriori reasoning process that judges it as inconsistent. If however H_f is judged as consistent with the additional CI, an expanded interpretation of H_f could be developed, providing a deeper situational understanding. This processing flow is depicted in figure 4 below. As noted in the figure, this overall process, which can be considered a "Process Refinement" operation, would be a so-called "Level 4" process in the context of the JDL Data Fusion Process Model.

VISUAL SENSOR NETWORK APPLICATIONS: CONTEXT-AWARE SURVEILLANCE SYSTEMS

A prototypical VSN application is surveillance with distributed camera networks. Modern surveillance systems have been named *third-generation surveillance applications*, a term that designates systems that resemble the nature of the human intelligent process of surveillance, which activates certain cognitive abilities, and that satisfy the requirements of modern surveillance, which are among others the management of a large number of cameras, the geographical spread of resources, and the need of many monitoring points (Regazzoni, Ramesh, & Foresti, 2001; Valera & Velastin, 2005). The ultimate goal of third-generation surveillance systems is to automatically achieve a high degree of understanding of the scene from multiple observations to barely require operator attention while cutting component cost. The existence of multiple cameras requires the development of IF procedures to integrate data generated at different locations and reasoning techniques to obtain a high-level and global interpretation of the scene.

CI plays a key role in the development of third-generation surveillance systems. Consistently to the definition presented in the previous sections, we can consider CI in this kind of applications as any external piece of knowledge used to complete the quantitative data about the scene computed by straightforward image-analysis algorithms can be considered context. As in (Kandefer & Shapiro, 2008), CI is therefore an "external constraint" (since it is not directly acquired by the primary system sensors) that "influences the behavior" of the fusion process (since it is used to guide and support DIF). Adapting the characterization by Bremond and Thonnat (1996), four sources of CI must be taken into account in visual DIF: (1) the scene environment: structures, static objects, illumination and other behavioral characteristics, etc.; (2) the parameters of the recording: camera, image, and location features; (3) historic information: past detected events; (4) soft information provided by humans.

This approach to the notion of CI may seem imprecise, but this is intentional. It is especially difficult to clearly distinguish which information is context and which information is perception, either from a computational or a cognitive perspective, in visual applications. Hence, we propose to create a symbolic model that blends perceived and contextual information to obtain an integrated description of the scenario. This knowledge model is encoded with formal ontologies, which are used with two purposes: representing the entities of the domain of interest, i.e., the objects and the relations (contextual and perceived) of the scene; and reasoning with them to perform DIF tasks, i.e., to derive implicit knowledge from the explicit facts to accomplish high-level scene recognition and to provide feedback to improve lower-level image processing procedures. From a practical perspective, this avoids the necessity of explicitly distinguishing between CI and other types of knowledge. In the remainder of this section, we firstly present some related research works based on the use of formal knowledge representations for CI exploitation in information fusion. Secondly, we describe a general architecture for HLIF in VSN and explain how knowledge is represented and exploited in such framework.

Related Work in Ontology-Based CI Exploitation and IF in VSNs

Symbolic models have been applied to acquire, represent, and exploit knowledge in IF, and particularly in visual IF. In the last years, the interest in ontologies has increased considerably (Nowak, 2003), and its use is becoming more and more frequent. Nevertheless, despite the fact that most of the current approaches combine contextual and perceptual information, they do not explicitly describe how context is characterized and integrated in the fusion loop.

Previous ontology-based fusion researches can be classified according to the levels defined by the JDL model. At image-data level (i.e., JDL level 0), one of the most important contributions is COMM (Core Ontology for MultiMedia), an OWL ontology to encode MPEG-7 data (Arndt, Troncy, Staab, Hardman, & Vacura, 2007). Similarly, the Media Annotations Working Group of the W3C is working in an OWL-based language for adding metadata to Web images and videos (Lee, Bürger, & Sasaki, 2009).

Other proposals are targeted at modeling video content at object level (i.e., JDL L1). For example, a framework for video event representation and annotation is described in (François, Nevatia, Hobbs, Bolles, & Smith, 2005). This framework includes two languages, namely VERL (Video Event Representation Language) and VEML (Video Event Markup Language), used to describe processes and to markup video sequences, respectively. Other authors have discussed and improved this approach to support the representation of uncertain knowledge (Westermann & Jain, 2007). The research work in (Kokar & Wang, 2002) presents a symbolic representation for the data managed by a tracking algorithm. Similarly, recently it has

been a first approximation to the development of an ontology for tracking data (Snidaro, Belluz, & Foresti, 2007).

Scene interpretation issues (i.e., JDL L2 and L3) are being dealt with ontologies as well. In (Neumann & Möller, 2008), a proposal for scene interpretation based on Description Logics and supported by the reasoning features of RACER inference engine is presented. The problem of representing high-level semantics of situations with a computable formalism is also faced in (Kokar, Matheus, & Baclawski, 2009), where an ontology encoding Barwise's situation semantics is developed. The approach in (Aguilar-Ponce, Kumar, Tecpanecatl-Xihuitl, & Bayoumi, 2007) defines a multi-agent architecture for object and scene recognition in VSNs. In addition, the authors propose the use of an ontology to communicate information between task-oriented agents.

