This document is published in:

Shahbazian, E.; Rogova, G. (eds.), 2010. *Human Systems Integration to Enhance Maritime Domain Awareness for Port/Harbour Security*, [Proceedings of the NATO Advanced Research Workshop, Opatija, Croatia, 8-12 December 2008], (NATO Science for Peace and Security Series, Sub-Series D: Information and Communication SecurityInformation and Communication Security – v. 28), IOS Press, pp. 79-92.

DOI: 10.3233/978-1-60750-621-8-79

Contextual and Human Factors in Information Fusion

Jesus GARCIA[a,1] Miguel A. PATRICIO[a], José M. MOLINA[a] and Luiz A. CARDOSO[b]
[a] *GIAA – University Carlos III de Madrid, Spain*
[b] *IPqM – Navy Research Institute, Brazil*

Abstract. Context and human factors may be essential to improve measurement processes for each sensor and the particular context in each sensor could be used to obtain a global definition of context in multisensor environments. The reality may be captured by human sensorial domain based only on machine stimulus and then generate a feedback which can be used by the machine, at its different processing levels, adapting its algorithms and methods accordingly. Reciprocally, human perception of the environment could also be modelled by context in the machine. In the proposed model, both machine and man take sensorial information from the environment and process it cooperatively until a decision or semantic synthesis is produced. In this work, we present a model for context representation and reasoning to be exploited by fusion systems. In the first place, the structure and representation of contextual information must be determined before being exploited by a specific application. Under complex circumstances, the use of context information and human interaction can help to improve a tracking system's performance (for instance, video-based tracking systems may fail when dealing with objects interaction, occlusions, crosses, etc.).

Keywords. Context representation, adaptive data fusion, human computer interaction.

1. Introduction

Surveillance network processing problems are relevant to Air Traffic Control (ATC), Vessel Traffic Services (VST) and many security and defence applications. These systems are usually meant to emulate human intelligent behaviour in surveillance tasks. In fact, before machine-implemented information fusion, maritime surveillance, for instance, was a fully non-automated activity, performed by human operators by ("mentally") fusing data from different sources, including that collected from direct sea observation, radio communication messages and, when available, air/surface surveillance radars. Intelligence and other contextual information always played an essential role in this task, conditioning human perceptual analysis and decision.

Along with the evolution of computational tools and software technology, the use of different types of radar and other sensors became widespread, many modern systems today being capable of autonomous fusion of target data from an assorted range of cameras or different radar sensors, either land based, shipborne or airborne. AIS (Automatic Identification System) data also become a valuable source of information to

[1] Corresponding Author.

be fused in surveillance applications, after its introduction by the International Convention for the Safety of Life and Sea (SOLAS), as a compulsory self-identification standard for a large class of commercial vessels [1].

In the near future, other higher lever data, like port activity, shipboard communication, social connections of the crew and passengers, detailed cargo description, news around the word (as posted in the Internet, for instance), are expected to be considered, increasing enormously the range of possibilities for data association and information fusion. The improvement of the fusion and analysis of maritime intelligence information is an eventual goal in maritime surveillance systems being depicted for the future [2].

Due to the recurrent asymmetrical threats observed in the world in the last years, the fusion models need to take into account increasingly more civilian sourced information, becoming prone to achieve very high levels of complexity. Because of that, and because of the above mentioned human-centred subjective criteria for user satisfaction, the demand for system performance in this framework draws attention to information fusion models with the potential to fully integrate and exploit both human and machine capabilities (Some may say that this model encompass some cheating, resembling that first chess machine with an dwarf player inside [3]. In our point of view, this is more like advanced chess, where the "team" is assembled by a man cooperating with a machine program [4]).

While contextual information can satisfactorily include high level human-modelled offline information to enhance a system's behaviour, there has been also some interest in the quest for models of online man-machine cooperation in information fusion and decision making. In [5] E. Blash has introduced a fuse-interaction model where the user can interfere at any hierarchical level of the JDL model [6]. In a more recent proposal [7] this possibility is extended to simultaneous cooperative learning, much in a symbiotic human-machine processing, in order to enhance information fusion. In order to address maritime surveillance and port security applications using sensor networks, we now consider a modelling approach where both full symbiotic Human-Computer reciprocal learning interaction and context representation are present.

