Data Fusion to improve trajectory tracking in Cooperative Surveillance Multi-Agent Architecture ⋆

Federico Castanedo, Jesús García, Miguel A. Patricio and José M. Molina

Applied Artificial Intelligence Group (GIAA)
http://www.giaa.inf.uc3m.es
Universidad Carlos III de Madrid
Avda. Universidad Carlos III 22, 28270-Colmenarejo, Spain

Abstract

In this paper we present an extension of Cooperative Surveillance Multi-Agent System (CS-MAS) architecture to incorporate dynamic coalition formation. A specific coalition formation using fusion skills is shown so that the fusion process is distributed now in two layers: (i) a global layer in the fusion center, which initialize the coalitions and (ii), local layer within coalitions, with the dynamic instantiation of a local fusion agent. There are several types of autonomous agents: surveillance-sensor agents, fusion center agent, local fusion agent, interface agents, record agents, planning agents, etc. Autonomous agents differ in their ability to carry out a specific surveillance task. A surveillance-sensor agent controls and manages individual sensors (usually video cameras). It has different capabilities depending on its functional complexity and limitation related to specific sensor nature aspects. In this work we add a new autonomous agent called local fusion agent to the CS-MAS architecture, addressing specific problems of on-line sensor alignment, registration, bias removal and data fusion. The local fusion agent it is dynamically created by the fusion center agent and involves several surveillance-sensor agents working in a coalition. We show how the inclusion of this new dynamic local fusion agent guarantee that, in a video-surveillance system, objects of interest are successfully tracked across the whole area, assuring continuity and seamless transitions.

Key words: Software Agents, Data Fusion, Coordination, Sensor Network Architecture, Distributed Vision, Surveillance Application

⋆ Funded by projects CICYT TSI2005-07344, CICYT TEC2005-07186, CAM MADRINET S-0505/TIC/0255 and IMSERSO-AUTOPIA

Email address: {`fcastane,jgherrer,mpatrici`}`@inf.uc3m.es`, `molina@ia.uc3m.es` (Federico Castanedo, Jesús García, Miguel A. Patricio and

1 Introduction

Video surveillance is without any question a powerful tool for public safety and security, and with the increasing need for more security in airports [1], sea environments [2,3], railways, underground [4–8], and other critical environments, the demand for video developments is growing rapidly. Typical examples of commercial surveillance systems are DETEC [9] and Gotcha [10]. They are usually based on what is commonly called motion detectors, with the option of digital storage of the detected events (input images and time-stamped metadata). For other surveillance applications (e.g. road traffic, ports, and railways), see [2,5,8,11,12]. Many of these systems require a wide geographical distribution that calls for camera management and data communication. In this way, [5] propose combining existing surveillance traffic systems based on networks of smart cameras. The term 'smart camera' (or 'intelligent camera') is normally used to refer to a camera that has processing capabilities (either in the same casing or nearby) so that event detection and event video storage can be done autonomously by the camera.

Visual sensor networks [13] are related to spatially distributed multi-sensor environments which raise interesting challenges for surveillance. These challenges concern to data fusion techniques to deal with the sharing of information gathered from different types of sensors [14], communication aspects [15], security of communications [15] and sensor management. These new systems are called "third-generation surveillance systems", which would provide highly automated information, as well as alarms and emergencies management. PRISMATICA [4] is an example of these systems. It consists of a network of intelligent devices that process sensor inputs. These devices send and receive messages to/from a central server module. The server module co-ordinates device activity, archives/retrieves data and provides the interface with a human operator. The design of a surveillance system with no server to avoid this centralization is reported in [16]. All the independent subsystems are completely self-contained, and all these nodes are then set up to communicate with each other without having a mutually shared communication point. As part of the VSAM project, [16] presents a multi-camera surveillance system based on the same idea as [17]: the creation of a network of 'smart cameras' that are independent and autonomous vision modules. The surveillance systems described above take advantage of progress in low-cost high-performance processors and multimedia communications. However, they do not account for the possibility of make temporal groups of 'smart cameras' which cooperate each other for specific tasks. For example, fusing information from cameras, as in this work. In this work we called dynamic coalitions as a temporal groups of agents working together. An agent may need to cooperate in order to achieve better and