All these research works focus on contextual scene recognition, but as previously mentioned, it is also interesting to apply this knowledge to refine image-processing algorithms (which corresponds to JDL L4), and particularly trackers. A preliminary approach to this topic has been presented in (Gómez-Romero, Patricio, García, & Molina, 2009b). In this paper, the authors overview some aspects of the current state of this research and discuss its contributions from an architectural and knowledge management point of view. In the next section, we elaborate on this approach to analyze general problems and solution to CI exploitation in HLIF of VSNs.

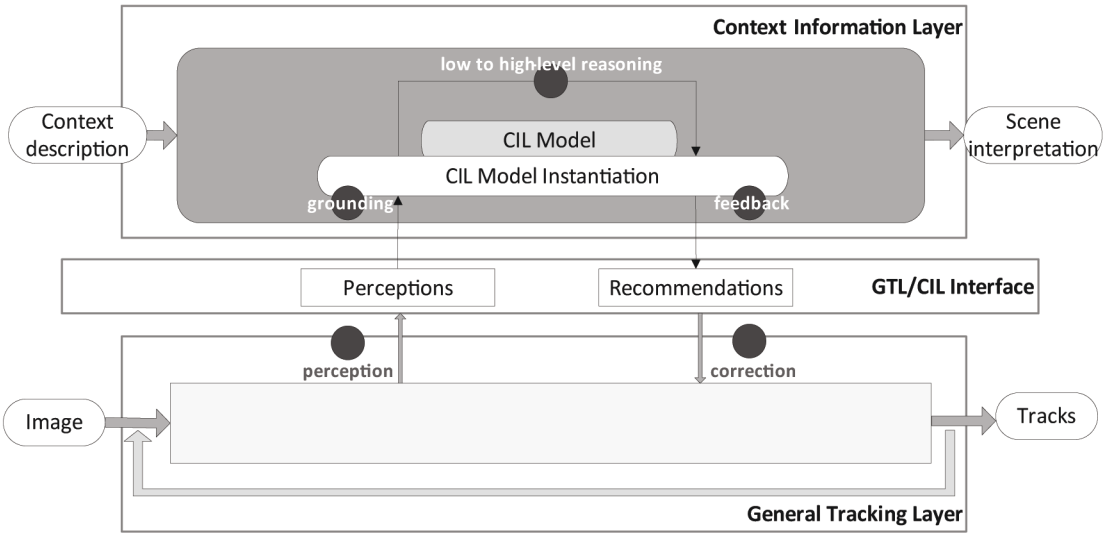
Architecture for High-Level Context-Based Visual Information Fusion

According to the description of the flow processes for CI exploitation discussed in section 5, VSNs will favor *a posteriori* schemas, since they facilitate the incorporation of CI at runtime in relation to the current situation estimate and, which is more important, CI cannot be easily integrated into the situation estimation procedures. This is especially relevant in the case of tracking, in which our proposal focuses, because it is not clear how CI can be directly applied within the classical tracking algorithms and, when possible, it entails the development of ad hoc solutions that can be hardly extended to similar domains, which rockets application costs. The matter is that there is a considerable gap between the CI that can be exploited to improve tracking and the form in which information can be provided to change the behavior of a tracking algorithm; the former is usually defined at a high abstraction level (e.g. "a person is hidden behind a column"), whereas the latter is usually a numerical parameter of a statistical procedure (e.g. noise models or variance matrices). An "A Priori" model of this correspondence would entail to *hardwire* the CI management logic to the particular software. Consequently, we adapt the notional process flow for "A Posteriori" CI exploitation depicted in figure 4 to decouple low (classical) and high-level (context-based) DIF procedures.

The general architecture of a CI-enhanced Fusion Node in a VSN is depicted in Figure 5. The schema shows the tracking system (the GTL, General Tracking Layer) and, built upon it, the context-based extension for HLIF (the CIL, Context Information Layer). Interestingly enough, the GTL and the CIL may not be in the same node of the VSN, in such a way that the HLIF may be performed not from the data obtained by a single node, but from the fused data obtained after combining the track results of independent nodes. The communication between the GTL and the CIL is performed through the generic GTL/CIL interface, which defines method to update the CIL model according to the GTL tracking data and to consult recommendations calculated by the CIL. This interface acts as a middleware that guarantees the interoperability and the independence of both layers.