Human perception of context and activity differs from the perceptions coming from computer sensors. To be capable of anticipating human behaviour, a context model then needs to bridge the gap between human understanding and computer perception. The structure and representation of this information must be determined before being exploited by a specific application [8]. Once a context model is built up and validated, it can be used to identify and correct erroneous or incomplete data coming from perceptual components, which can be predicted and corrected knowing the situation, i.e. the current state within the context model [9]. Context could be useful to improve measurement process for each sensor and the particular context in each sensor could be fused to obtain a global definition of context in multisensor environments.

In this proposal, extensions to the well known JDL architecture are defined to enhance reasoning with external information, besides sensorial input, handling both contextual information and human interaction, as indicated in figure 1.

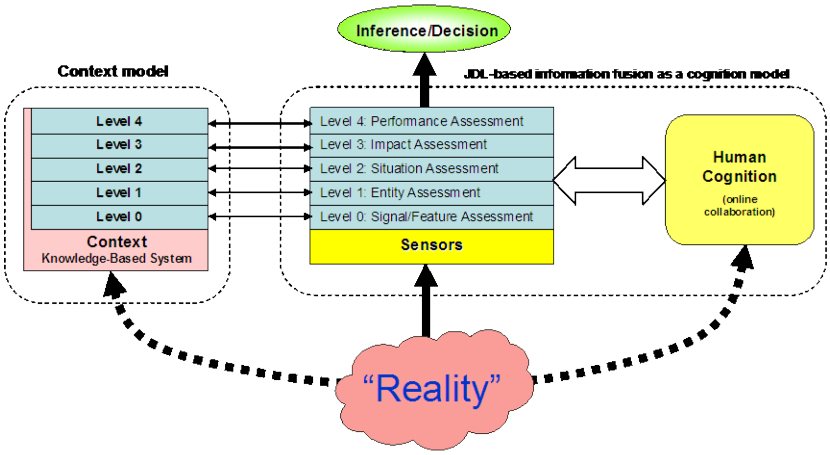


Fig. 1. Extended architecture for Fusion Enhancement.

A direct application of contextual information introduces improvements at the lowest level of fusion systems, in charge of detection and tracking all interesting objects. In this work, the improvement in surveillance capabilities of Visual Sensor Networks (VSN) using a Context Model (static and dynamic) is explored, leading to the proposal of a combined Context-HCI information fusion model. Video-based tracking systems may fail when dealing with complex scenarios, those in which objects enter and/or leave the scene, objects interact with each other producing occlusions, crosses, unions, separations, etc [10][11]. It is under this type of circumstances when the use of context information can help to improve a tracking system's performance [12][13]. In particular, the tasks of data association and track fusion can be improved considering the human input and a high-level reasoning process based on the available contextual information. In the next section we explain our JDL extension to take into account the user interaction to adapt the fusion chain. Section 3 presents the general architecture to integrate context in Level 1 fusion systems, with the details for context representation in the case of visual sensor networks is presented in section 4. Finally section 5 presents the conclusions and future works.

2. HCI in Maritime Surveillance Fusion Systems

The JDL Data Fusion Model can be thought of as a general framework for emulating human reasoning, concerning environment perception, thus providing an automated decision making tool. In the model, the complex and apparently parallel mental processes exhibited by human intelligence are hierarchically organized as a sequence of growing abstraction levels, each taking input from the previous, lower levels. The interaction of human actions into the JDL architecture allows accepting direct human guidance at any level, exhibiting supervised learning properties and treating the human actor as an external Level 4 agent.

In the proposed model (see diagram in figure 2), both machine and man take sensorial information from the environment and process it until a decision or semantic synthesis is produced. The machine works hierarchically, executing from low level signal processing and feature extraction algorithms, at the bottom layer, to high level programming strategies, at the top layer. On the man side, perception and reasoning are similarly developed, to include perception induced by stimulus generated at intermediate machine processing levels. In some cases, the reality may lay out of human sensorial domain, with man relying only on machine-translated stimulus. The machine, reciprocally, receives feedback at its different processing levels, adapting its algorithms and methods accordingly.