José M. Molina).

more accurate performance, or need additional capabilities that it does not have. This cooperation takes place doing a coalition formation which it is created by the fusion center agent. By cooperation we means sharing data and resolving conflicts. Current research is focusing on developing surveillance systems that consist of a network of cameras (monocular, stereo, static or PTZ (pan/tilt/zoom)) which uses in a static design, information from neighboring cameras. In a coalition, consistency across multiple camera views (either with shared or disjoint fields of view) can only be maintained when spatial and time coherence is achieved. Otherwise, biased local tracks may lead to erroneous inconsistency alarms and splitting effects corresponding to views from different cameras, or instabilities such as a "zig-zag" effect in the estimated trajectories.

In [13], authors have developed a novel multi-agent framework for deliberative camera-agents forming visual sensor networks. In this framework, each camera is represented and managed by an individual software agent, called surveillance-sensor agent. A software agent [18] is a computational process which has several characteristics: (1) *"reactivity"* (allowing agents to perceive and respond to a changing environment), (2) *"social ability"* (by which agents interact with other agents) and (3) *"proactiveness"* (through which agents behave in a goal-directed way). Surveillance-sensor agents are located at the same level (sensor layer), so that it allows the coordination of the execution among surveillance-sensor agents. Each surveillance-sensor agent knows only part of the information (partial knowledge due to its limited field of view), and has to make decisions with this limitation. The distributedness of this type of systems supports the surveillance-sensor agents' proactivity, and the cooperation required among these agents to accomplish surveillance justifies the sociability of surveillance-sensor agents. The intelligence produced by the symbolic internal model of surveillance-sensor agents is based on a deliberation about the state of the outside world (including its past evolution), and the actions that may take place in the future. The architecture used in [13] to describe the behavior of surveillance-sensor agents was Belief-Desire-Intention (BDI) model [19].

In this work, authors extend the architecture presented in [13], Cooperative Surveillance Multi-Agent System (CS-MAS). This architecture is a logical framework of autonomous agents working in sensor network environments and it is implemented using the BDI model. It is composed of several agent types: surveillance-sensor agent, that controls a specific sensor; fusion center agent, which has a global view of the environment being monitoring; interface agent, that uses as input/output of surveillance system; record agent, which is in charge of record a specific Digital Video Record (DVR) device [20]; etc. The fusion center agent has the global view of the environment being monitoring, also it is in charge of the creation of the local fusion agent and the coalition. The criteria for the creation of coalitions are: prediction of hand-over between surveillance-sensor agents, events triggered with contextual information, and the detection of nearby tracks.

Each agent has different capabilities for specific surveillance tasks. In the case of surveillance-sensor agent, for instance, its capabilities will depend on the sensor nature and, in general, the abilities of a generic agent. Then, the concept of coalition appears when a group of autonomous agents choose to work together to achieve a temporal common goal. The process to make a coalition is called Coalition Formation (CF), and it is widely studied [21–24], although there are few works related to surveillance systems [20]. A CF starts at a certain moment to achieve one task, and when this task ends the coalition breaks off. In Fig. 1 we show an illustration of a surveillance deployment with overlapped field of views, these overlapped areas could be exploited with a coalition in order to obtain more accurate results and guarantee a coherent monitoring in the global area. The tracking algorithms implemented in the surveillance-sensor agents have to deal with motion detection errors and complex object interactions (merging, occlusions, fragmentation, etc.). The new local fusion agent combines the information inferred by the individual surveillance-sensor agents to maximize the final information content about the area to guard. In this example, the coalition would be created by the fusion center agent when it predicts an object will enter in the common area of several sensors (S1, S2, S3). Then, the local fusion agent is dynamically created and performs registration and data fusion of the information of the surveillance-sensor agents that have joined to the coalition.

The main contributions of this work are: (1) the description of the software architecture able to make dynamic coalitions in visual sensor networks, (2) the presentation of a general coalition formation framework and finally (3) the example of a coalition formation process for a specific surveillance task: data fusion for tracking in overlapped areas of surveillance-sensor agents.