The GTL is an implementation of a classical tracking procedure –we will assume the tracker described by Patricio et al. (2008), but the HLIF

Figure 5. Architecture of a HLIF node of a VSN

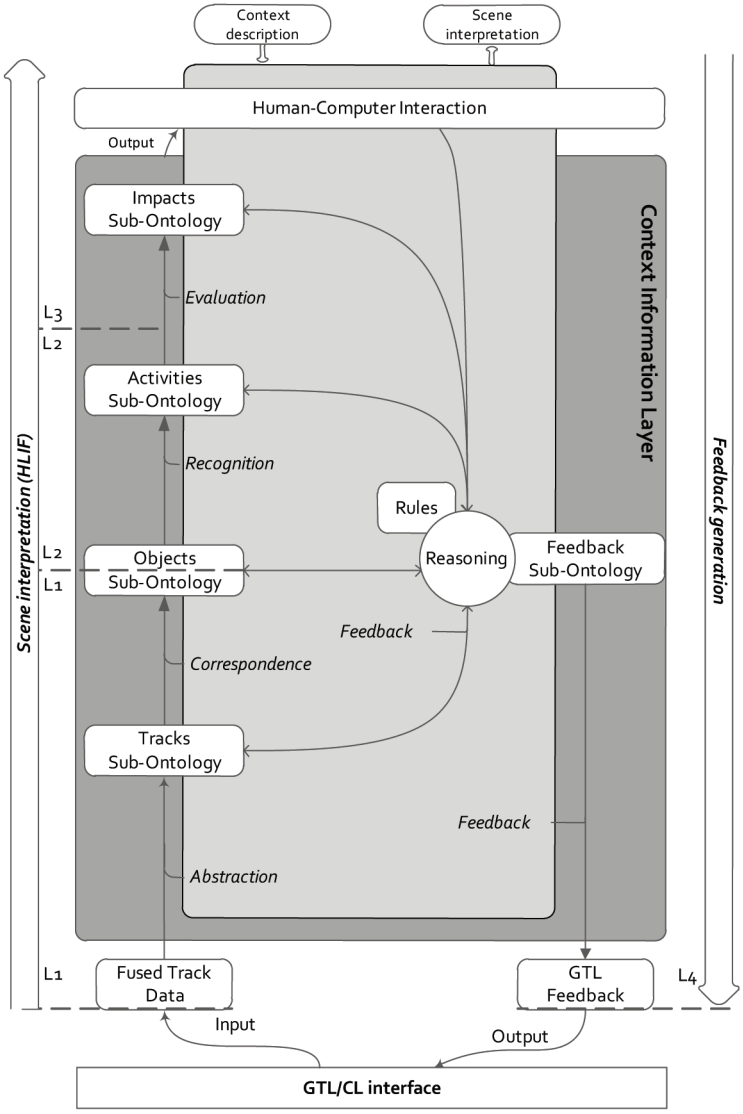


architecture is independent of the concrete algorithms. It sequentially executes various image-processing algorithms with a video sequence in order to track all the targets within the local field. The GTL is arranged in a pipelined structure of several modules, which correspond to the successive stages of the tracking process: (1) detection of moving objects; (2) blob-to-track multi-assignment; (3) track initialization/deletion; (4) trajectory analysis.

Essentially, the CIL manages the CIL model, which is the abstract representation of the scene. The CIL receives from the GTL tracking data, reasons with it, and provides as a result a high-level description of the situation. The CIL model is implemented as a set of interrelated ontologies organized according to the JDL model. The contents and the use of these ontologies are described in section 6.3. The information flow in the CIL is as follows (see Figure 6):

1. The GTL calls the GTL/CIL interface methods when tracks are created, deleted, or updated.
2. Updated track information is conveniently transformed into a symbolic representation, and the CIL model is updated with it. This triggers a reasoning procedure developed in four steps, which leads to a high-level representation of the scene from the low-level tracking data. In this process, CI obtained from several sources is applied to create and validate interpretation hypothesis. The reasoning procedure clearly reflects the task levels described in the JDL model, from L0 to L3:
   a. Abstraction of acquired tracking data, i.e., transformation of tracking data into ontology instances (*grounding*).
   b. Correspondence between tracks and objects, i.e., object identification from track properties.
   c. Recognition of activities from object descriptions, i.e., scene interpretation from object properties.
   d. Evaluation of threats, i.e., activity impact assessment.
3. Recommendations are created by reasoning with the updated scene model and CI –which can be seen as a L4 task. These recommendations are provided as a feedback to the GTL/CIL interface.

Figure 6. Process flow and knowledge structure of the CIL of a HLIF node of the VSN



4. The GTL calls the GTL/CIL interface methods to consult recommendations from the CIL, and converts them to concrete actions to correct its behavior.

Ontological Representation of CI for Visual DIF

The knowledge of the CIL model is structured in four layers, as depicted in figure 6, from tracking data to impacts and threats. The knowledge is closely related to two modules of the architecture: camera data acquisition and feedback generation.

- **Camera data (L0).** Sequence provided by the cameras (in some processable video format).
- **Tracking data (L1).** Output of the tracking algorithm represented symbolically. Example: frames, tracks and track properties (color, position, velocity, etc.).

- **Scene objects (L1-L1/2).** Objects resulting from making a correspondence between existing tracks and possible scene objects. For example, a track can be inferred to correspond to a person (possibly by applying CI). Scene objects include static elements, which may be defined a priori, and dynamic objects, which may be defined a posteriori. Example: person, door, column, window, etc.
- **Activities (L2).** Description of relations between objects which last in time. Example: grouping, approaching, picking/leaving an object, etc.
- **Impacts and threats (L3).** Cost or threat value assigned to activities.
- **Feedback and process improvement (L4).** Abstract representation of the suggestions provided to the tracking procedure.