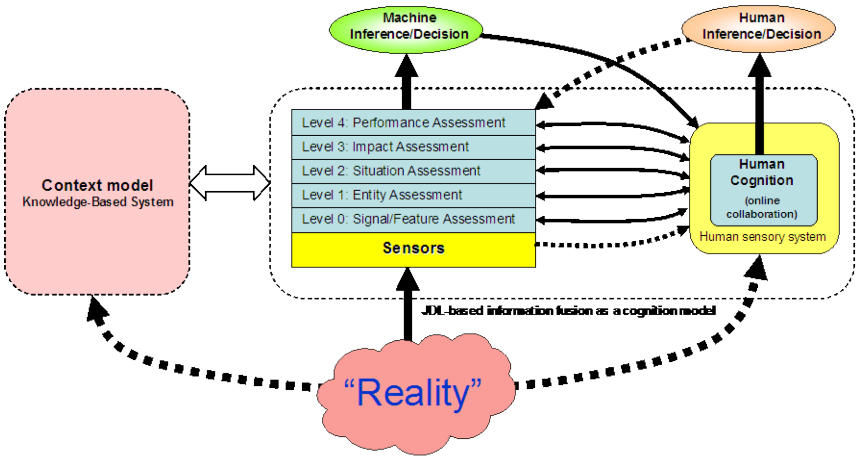


Fig. 2. Information fusion boosting through human-computer interaction

A fusion-based automated surveillance system, operating over a sensor network, can thus be thought of as implementing an Artificial Intelligence (AI) model of Human Cognition that produces:

- The fusion of sensor information, with the elimination of redundant input information through integration and generalization;
- The improvement of accuracy and reduction of uncertainty in sensor information [14]; and
- A compiled representation of the environment (within a delimited space-time volume), the entities present in it, comprehension of their meaning, their state, and predicted short future evolution, based on all above sensor-fused information, thus inducing what is called "situation awareness" [15].

The purpose of achieving "situation awareness" is usually to allow some decision making process to occur, either agent (automatic) or human based, aiming optimization of decision latency and correctness, according to mission previously assigned objectives. Thus, a fourth computational task for the AI processing could be described as to produce a decision or a semantic evaluation of the input space. Such systems will often carry, relatively to those based on human intelligence only, some expectancy for exceeding performance levels, as given by either:

(i) Shorter response time requirements; or
(ii) higher accuracy requirements; or
(iii) higher throughput (expressed in terms of larger scan volume, larger target quantity or faster scan rate); or

(iv) higher confidence levels; or

(v) improved repeatability requirements (consistent behaviour in a range of situations where human cognition might fail, for example, as a result of fatigue or stress); or

(vi) a simultaneous combination of these requirements.

Under this circumstances, it is remarkable that, in spite of the urge for automation, replacing some routine human action in the information gathering and processing by machine implemented algorithms, to outperform human capacity, the kernel of the decision process usually (and also by requirement) relies on a human being itself, namely, the system user or operator. The same goes for the evaluation criteria of system performance. This “human-has-the-right-to-final-judgement” requirement elicits that a certain hierarchical modelling is implied in the system design, suggesting human intelligence would be self-perceived as hierarchically oriented.

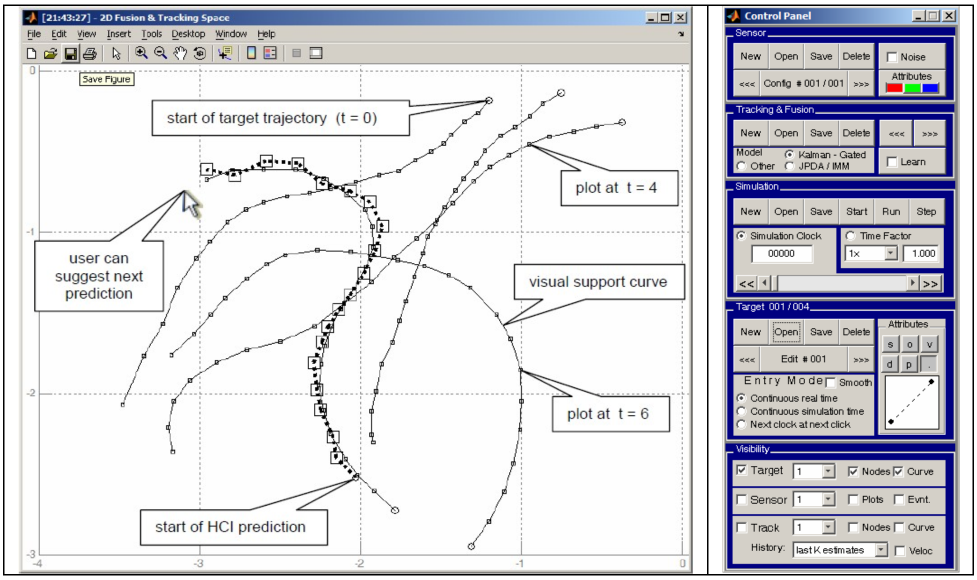


Fig. 3. Interface for Human feedback in low-level fusion process

The capacity of the fusion-based automated surveillance system to obtain the situation awareness as human operator relies in the definition of the context information and the capacity of the system to infer the real situation from this context information. This operation needs specifically developed interfaces to adapt low-level operations to human perception (see an example in figure 3).