In section 2 a brief description of the Cooperative Surveillance multi-agent architecture is shown. Section 3 gives the formal description of the architecture and the coalition formation mechanism. Then, the details of a coalition formation process for data fusion is described. The fusion process to track the objects of interest when the coalition is established is described in Section 5. Then, we illustrate, in a sample scenario, the fusion capability of our extended CS-MAS and evaluate its performance in coordinated tracking. Finally we include some conclusions and future work.

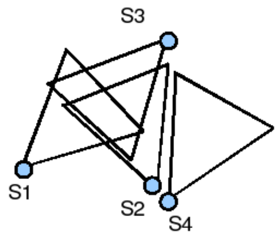

Fig. 1. An example of a surveillance deployment with overlapped areas. The sensors S1, S2 and S3 could form a coalition to monitor the overlapped area.

2 Cooperative Surveillance Multi-Agent Architecture

2.1 General overview

Next, we show a brief description of the different types of the autonomous agents belonging to the multi-agent system (see Fig. 2):

- Surveillance-Sensor Agent: It tracks all the targets moving within its local field of view (FoV) and sends data to the fusion center agent. It also sends information to the context agent. It is coordinated with other agents in order to improve surveillance quality. It has different roles (individualized agent, object recognition agent, face recognition agent) each one with different specific capabilities. It is possible to change the role but at any given moment it could be only in one role.
- Center Fusion Agent: Integrates the information sent from the associated surveillance-sensor agents. It analyzes the situation in order to manage the resources and coordinating the surveillance-sensor agents. This agent has the global view of the environment being monitoring by all the surveillance-sensor agents. It is in charge of creating the dynamic coalitions of surveillance-sensor agents using contextual information and the prediction of certain situations requiring a cooperative fusion process.
- Local Fusion Agent: It is dynamically created by the fusion center agent and performs data fusion of the surveillance-sensor agents information in a certain coalition. By coalition we means a group of autonomous agents working temporarily together.
- Record Agent: This type of agent belongs to an specific camera only with recording features [20].
- Planning Agent: It has a general vision of all the scene. It makes inferences on the targets and the situation.
- Context Agent: It provides dependent information of the context where the monitoring is being made. This agent indicates the semantic distance between different surveillance-sensor agents. By semantic distance we mean that in spite of the physical distance between two cameras, they are moni-

toring the same scene and maybe the same targets. The context agent stores information about static objects which could provoke partial occlusions to the tracked targets but it also stores dynamic information about the scene (i.e. A truck is occluding one target) [25].

- Interface Agent: The input/output agent interface of the multi agent system. It provides a graphical user interface which shows the evolution of the targets that are being tracked.