The levels L1-L4 corresponds to an ontology in the CIL knowledge representation model. We provide a reference version of these ontologies[4] in the standard ontology language OWL (McGuiness & van Harmelen, 2004) that must be refined in a concrete application. For instance, in a particular application it may be interesting to define a concept to represent a Mirror as a specialization of StaticObject and ReflectingObject, which are concepts of the scene objects ontology. For a more extensive description of these ontologies, see (Gómez-Romero, Patricio, García, & Molina, 2009c).

In their definition, the ontologies of the CIL model only contain descriptive knowledge; i.e., axioms defining concepts and relations. They offer a vocabulary to describe scenes that must be instantiated in a concrete application according to the entities appearing in the scene, both static and dynamic. Thus, before starting the analysis of a video sequence, it is necessary to annotate the static elements of the scenario. Annotating the scenario means to create instances of the ontologies describing scenario properties: object position and size, possible occlusions, enter and exit zones, or any other contextual knowledge that we may consider necessary (see Figure 7).

Additionally, contextual rules must be loaded into the reasoner. Ontology instances corresponding to the dynamic entities of the scene are created and deleted during the processing of the sequence as a result of the track information provided by the GTL and additional reasoning processes (see Figure 8). In the next section, we

Figure 7. Initial mark-up of the scenario[5]

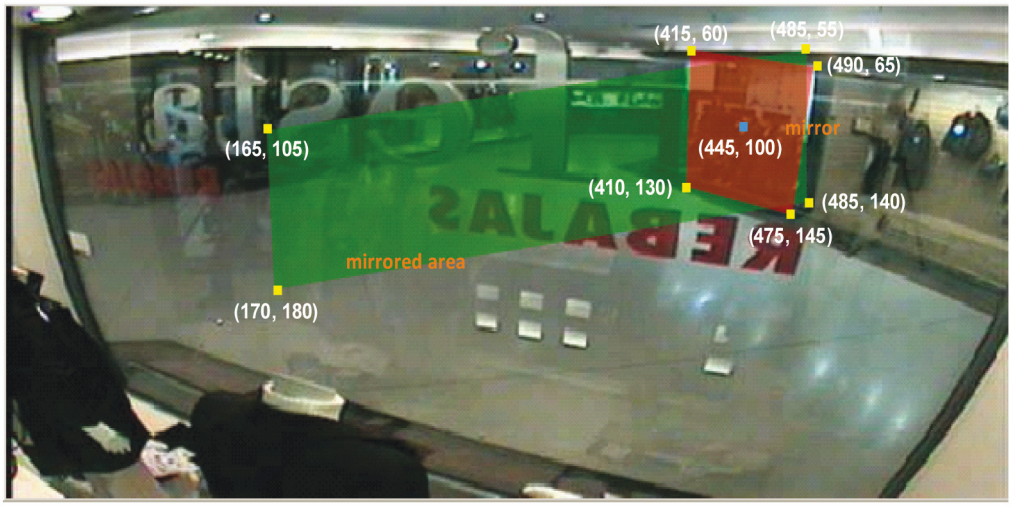
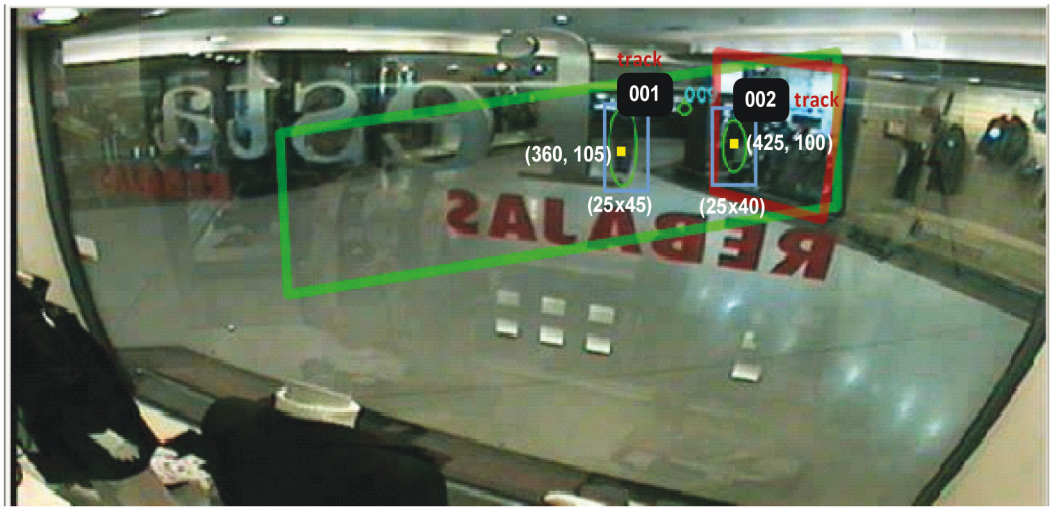


Figure 8. Additional instances created during video processing



explain the role of reasoning rules in the architecture and how instances are inferred to obtain the high-level description of the scenario.