3. Architecture for Context Integration in Information Fusion at Sensor Level

In our proposal, the context represents the previous knowledge of the human operator to interact/adapt the system to real situations. In low-level interaction, context represents the knowledge of human operator to adapt the tracking system at the Fusion Center integrating the contextual information and the tracking system's information of each sensor, treating it with low level concepts and context information to interpret what is happening in the scene and thus improving the system's output and also to improve each local tracking system. The use of context main objective of this work is to develop an architecture for surveillance networks in complex scenarios and proposes a methodology to develop context-based fusion system (an example in visual sensor networks is described in next section). The architecture proposed in this work is based on two-layer data processing modules to improve the association process of a tracking system (see figure 4):

- GTL (General tracking layer). Generic Multipurpose Tracking Process for Video Surveillance Systems
- CL (Context layer). Symbolic reasoning to manage the symbolic interface between GTL modules, asses situation and take the appropriate decisions

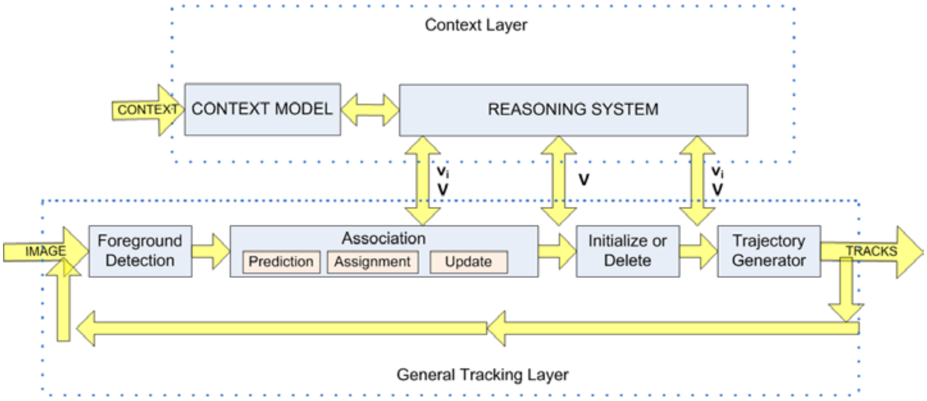


Fig. 4. Context-based Fusion Architecture to interact on Level 1.

The inclusion of context in a fusion system requires the definition and development of several concepts (see Figure 5):

- Describing a formal Context Model to represent information for the specific scenario that can help interpret what is happening in the environment. This Context Model considers both static and dynamic contextual information.
- Propose a reasoning schema able to improve a tracking system performance using the specific information of the Context Model.
- Define the adaptation of the tracking level from the context level through an interface.

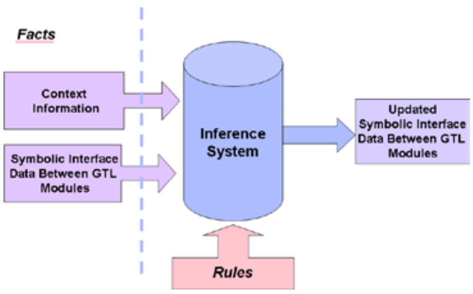


Fig. 5. Conceptual representation of Context Concepts in Level 1.

4. The role of Context in Visual Sensor Networks

The concept of Visual Sensor Network (VSN) is the current paradigm to deploy visual surveillance systems based on distributed fusion. For instance, Fig. 6 illustrates a domestic VSN with three indoor cameras with overlapped Field of Views. These overlapped areas could be exploited by the Fusion Centre in order to get more accurate results and guarantee coherent monitoring in the global area. The tracking algorithms implemented in the surveillance-sensor nodes have to deal with motion detection errors and complex object interactions (merging, occlusions, fragmentation, etc.). The Fusion Center combines the information inferred by the individual surveillance-sensor agents to maximize the final information content about the area to be guarded. We have previously proposed a fusion architecture for distributed solutions in visual sensor networks [16].