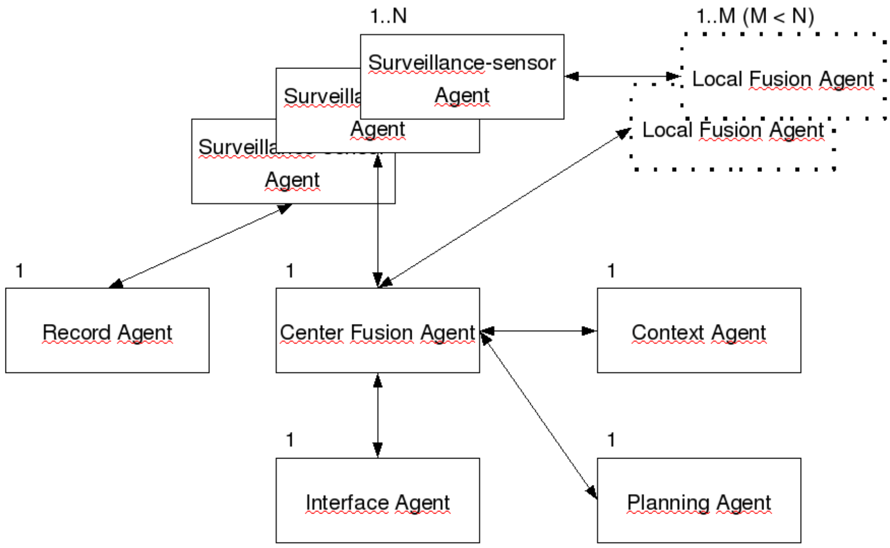

Fig. 2. Different types of agents in CS-MAS

In Fig. 3 the new CS-MAS architecture is depicted. It has three layers: (1) sensor, (2) local fusion Layer and (3) fusion center layer. In sensor layer, each sensor is controlled by an autonomous agent. At this level, autonomous agents can cooperate with others agents (through dynamic coalitions) to use capabilities of other agents in order to carry out tasks that they are not able to achieve alone [13] or to improve the same capability. In this work, we develop a local fusion layer in the CS-MAS architecture. This layer includes a new dynamic local fusion agent which is in charge of fuse several sensor agents data with the specific goal of achieving better performance or accuracy for specific surveillance tasks. The local fusion agent is dynamically created by the fusion center agent as we show further in this paper. Finally, the center fusion layer is composed by the center fusion agent, which has a global view of the environment being monitoring.

In video surveillance systems, trajectory tracking is employed to identify individual objects and keep a temporal history of their evolution within the guarded areas. So, we present how our CS-MAS architecture improves trajectory tracking by means of data fusion from several neighboring surveillance-sensor agents (camera agents in a visual sensor network) which are in a coalition. All the sensors that have a common target form the neighborhood and the neighborhood is the set of sensors that could make a coalition. One of the aims of the fusion center agent is to guarantee that objects of interest are successfully tracked across the whole area, assuring continuity and seamless transitions. Besides, the tracking problems presents with specific surveillance-sensor agents (false alarms, uncertainty in data, for example) are solved through cooperative tracking.

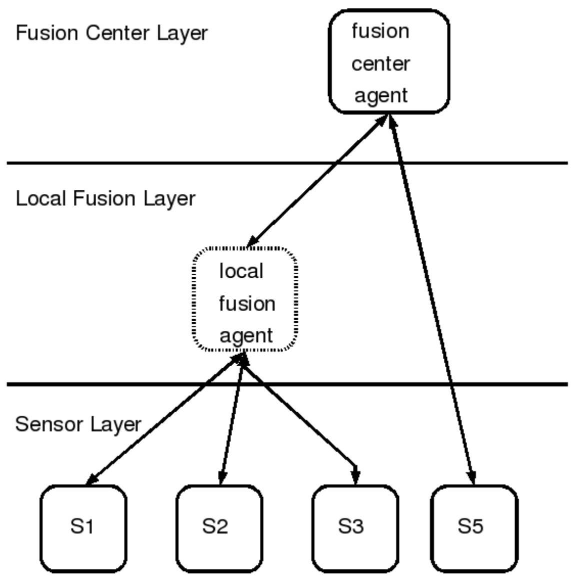

Fig. 3. CS-MAS Logical Layers. In sensor layer, C1, C2 and C3 are examples of surveillance-sensor capabilities and memory is the knowledge of the agent. Coalition layer, depicts an example of three surveillance-sensor agents working together.

2.2 BDI Description

In order to act rationally, the BDI model [19] represents internally the situation faced and the mental state in form of beliefs, desires and intentions. Each agent has its owns set of beliefs, desires and intentions. The state of the agent at any given moment is a triple (B, D, I), where $B \subseteq Beliefs$, $D \subseteq Desires$ and $I \subseteq Intentions$. The cooperation between autonomous agents takes place for the purpose of improving their local information, and it is achieved by message exchange. Currently we are not using any ontology for the information exchanged between the autonomous agents. However, the content of each message is a specific implementation of a Java class which describes the information. Therefore agents exchange the information using instances of Java classes and FIPA standard messages. If we focus in the surveillance-sensor agent, CS-MAS establishes its beliefs, desires and intentions as:

Beliefs. Surveillance-sensor agent beliefs should represent information about the outside world, like objects that are being tracked, other known autonomous agents who are semantically close and their execution state, and geographic information including location, size and trajectory of the tracked objects, location of other elements that might require special attention, such as doors and windows, and also obstacles that could occlude targets of interest (for instance, tables, closets). On the other hand, its own and neighborhood capabilities are represented in the Surveillance-Sensor Agent beliefs.