Abductive Reasoning for HLIF (Interpretation and Feedback)

Standard ontology reasoning procedures can be performed within the CIL ontologies to infer additional knowledge from the explicitly asserted facts (tracking data and CI). By using a DL inference engine, tasks such as classification or instance checking can be performed. Nevertheless, monotonicity of ontology languages forbids adding new knowledge to the models while reasoning, which is required in scene interpretation. Actually, scene interpretation is a paradigmatic case of abductive reasoning, in contrast to the DL deductive reasoning: abductive reasoning takes a set of facts as input and finds a suitable hypothesis that explains them (sometimes with an associated degree of confidence or probability). This is what is needed in this case: we want to figure out what is happening in the scene from the observed and the contextual facts. In terms of the architecture of the CIL, scene interpretation can be seen as an abductive transformation from instances of a lower level ontology to instances of a higher level ontology. Abductive reasoning is not directly supported by ontologies (Elsenbroich, Kutz, & Sattler, 2006), but it can be simulated by using customized procedures or, preferably, by defining transformation rules in a suitable query language. The RACER inference engine allows abductive reasoning (Häarslev & Möller, 2001), and therefore it is a good choice to implement the reasoning procedures within the ontologies.

In the architecture, abductive rules formally represent contextual, heuristic and common sense knowledge to accomplish HLIF and low-level tracking refinement. Accordingly, we have two types of rules: bottom-up rules and top-down rules. *Bottom-up* rules are used in scene interpretation and obtain instances of an upper-level ontology from instances of a lower-level ontology. For instance, some rules could be defined to identify objects from track measures; i.e., to obtain instances of the scene objects ontology from instances of the tracking data ontology. An example rule may be: "create a person instance when an unidentified track larger than a predefined size is detected inside a region of the image". *Top-down* rules create suggested action instances from the current interpretation of the

scene, the historical data, and the predictions. These actions are used to adapt hypothesis at a lower-level to interpretations of a higher-level, which means the creation of instances of a lower-level ontology from instances of an upper-level ontology. Eventually, the top-down rules may create instances of the feedback ontology, which are retrieved and interpreted by the GTL through the GTL/CIL interface, resulting (if not discarded) in corrections to the low-level fusion procedure: tracking parameters, data structures, etc. Please notice that recommendations can be generated at different abstraction levels, either to modify instances of a lower-level ontology or to create new instances of the recommendations ontology. An example rule may recommend "to ignore a track associated to a person which is inside an area marked as a mirror". More details of the use of RACER abductive rules in the architecture can be found in (Gómez-Romero, García, Patricio, & Molina, 2009a).

FUTURE RESEARCH DIRECTIONS IN SENSOR-NETWORK-BASED CONTEXTUAL EXPLOITATION

In spite of the recent advances in contextual IF, there are several topics that require further research. Broadly speaking, the existing approaches are quite specific and it is difficult to apply the advances obtained in one application domain to another. The architecture presented in this chapter is an initial attempt to establish a common terminology and a reference organization for such systems, but it is necessary to advance in the development of frameworks that allow for the implementation of DIF applications with less effort. An IF framework should provide an adaptable infrastructure where specific procedures can be easily reused and/or integrated, especially those based on AI techniques, which are likely to play a key role in the next generation fusion applications. We envision a possible approach to a general IF multi-layer framework with: a front-end that manages hard and soft sensor inputs; an initial layer for detection, semantic labelling, and flow control, based on an intelligent repository of pluggable algorithms; a fusion layer, composed of several interrelated fusion nodes that process information at different JDL levels and incorporate CI to the process; and a presentation layer to convey the results through appropriate visualization interfaces. More discussions and realizations of general frameworks for DIF will be certainly useful to foster the creation of competitive solutions while cutting development costs in critical application areas.

There is also a promising field of study concerned with the design and the development of these architectures targeted to highly-distributed IF problems. The architecture presented in this chapter does not compel any task distribution schema, since we have assumed that IF procedures are independent, and that they can be performed by independent processes as long as proper communication mechanisms are offered. For the sake of simplicity, we have limited data alignment at tracking level, but it could be possible to combine estimations performed by single IF nodes at different levels, in such a way that the system would be able to obtain a combined view of the scenario from the detected objects or the recognized situations, instead of only the track data.

Recently, the multi-agent paradigm has been proposed to design and implement decentralized DDF schemas. The multi-agent paradigm provides a theoretical and practical framework to achieve successful communication and cooperation among the distributed components of the system. Multi-agent systems are argued to be well suited for the development of distributed systems in dynamic environments, as it occurs with VSNs. As a matter of fact, the notion of agent suits very well to the concept of intelligent camera. On the one hand, nodes in the VSNs are autonomous, in the sense that they have processing capabilities to acquire and process information in its field of view. On the

other hand, the social abilities of agents provide the necessary means to implement ISS protocols and to cooperate in the overall objective of the VSN. In addition, agent-based standard communication protocols are the support to achieve interoperation with other systems. Last but not least, there are several multi-agent frameworks that hide particular communication details and provide an easy way for developing distributed systems.