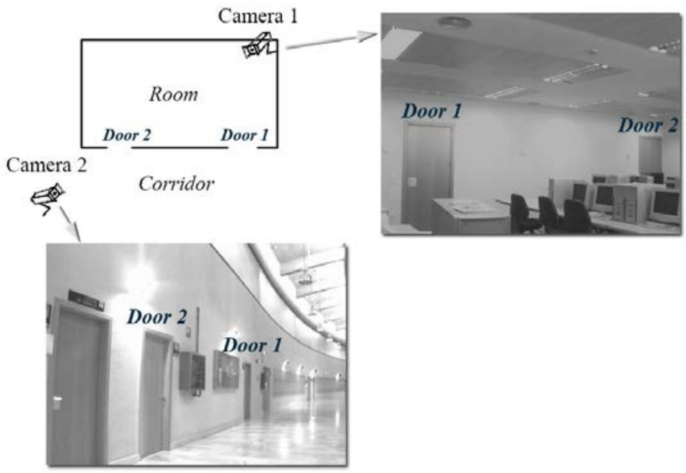


Fig. 6. Indoor visual sensor network example

4.1. Context Model

The Context Model (CM) is an implementation of specific context information that helps understand and improve the tracking system [17]. This information is obtained from the input Scenario's Context and from the Fusion Center. The information obtained from the input in each camera is static information about the scenario, while the information obtained and updated from the Fusion Center is dynamic information about the moving objects. CM maintains the following context information which helps reasoning about what is really happening in the scene, and thus, it helps improve the tracking system. Let $\mathfrak{K}$ be the set of objects which enclose the context information for a specific scenario. $\mathfrak{K}$ is described as a set of regions of interest $\mathfrak{R}$, static objects $\mathfrak{S}$ and tracks information $\mathfrak{I}$ (see Fig. 7). Regions of interest $\mathfrak{R}$ can be described as areas in a scene where specific characteristics can be found. Static objects $\mathfrak{S}$ are objects in the scene where tracks will be initialized or deleted. Finally, tracks information $\mathfrak{I}$, stands for consistent tracks properties, such as size, position, color, etc., of the objects that are in the scene.

$\mathcal{R} = \{r_1, \ldots, r_n\}$
$\mathcal{S} = \{s_1, \ldots, s_m\}$
$\mathcal{I} = \{i_1, \ldots, i_p\}$
$\mathcal{K} = \mathcal{R} \cup \mathcal{S} \cup \mathcal{I}$

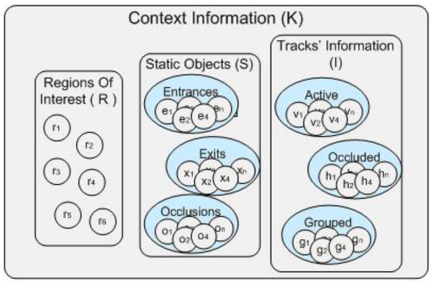


Fig. 7. Context Information

Regarding the set of regions of interest $\mathcal{R}$, a region of interest r_i can be described as:

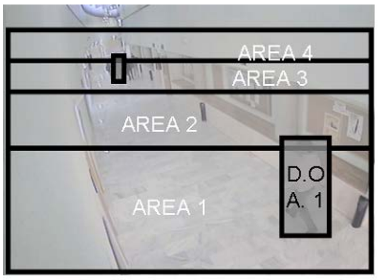
$$r_i = \{z_{r_i}, t_{r_i}, s_{r_i} d_{r_i}\}$$

where:

- z_{r_i} represents this region of interest zone as polygons, because polygons are simple and flexible enough to match human knowledge.
- t_{r_i} is the r_i's temporal information. This temporal information indicates at what time interval this region's characteristics are relevant. Therefore, there could be different regions of interest with the same zones, but different temporal information.
- s_{r_i} gives illumination geometries' information. For instance, in r_i at the time interval t_{r_i}, the shadows and reflections size and direction.
- d_{r_i} stands for the dynamic objects' features that are expected to be observed in a specific region of interest at a specific time interval. This information allows the CL to reason about what is happening in the scene.

There are three types of static objects $\mathcal{S}$ considered: entrances $\mathcal{E}$ and exits $\mathcal{X}$ and occlusion objects $\mathcal{O}$. All of these types of static objects are described by polygons. For simplicity, the areas described in the examples presented are represented by rectangles.