Desires. Surveillance-sensor agent has two main desires as the final goal of a surveillance-sensor agent is the correct tracking of moving objects: permanent surveillance and temporary tracking. The corresponding surveillance plan is: the surveillance-sensor agent permanently capture images from the camera until an intruder is detected (or announced by a warning from another surveillance-sensor agent). On the other hand, the tracking plan is initiated by some event (detection by camera/warning from another agent), and it runs a tracking process internally on the images captured from the camera until the tracking is no longer possible. Also it has the desire of cooperate with other agents and to correct the information with the feedback messages.

Intentions. Intentions are the basic steps the agent choose at any moment in order to achieve it desires. There are two basic intentions: external and internal actions. External actions correspond to communication acts with other autonomous agents that implement different cooperative dialogs, while internal actions involve commands to the tracking system, and even to the camera.

The foundation for most implemented BDI systems is the abstract interpreter proposed by Rao and Georgeff [26]. Although many ad-hoc implementations of this interpreter have been applied to several domains, such as dMARS [27], the release of JADEX [28] has recently gained a quick acceptance. JADEX is an extension of JADE [29], which facilitates FIPA communications between agents, and it is widely used to implement intelligent and software agents. We choose JADEX as the underlying framework to develop CS-MAS.

3 CS-MAS: Formal Representation

Next, we present a formal representation of the beliefs, desires and intentions proposed to make coalitions and coordinate the relationships among surveillance-sensor agents, fusion center agent and local fusion agents.

Let n be the number of autonomous agents in the set A representing the multi-agent system, $A = \{A_1, A_2, \ldots, A_n\}$.

Each agent A_i has a set of m possibles capabilities $C = \{C_1, C_2, ..., C_m\}$. In surveillance systems, these capabilities are for example, tracking capability, event recognition capability, recording capability, projection capability, fusion capability, etc.

If we particularize to our surveillance multi agent system, S is the set of autonomous surveillance-sensor agents, $S = \{S_1, S_2, \ldots, S_n\}$.

For each surveillance-sensor agent in the CS-MAS framework, *Beliefs* represents the knowledge about its own capabilities, neighbor surveillance-sensor agents capabilities and the environment information gathered by its sensor.

Let Θ_i be the neighborhood of a surveillance-sensor agent S_i, where $(\Theta_i \subseteq S) \wedge (\Theta_i \neq \emptyset)$. The neighborhood of an agent S_i are all the other surveillance-sensor agents which share the field of view and they could apply the same capability (i.e. tracking the same target) at this moment. The distance between the surveillance-sensor agents is a semantic distance. For example, if two surveillance-sensor agents are monitoring the same scene, but one of them have an occlusion of the target (i.e. a truck behind the target) it causes an infinite semantic distance between them. So, the neighborhood is dynamic, and it is completely defined only in the moment when the coalitions are formed, when all agents confirm that they are able to apply a common capability to the same object.

We can represent each surveillance-sensor agent *Beliefs*, $\forall i : S_i \in S$, as:

- $\forall i \cdot (Bel\ S_i\ \{O_1(t), ...O_n(t)\})$, the information of the current surveillance-sensor agent about environment at time t. By environment information we means a list of detected objects (tracks). For each track the surveillance-sensor agent's beliefs are: position, size, velocity, etc.
- $\forall i \cdot (Bel\ S_i\ C_i)$, the knowledge about its own capabilities.
- $\forall i \cdot (Bel\ A_S\ \forall j \in \Theta_i \cdot (Bel\ S_j\ C_j)$, the surveillance-sensor agent S_i knows its neighbor surveillance-sensor agents capabilities.