In this chapter, we have proposed ontologies to represent and reason with contextual and factual knowledge for HLIF. Nevertheless, classical ontologies do not allow the representation of uncertain or vague knowledge, despite it is inherent to several fusion applications, especially when it comes to CI. We consider that it will be convenient to adapt current approaches to fuzzy and probabilistic ontologies to IF, thus facilitating the creation of knowledge bases able to deal with imprecise information. This is particularly relevant in IF systems that have textual –or, in general, subjective human-generated input data–, which are more and more frequent in cutting-edge fusion domains (e.g. modern asymmetric warfare). Modelling the vagueness and the uncertainness of this knowledge will allow its integration with traditional sensor sources and obtaining more accurate estimates and assessments.

CONCLUSION

Information Fusion aims at improving the state estimations performed by a single-node system by the combination of the observations acquired at distributed sensor nodes. Information Fusion procedures are usually classified according to the JDL model, the prevailing theory to describe fusion systems, which establishes a common vocabulary to facilitate communication and understanding among the heterogeneous specialists interested in Information Fusion, and defines a functional model to structure fusion applications. The JDL model defines five levels of processing, ranging from the less abstract signal processing procedures to the more abstract situation recognition procedures. High-level information fusion is a term that designates the achievement of a symbolic interpretation of the perceived scenario in terms of situation descriptions and assessments.

Despite the success of low-level data fusion techniques, there is still a lack of systematic approaches to high-level information fusion, mainly because there are several information sources that need to be modeled, managed, and exploited in this kind of applications. Modern applications involving high-level information fusion require the employment and exploitation of all available information due to the complexities of these problems and the frequent difficulties of achieving sufficiently-rich observational data. Contextual information can often help in the formation of fused state estimates that are improvements over those developed from observational data alone, but of course such exploitation comes at a price of complexity that has to be factored into the cost-effectiveness of any prototype design. In this chapter, we have reviewed the basic design frameworks for such contextual exploitation strategies. We have described an "a priori" and an "a posteriori" schema to the exploitation of context information; the former is targeted at applications domains where the context can be easily introduced and applied in the fusion process, whereas the latter is more appropriate when the context is not known at design time, or it cannot be directly applied to guide the fusion.

Visual sensor networks are a prototypical case of the second schema, since video cameras provide a huge amount of information that is difficult to model, to interpret, and to exploit, especially at a high abstraction level. We have studied the requirements and the issues that arise in contextual high-level fusion, and provided a firsthand architectural proposal to discuss on the specific tasks and workflow that should be considered when designing and implementing a high-level information fusion system in a visual sensor

network. We have proposed the use of ontologies to represent and reason with visual knowledge in visual sensor networks, since they provide several advantages over ad hoc mechanisms: formalization, modularization, reutilization, availability of tools, etc. Information fusion is therefore seen as a model construction task accomplished by creating hypothesis through abduction on the perceived information based on the premises established by context knowledge.

However, as we have highlighted in the future research section, much remains to be done in defining and designing more robust and efficient approaches for use of contextual information, especially when visual sensor networks are involved. The fusion community needs to caucus on this issue and develop holistic research strategies that will result in unified design patterns for contextual exploitation.

ACKNOWLEDGMENT

The UC3M Team gratefully acknowledges that this research activity is supported in part by Projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, CAM CONTEXTS (S2009/TIC-1485) and DPS2008-07029-C02-02.

UC3M also thanks Prof. James Llinas for his helpful comments during his stay, which has been supported by the collaboration agreement 'Chairs of Excellence' between University Carlos III and Banco Santander.

Prof. James Llinas gratefully acknowledges that this research activity is supported by a Multidisciplinary University Research Initiative (MURI) grant (Number W911NF-09-1-0392) for "Unified Research on Network-based Hard/Soft Information Fusion", issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery.

REFERENCES

Aguilar-Ponce, R., Kumar, A., Tecpanecatl-Xihuitl, J. L., & Bayoumi, M. (2007). A network of sensor-based framework for automated visual surveillance. *Journal of Network and Computer Applications*, *30*(3), 1244–1271. doi:10.1016/j.jnca.2006.04.011

Arambel, P. (2008). Structured pedigree information for distributed fusion systems. *Proceedings of the SPIE* (17).

Arndt, R., Troncy, R., Staab, S., Hardman, L., & Vacura, M. (2007). COMM: Designing well-founded multimedia ontology for the Web. *Proceedings of the 6th International Semantic Web Conference (ISWC '07)* (pp. 30-43), Busan, South Korea.

Baader, F., Calvanese, D., McGuiness, D. L., Nardi, D., & Patel-Schneider, P. F. (2003). *Description logic handbook: Theory, implementation, and applications*. Cambridge, UK: Cambridge University Press.

Barron, J., Fleet, D., & Beauchemin, S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, *12*(1), 42–77. doi:10.1007/BF01420984

Bishop, C. (2004). *Pattern recognition and machine learning*. New York, NY: Springer.