For instance, Fig. 8 shows an example indoor scenario, on the left the areas and some of the dynamic objects are depicted, and on the right the static objects. The scene is divided in four areas, depending on their dynamic object's sizes. Therefore, when an object moves from Area 1 to Area 4 its size decreases. In this scene there are no occlusion objects, but instead seven entrances and exits are established.



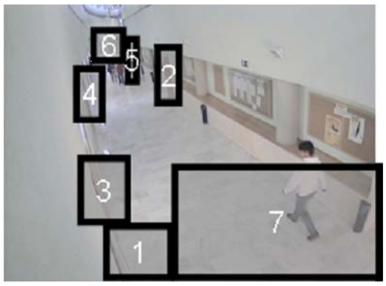


Fig. 8. The left image shows the areas and some of their dynamic objects in CORRIDOR, and the right image shows the static objects.

4.2. Sensor and Information Fusion with Context Reasoning in VSN

For each i-th local processor associated to one sensor, at the lowest level, CM maintains tracks information $l^{i}[k]$ for each frame k, every time a new track is initialized, updated or deleted. This tracking information is an extension of conventional systems, with specific information needed by the Reasoning System in the Fusion Center to represent the dynamic contextual information. Tracking information at i-th sensor, $l^{i}[k]$, has been classified in three types of tracks: active $v^{*i}[k]$, occluded $h^{i}[k]$ and grouped $g^{i}[k]$, defined for each sensor:

$$l^{i}[k] = \{v^{*i}[k], h^{i}[k], g^{i}[k]\}$$

Active tracks, $v^{*i}[k]$, are those containing the tracks sent by each i-th local processor.

$$v_j^{*i}[k] = \{v_1^{*i}[k], v_2^{*i}[k], \ldots, v_M^{*i}[k]\}$$

Each of these active tracks $v_j^{*i}[k]$ are described by a set of features such as a unique identifier, velocity, position, color ... and the list of grouped tracks, $l_j^{i}[k]$ (see Fig. 9),

$$v_j^{*i}[k] = \{v_j^{i}[k], l_j^{i}[k]\}$$

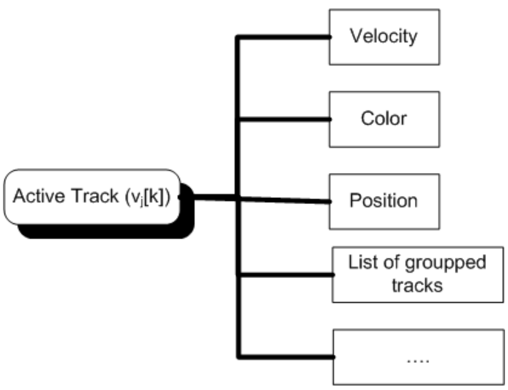


Fig. 9. Active Tracks

So, the Fusion Center maintains a set of occluded tracks $h^{i}[k]$, and a set of grouped tracks $g^{i}[k]$ for each sensor, updated to the last frame k.

$$h^i[k] = \{h_1^i[k], \ldots, h_q^i[k]\}$$

$$g^i[k] = \{g_1^i[k], \ldots, g_u^i[k]\}$$

An active track $v_j^{*i}[k]$, as described earlier, is represented by a set of features and $l_j^i[k]$, where $l_j^i[k]$ is a list that contains the identifiers of the tracks that are grouped with that active track $v_j^{*i}[k]$. For instance, $g_j^i[k]$ can be a grouped track at frame k, that was an active track $v_j^i[w]$ at frame w, but that at frame k it has been grouped with an active track $v_j^i[k]$.

In order to make proper use of the contextual information, a knowledge base with all the contextual information integrates the knowledge coming from each sensor, as depicted in Fig. 10.