The fusion center agent *Beliefs* can be represented as:

- $\forall i \cdot (Bel\ S_i\ \{O_1(t), ..., O_n(t)\})$, the fusion center agent receives the tracks of all the surveillance-sensor agents which are being monitored in the environment. Therefore this agent has the global view knowledge.
- $\forall i \cdot (Bel\ S_i\ Context())$, the context information of each surveillance-sensor agent, which are used to infer situations for making coalitions.
- Let k be equal the number of local fusion agents at one moment in the multi agent system and F the set of autonomous local fusion agents $F = \{F_1, F_2, \ldots, F_k\}$, then $\forall k \cdot (Bel\ \forall S_i \in\ Coalition(F_k))$, the fusion center agent knows the surveillance-sensor agents involved in a coalition for fusion purposes.
- $\forall i \cdot (Bel\ S_i\ Capabilities())$, the fusion center agent knows the capabilities of all the surveillance-sensor agents in CS-MAS. It also knows the neighbor-

hood of each agent. The definition of neighborhood it is based on a semantic distance rather than the physical distance of each sensor. Therefore each agent neighbours are defined in the design of the surveillance system.

We can represent each local fusion agent *Beliefs* as:

- Let Δ_i be the subgroup of surveillance-sensor agents that are being fused by the local fusion agent F_i, where $(\Delta_i \subseteq S) \wedge (\Delta_i \neq \emptyset)$, then $\forall i \cdot (Bel\ F_i\ \forall j \in \Delta i \cdot (Bel\ S_j\ \{O_1(t), ...O_n(t)\})$, all the local fusion agents knows the tracks information of the surveillance-sensor agents that are being fused.

3.1 Agents Definition in CS-MAS

Temporarily, the autonomous agents are able to work together forming a group with neighbor agents in order to act cooperatively and achieve their collective goals. In CS-MAS architecture, this cooperation mechanism is carried out by coalition formation. The coalition formation process is initiated by the fusion center agent. This agent detects the necessity of making a coalition. In the context of data fusion, coalitions are dynamically created in the coalition layer as temporal groups of surveillance-sensor agents working together. When the fusion center agent detects deviations in the tracking process or when an object could be tracked by two or more agents, a local fusion agent is dynamically created by the fusion center agent.

Let O be the set of targets at time t: $O = \{O_1^t, O_2^t, ..., O_j^t\}$.

Definition 1 *Apply* (S_i, C_l, O). It is a function that apply capability l of surveillance-sensor agent i on the set of targets O at time t.

$$Apply : S_i \times C_l \times O_j^t \rightarrow Boolean \tag{1}$$

Definition 2 Coalition at time t is a triple $\Psi_i =< Co_i,\ C_l,\ O >$. Where $Co_i \subseteq A$ is a subset of autonomous agents such that at time t $\forall j \in Co_i$ $Apply\ (A_j, C_l, O)\ \ is\ true$. So, the group of agents in coalition Ψ_i work together temporarily in the same specific action for a group of targets (O)..

At this level, *Desires* capture the motivation of the agents, the final goal of each surveillance-sensor agent is the permanent surveillance of its environment. So, the *Desire* of our surveillance-sensor agents is:

- $\forall i \cdot (Des\ S_i\ Surveillance(O_k))$.

The *Desires* of the fusion center agent are: (1) to obtain a global view of all the objects being tracked, and (2) to make dynamic coalitions:

- $\forall i \cdot (Des\ Ia_i\ IntegrateData(O_k))$
- $\forall i \cdot (Des\ Ia_i\ MakeCoalitions(\Psi_i))$

The *Desires* of the local fusion agent are: (1) to make a consistency checking from the provided information , (2) compute corrections for time-space alignment (registration) and (3) to fuse the information received by the surveillance-sensor agents involved in the coalition:

- $\forall i \cdot (Des\ F_i\ ConsistencyChecking(O_k))$
- $\forall i \cdot (Des\ F_i\ Registration(O_k))$
- $\forall i \cdot (Des\ F_i\ FuseData(O_k))$ The data fusion is performed after a track-to-track association (match the same track from different surveillance-sensor agents) and vector combination.