Bremond, F., & Thonnat, M. (1996). A context representation for surveillance systems. In *Proceedings of the Workshop on Conceptual Descriptions from Images at the 4th European Conference on Computer Vision (ECCV '96)*. Cambridge, UK.

Brower, J. (2001). Relations without polyadic properties: Albert the Great on the nature and ontological status of relations. *Archiv für Geschichte der Philosophie*, *83*, 225–257. doi:10.1515/agph.2001.001

Castanedo, F., García, J., Patricio, M. A., & Molina, J. M. (2009). Designing a visual sensor network using a multi-agent architecture. In *Proceedings of the 7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS '09)* (pp. 430-439), Salamanca, Spain. Berlin/Heidelberg, Germany: Springer.

Castanedo, F., García, J., Patricio, M. A., & Molina, J. M. (2010). Data fusion to improve trajectory tracking in a Cooperative Surveillance Multi-Agent Architecture. *Information Fusion*, *11*(3), 243–255. doi:10.1016/j.inffus.2009.09.002

Chang, K., Chong, C. Y., & Bar-Shalom, Y. (1986). Joint probabilistic data association in distributed sensor networks. *IEEE Transactions on Automatic Control*, *31*(10), 889–897. doi:10.1109/TAC.1986.1104143

Chong, C. Y., Mori, S., & Chang, K. C. (1990). Distributed multitarget multisensor tracking. In Bar-Shalom, Y. (Ed.), *Multitarget-Multisensor Tracking: Advanced Applications* (*Vol. 1*, pp. 247–295). Norwood, MA: Artech House.

Das, S. (2008). *High-Level Data Fusion. London*. Boston: Artech House Publishers.

Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing Journal*, *5*(1), 4–7. doi:10.1007/s007790170019

Ekbia, H. R., & Maguitman, A. G. (2001). Context and relevance: A pragmatic approach. In *Proceedings of the 3rd International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT '01)* (pp. 156-169). Dundee, UK: Springer.

Elsenbroich, C., Kutz, O., & Sattler, U. (2006). A case for abductive reasoning over ontologies. In *Proceedings of the OWL Workshop: Experiences and Directions (OWLED '06),* Athens, GA, USA.

François, A., Nevatia, R., Hobbs, J., Bolles, R., & Smith, J. (2005). VERL: An ontology framework for representing and annotating video events. *IEEE MultiMedia*, *12*(4), 76–86. doi:10.1109/MMUL.2005.87

Gómez-Romero, J., García, J., Patricio, M. A., & Molina, J. M. (2009). Towards the implementation of an ontology-based reasoning system for visual information fusion. In *Proceedings of the 3rd Skövde Workshop on Information Fusion Topics (SWIFT '09)* (pp. 5-10), Skövde, Sweden.

Gómez-Romero, J., Patricio, M. A., García, J., & Molina, J. M. (2009). Context-based reasoning using ontologies to adapt visual tracking in surveillance. In *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '09)* (pp. 226-231), Genoa, Italy.

Gómez-Romero, J., Patricio, M. A., García, J., & Molina, J. M. (2009). Ontological representation of context knowledge for visual data fusion. In *Proceedings of the 12th International Conference on Information Fusion (FUSION '09)* (pp. 2136-2143), Seattle, WA, USA.

Gruber, T. (1993). A translation approach to portable ontology. *Knowledge Acquisition*, *5*(2), 199–220. doi:10.1006/knac.1993.1008

Häarslev, V., & Möller, R. (2001). Description of the RACER system and its applications. In *Proceedings of the International Workshop on Description Logics (DL '01)*. Stanford University, CA, USA.

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, *85*(1). doi:10.1109/5.554205

Heikkila, J., & Silven, O. (1997). *A four-step camera calibration procedure with implicit image correction* (pp. 1106–1112). IEEE Computer Vision and Pattern Recognition.

Henricksen, K. (2003). *A framework for context-aware pervasive computing applications* (PhD Thesis, University of Queensland).

Hongeng, S., Nevatia, R., & Bremond, F. (2004). Video-based event recognition: Activity representation and probabilistic recognition methods. *Computer Vision and Image Understanding*, *96*(2), 129–162. doi:10.1016/j.cviu.2004.02.005

Hu, W., Xie, D., Tan, T., & Maybank, S. (2004). Learning activity patterns using fuzzy self-organizing neural network. *IEEE Transactions on Systems, Man, and Cybernetics*, *34*(3), 1618–1626. doi:10.1109/TSMCB.2004.826829

Julier, S. (2009). Estimating and exploiting the degree of independent information in Distributed Data Fusion. In *Proceedings of the 12th International Conference on Information Fusion (FUSION '09),* Seattle, WA, USA.

Julier, S., & Uhlmann, J. (2001). General decentralized data fusion with Covariance Intersection. In Liggins, M. E., Hall, D., & Llinas, J. (Eds.), *Handbook of multisensor data fusion* (2nd ed., pp. 319–344). Boca Raton, FL, USA: CRC Press. doi:10.1201/9781420038545.ch12

Kandefer, M., & Shapiro, S. C. (2008). A categorization of contextual constraints. *Biologically inspired cognitive architectures:* Papers from the AAAI Fall Symposium (pp. 88-93). Menlo Park, CA, USA: AAAI Press.