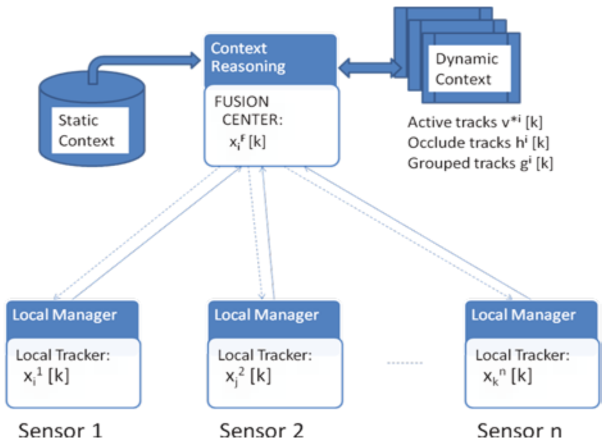


Fig. 10. Context-based fusion

This knowledge base is composed of predicates such as *is_a_track(), is_a_occlusion(), is_hidden(), are_grouped()*, etc. This system is responsible for the coherence of local/global context information. As a part of the fusion process, this reasoning system [18][19] takes into account the correspondences among data from different sensors (data association), applies an inference process and improves the track fusion process and the global result (coherence/fusion information) is send back to every local process to adapt the local behaviour accordingly to the global conditions.

For instance, considering the mono-sensor indoor scenario presented above, the contextual information is inserted as a set of facts in the knowledge base, and remains within the knowledge base. The static facts are first inserted at the knowledge base, these initial facts c_i represent the contextual information. Entrances, exits, areas and dynamic objects are represented by polygons (height and width), and all but the dynamic objects have a position, as illustrated in Figure 11.

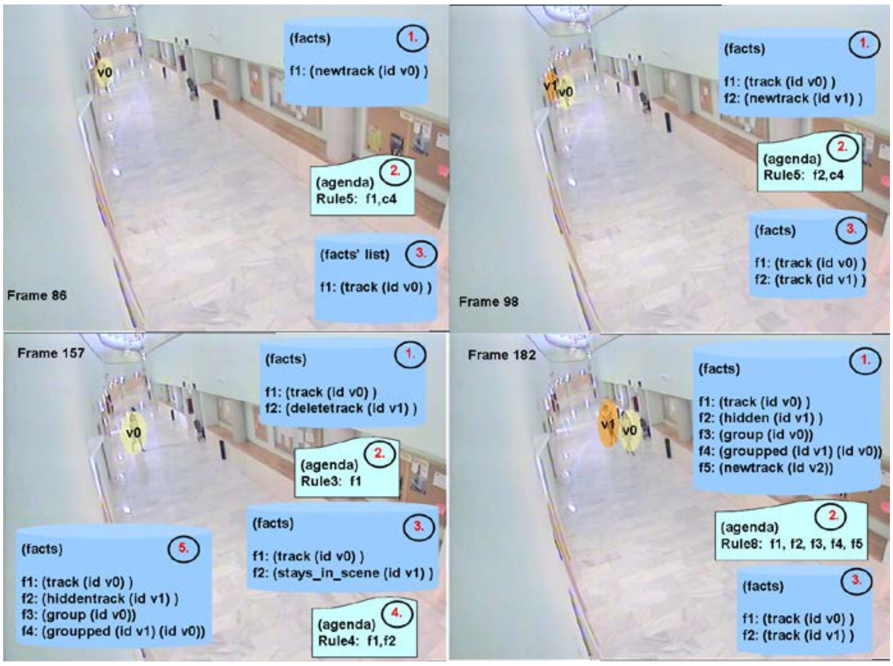


Fig. 11. Systems state at frames 86, 98, 157 and 182

Once the initial facts have been loaded into the system, the tracking system analyzes the video input. The detection of new tracks adds a new fact into the knowledge base, this fact labelled as f_1 is (*newtrack (id v0)*). Once this fact f_1 is inserted into the Fusion Center, a rule is activated (as shown in Figure 11, frame 98), due to two primitive facts f_1 and c_4. The execution of this rule eliminates the previous f_1 and introduces a new fact f1: (*track (id v)*)) to confirm the track.

When a new object enters the scene, right before the new object has been detected, the only fact in the knowledge base made reference to the existence of a track with id 0, as the sensor detects the new object, it adds a new fact into the knowledge base f_2 (frame 98). A rule modifies the knowledge base by eliminating the fact introduced previously and inserting a fact that establishes the existence of a track with id 1 in the scene. At frame 157 track v_1 disappears from the scene, hidden behind track v_0, at that point the sensor processor deletes the track v_1 and inserts into the knowledge base fact f_3 (frame 157). This new fact activates rule 3, which modifies the knowledge base as shown in at frame 157, box 3. The facts establish that there is a track with id 0 visible in the scene and another track with id 1 in the scene but not visible. This modification of the knowledge base activates rule 4 which modifies the knowledge base to establish that the track with id 0 is a group and grouped with it is the track with id 1. Finally, the situation is restored later at frame 182, when a new track is initialize but the system reassigns it to the hidden track kept in the group.