Intentions are the basic steps the agent has chosen to do in order to achieve its *Desires*. The surveillance-sensor agents intentions are:

- $\forall i \cdot (Int\ S_i\ \exists j \in \Theta_i \cdot (AcceptCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of surveillance-sensor agent S_i to accept making a coalition with other surveillance agent S_j that involve the set of targets O_k at time t in order to apply the capability C_l. This intention triggers the Accept-Coalition message which is described in the Protocol for Coalition Formation section.
- $\forall i \cdot (Int\ S_i\ \exists j \in \Theta_i \cdot (DenyCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of surveillance-sensor agent S_i to deny making a coalition with other surveillance-sensor agent S_j that involves the target O_k at time t in order to apply the capability C_l. This intention triggers the Reject-Coalition message which is described in section 3.2.
- $\forall i \cdot (Int\ S_i\ LeaveCoalition(S_i,\ O_k^t,\ C_l)))$, the intention of surveillance-sensor agent S_i to leave coalition that involves the target O_k at time t and the capability C_l. This intention triggers the Cancel-Coalition message.
- $\exists i \cdot (Int\ S_i\ Tracking(O_k))$, in the multi agent system exists at least one surveillance-sensor agent with tracking capability.
- $\exists i \cdot (Int\ S_i\ Recognition(O_k))$, in the multi agent system exists at least one surveillance-sensor agent with recognition capability.
- $\forall i \cdot (Int\ S_i\ Projection)$, the surveillance-sensor agents have the Projection capability. The projection capability projects the local tracking information into a global common reference.
- $\forall i \in \Psi_i \cdot (Int\ S_i\ \cdot (SendTargetInfo(F_j,\ O_k^t,\ C_l)))$, all the surveillance-sensor agents involved in a coalition can communicate to the local fusion agent F_j the information about target O_k at time t. This intention triggers the Inform-Coalition message which is described in section 3.2.
- $\forall i \cdot (Int\ S_i\ \cdot (SendTargetInfo(Fc,\ O_k^t,\ C_l)))$, all the surveillance-sensor agents can communicate to the fusion center agent Fc the information from the capability C_l applied to target O_k at time t. For example, if the surveillance-sensor agent is performing the tracking capability it sends track-

ing information to the fusion center agent. Therefore the fusion center agent has a global view of the tracking being performed by all the surveillance-sensor agents.

On other hand, the fusion center agent intentions are:

- $\forall i \cdot (Int\ Ia_i\ \forall j \in \Theta_j \cdot (MakeCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of the fusion center agent Ia_i to make a coalition with the surveillance-sensor agents S_j belongs to Θ_j that involves the target O_k at time t in order to apply the capability C_l. For example, in this work we apply the coalition formation process for the fusion capability, but coalitions with different purposes could be implemented. This intention involves the creation of the local fusion agent in the system if the Call-for-Coalition message was succeeded.
- $\forall i \cdot (Int\ Ia_i\ \forall j \in \Theta_j \cdot (AskForCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of the fusion center agent Ia_i to ask for make a coalition to the surveillance-sensor agents S_j belongs to Θ_j that involves the target O_k at time t in order to apply the capability C_l. This intention send it the Call-for-Coalition message with the objective to make the coalition.
- $\forall i \cdot (Int\ Ia_i\ \forall j \in \Theta_j \cdot (ReAskForCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of the fusion center agent Ia_i to re-ask for make a coalition to the surveillance-sensor agents S_j belongs to Θ_j that involves the target O_k at time t in order to apply the capability C_l. This intention involves to repeat the process for the establishment of the coalition, since the surveillance-sensor agents could deny it. If a DenyCoalition() message is received, the fusion center agent repeat the process after some delay (which means send again the Call-for-Coalition message)
- $\forall i \cdot (Int\ Ia_i\ \forall j \in \Psi_j \cdot (CreateFusionAgent(S_j,\ O_k^t,\ FusionCapability)))$, the intention of the fusion center agent Ia_i to make a local fusion agent which fuse the information of the surveillance-sensor agents in the coalition.
- $\forall i \cdot (Int\ Ia_i\ \forall j \in \Psi_j \cdot (DestroyCoalition(S_j,\ O_k^t,\ C_l)))$, the intention of the fusion center agent Ia_i to destroy a coalition with the surveillance-sensor agents S_j belongs to Ψ_j that involves the target O_k at time t in order to apply the capability C_l. In this work, when the number surveillance-sensor agents involved in a coalition reach only one agent, the coalition is destroyed and therefore the local fusion agent.