Kokar, M., Matheus, C., & Baclawski, K. (2009). Ontology-based situation awareness. *Information Fusion*, *1*, 83–98. doi:10.1016/j.inffus.2007.01.004

Kokar, M., & Wang, J. (2002). Using ontologies for recognition: An example. In *Proceedings of the 5th International Conference on Information Fusion* (2, pp. 1324-1330).

Lee, W., Bürger, T., & Sasaki, F. (2009). *Use cases and requirements for ontology and API for Media Object 1.0*. Retrieved from W3C Working Draft: http://www.w3.org/ TR/ media- annot-reqs/

Liggins, M., Hall, D., & Llinas, J. (2009). *Handbook of multisensor data fusion* (2nd ed.). Boca Raton, Florida, USA: CRC Press.

Lipton, A., Fujiyoshi, H., & Patil, R. (1998). Moving target classification and tracking from real-time video. In *Proceedings of the 4th IEEE Workshop Applications of Computer Vision (WACV '98)* (pp. 129-136). Princeton, New Jersey.

McGuiness, D. L., & van Harmelen, F. (2004). *OWL Web ontology language overview.* Retrieved from W3C Recommendation: http://www.w3.org/ TR/ owl-features/

Neumann, B., & Möller, R. (2008). On scene interpretation with Description Logics. *Image and Vision Computing*, *26*, 82–101. doi:10.1016/j.imavis.2007.08.013

Nowak, C. (2003). On ontologies for high-level information fusion. In *Proceedings of the 6th International Conference on Information Fusion (FUSION 2003)* (1, pp. 657-664). Cairns, Australia.

Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *Proceedings of the 46th Conference in Decision and Control* (pp. 5492-5498). New Orleans, LA, USA.

Patricio, M. A., Castanedo, F., Berlanga, A., Pérez, Ó., García, J., & Molina, J. M. (2008). Computational intelligence in visual sensor networks: Improving video processing systems. In *Computational Intelligence in Multimedia Processing: Recent Advances* (pp. 351–377). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-76827-2_14

Piccardi, M. (2004). Background subtraction techniques: a review. In *Proceedings of the IEEE International Conference on Systems, Man & Cybernetics* (4, pp. 3099-3104). The Hague, Netherlands.

Regazzoni, C. S., Ramesh, V., & Foresti, G. L. (2001). Scanning the issue/technology. Special Issue on Video Communications, Processing, and Understanding for 3rd Generation Surveillance Systems. *Proceedings of the IEEE*, *89*(10), 1355–1367. doi:10.1109/5.959335

Shi, Y., Wan, Y., & Ge, Q. (2009). *A unified out-of-sequence measurements fusion algorithm.* 1st International Workshop on Database Technology and Applications, Hubel, China.

Snidaro, L., Belluz, M., & Foresti, G. L. (2007). Domain knowledge for surveillance applications. In *Proceedings of the 10th International Conference on Information Fusion (FUSION '07)* (pp. 1-6), Quebec, Canada.

Staab, S., & Studer, R. (2009). *Handbook on ontologies*. Berlin/Heidelberg, Germany: Springer-Verlag. doi:10.1007/978-3-540-92673-3

Tsai, R. (1987). A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, *3*(4), 323–344. doi:10.1109/JRA.1987.1087109

Valera, M., & Velastin, S. A. (2005). Intelligent distributed surveillance systems: A review. *IEEE Proceedings - Vision. Image, and Signal Processing*, *152*(2), 192–204.

Wang, L., Hu, W., & Tan, T. (2003). Recent developments in human motion analysis. *Pattern Recognition*, *36*(3), 585–601. doi:10.1016/S0031-3203(02)00100-0

Westermann, U., & Jain, R. (2007). Toward a common event model for multimedia applications. *IEEE MultiMedia*, *14*(1), 19–29. doi:10.1109/MMUL.2007.23

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 161–195. doi:10.1109/34.888718

Zimmermann, A., Lorenz, A., & Oppermann, R. (2007). An operational definition of context. In *Proceedings of the 6th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT '07)* (pp. 558-571), Roskilde, Denmark. Berlin/Heidelberg, Germany: Springer-Verlag.

ENDNOTES

1. The literature also uses "Sensor Fusion" and "Data Fusion" but we do not enter into a discussion of the subtle and non-standardized differences between these notions and the term "Information Fusion". The fundamental functions of each of these processes are largely the same.
2. "JDL" is an abbreviation for the Joint Directors of Laboratories, a US Defense Department organization that existed in the 1980´s within which there was a Data Fusion Group that standardized many of the still-used taxonomic and functional characterizations of IF.
3. Note that in some applications, the observational capability can be of other types, to include human observers.
4. http://www.giaa.inf.uc3m.es/miembros/jgomez/ontologies/
5. The example frames have been extracted from the PETS2002 dataset. The underlying task in this dataset was to track pedestrians in indoor video sequences of a shopping mall (http://www.cvg.cs.rdg.ac.uk/PETS2002/pets2002-db.html).