5. Conclusion and further work

In this contribution, we presented a context model handled by a reasoning system to improve the data acquisition in local sensors and drive the fusion process. This context-based system models the human capacity to adapt the low-level processing. The reasoning capabilities of the system have been tested with isolated sensors and are currently under experimentation, to analyze the potential improvement in fusion performance in multi-camera environments. Improvement at higher levels of cognitive processing is an open possibility also implied in the model.

References

[1] M. Guerriero, P. Willet, and S. Coraluppi. Radar-AIS Data Fusion and SAR tasking for Maritime Surveillance. ICIF 2008 – Cologne, Germany, 1-3 Jul. 2008.

[2] United States Coast Guard. The U.S. Coast Guard Strategy for Maritime Safety, Security, and Stewardship. Washington D.C., 2007.

[3] T. Standage. The Turk: the Life and Times of the Famous Eighteenth-Century Chess-Playing Machine. Walker & Company, 2002.

[4] The Advanced Chess Organization (CCO). Advanced chess description, 2004. On line at http://advancedchess.netfirms.com/description.htm.

[5] E. P. Blasch and S. Plano. Level 5: User refinement to aid the Fusion Process, in Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Dasarathy., B. V. (Ed.), Proceedings of SPIE Vol. 5099, 2003.

[6] D. L. Hall and J. Llinas. An Introduction to Multisensor Data Fusion. Proceedings of the IEEE, Vol. 85, No. 1, January 1997.

[7] L. A. Lisboa Cardoso, Jesús García, José M. Molina. Requirements for supervised fusion adaption. International Symposium on Distributed Computing and Artificial Intelligence, DCAI 2008, University of Salamanca, Spain, 22th-24th October 2008.

[8] O. Brdiczka, P.C. Yuen, S. Zaidenberg, P. Reignier, and J.L. Crowley. Automatic acquisition of context models and its application to video surveillance. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), 2006.

[9] L. Morency, C. Sidner, C. Lee, and T. Darrell. Contextual recognition of head gestures. Proc. of ICMI, 2005.

[10] M. Xu and T. Ellis. Augmented tracking with incomplete observation and probabilistic reasoning. Image and Vision Computing, vol. 24:pp. 1202–1217, 2006.

[11] Tao Zhao and Ram Nevatia. Tracking multiple humans in complex situations. IEEE Trans. Pattern Analysis and Machine Intelligence, 26(9):1208–1221, 2004.

[12] A.M. Sanchez, M.A. Patricio, J. Garcia, and J.M. Molina. Video tracking improvement using context-based information. In The 10th International Conference on Information Fusion, Quebec, July 2007.

[13] A. Torralba, K.P. Murphy, W.T. Freeman, and M.A. Rubin. Context-based vision system for place and object recognition. Proc. of ICCV, 2003.

[14] K. Alexiev and I. Nikolova. Methods for Data and Information Fusion, in Scientific Support for the Decision Making in the Security Sector, O. Kounchev et al. (Eds.), IOS Press, 2007.

[15] M. R. Endsley. Theoretical Underpinnings of Situation Awareness: A Critical Review. Situation Awareness Analysis and Measurement, Lawrence Erlbaum Associates, 2000.

[16] F. Castanedo, J. Garcia, M.A. Patricio and J.M. Molina. Data Fusion to improve trajectory tracking in Cooperative Surveillance Multi-Agent Architecture. Information Fusion. Special Issue on Agent-based Information Fusion (to appear). Aug. 2008.

[17] A.M. Sanchez, M.A. Patricio, J. Garcia, and J.M. Molina. Occlusion Management using a Context-based Tracking System. In 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence (AITAmI'08) Special Session on Vision-based Reasoning Co-located event of European Conference on Artificial Intelligence. Patras, Greece, 21-22 July 2008.

[18] A. M. Sanchez, M. A. Patricio, J. Garcia, JM Molina. "A Context Model and Reasoning System to Improve Object Tracking in Complex Scenarios". Expert Systems With Applications. Vol 41 (in press)

[19] Ana M. Sánchez, Miguel A. Patricio, J. García and J.M. Molina. Extending Local Contextual Reasoning to Multicamera Environments. SWIFT 2008 - Skövde Workshop on Information Fusion Topics. SWIFT 2008. Skovde, Sweden, 4-6 Nov 2008.