Local fusion agent's intentions in F are similar, but two important *Intentions* are:

- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (FusionTargetInfo(S_j,\ O_k^t,\ C_l)))$, the intention to receive and fuse information from the surveillance-sensor agent S_j belonging to the same coalition about the target O_k. at time t.
- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (InformBrokenCoalition(S_j,\ O_k^t,\ C_l)))$, the intention to inform to the surveillance-sensor agent S_j that the coalition is broken. This intention is is also used to inform about that the coalition is broken to

the fusion center.

- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (ConsistencyChecking(S_j,\ O_k^t,\ C_l)))$, all the local fusion agents have the intention to check the consistency of all the surveillance-sensor agents involved in a coalition (Ψ_j) for an specific capability.
- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (Registration(S_j,\ O_k^t,\ C_l)))$, all the local fusion agents have the intention to perform a registration process of all the surveillance sensor agents involved in a coalition (Ψ_j).
- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (TrackToTrackAssocciation(S_j,\ O_k^t))$, all the local fusion agents have the intention to carry out a track-to-track association between the tracks received by the surveillance-sensor agents in the coalitio (Ψ_j).
- $\forall i \cdot (Int\ F_i\ \forall j \in \Psi_j \cdot (SendFeedbackInfo(S_j,\ O_k^t,\ C_l)))$, the intention to send to each surveillance-sensor agent S_j in the coalition Ψ_j feedback information about the capability accuracy regarding to the other agents in the coalition.

3.2 Protocol for Coalition Formation

In [30], the author argues that the first step in a cooperative problem solving process begins when some agent recognizing the potential for cooperative action. In CS-MAS it begins when the fusion center agent detects the need to make a coalition, therefore the fusion center agent is the initiator of the coalition. The needs for coalition formation and data fusion are centralized in the fusion center agent. It is in charge of recognizing the potential for making coalitions in situations of tracking deviations of individual surveillance-sensor agents, hand-overs between adjacent coverage areas, and events depending on contextual situations.

In the initial moment the cooperation will exist only in the mental state of the fusion center agent that initiates the process, we call this recognition of cooperation: $\exists Fc \cdot (Bel\ Fc\ \exists j \in \Theta_j\ \cdot (Bel\ A_j\ MakeCoalition(A_i, O_k^t, C_l)))$ that means that fusion center agent Fc belief that another agent A_j exist and may want to make a coalition (with other agent A_i) for the groups of targets $\{O_k\}$ at time t and applying capability C_l.

Then the fusion center agent Fc send a set of messages to other agents (all of them shared the same neighborhood), in this case A_j, in order to complete a coalition with success. Each message is performed by the $MakeCoalition$ intention and using the concept of neighborhood (agents with a small semantic distance).

- *Call-for-Coalition.*
  $< Fc,\ cfp\ (A_j,\ < A_j, MC >, Ref\ A_j O_k^t C_l,\ \phi(A_j, O_k^t, C_l)) >$

$\equiv < Fc, query - ref\ (A_j, Ref\ A_j O_k^t C_l (I_{Fc}\ Done$
$(< A_j, MC >, \phi(A_j, O_k^t, C_l)) \Rightarrow$
$(I_{Fc}\ Done\ (< A_j, MC >,\ \phi\ (A_j, O_k^t, C_l))))) >$
where MC stands for *MakeCoalition* action. cfp (Call-for-Proposal) is used to check the availability of an agent to perform the *MakeCoalition* action (see FIPA Communicative Act Library Specification [31]) $query - ref$ is the action of asking another agent for the object referred to by a referential expression. Ref is the referential expression, in this case apply a specific capability to a specific object at time t $(A_j O_k^t C_l)$. *Done* indicates the action to be done, in this case make the coalition. I indicates the intention of performing some action.

We assume that all the surveillance-sensor agents asked to perform a coalition formation are consistent each other. For example, if Fc asked for coalition with S_1, S_2 and S_3, all of them must be able to form a coalition. Therefore the fusion center agent must have a table with all the possible options for coalition formation. These messages fit the FIPA standard that adds a performative to each communicative act.

When the agent A_j receives the message it has two possibilities: accept or reject the coalition proposal. The Accept-Coalition message is performed by the *AcceptCoalition* intention and the Reject-Coalition by the *DenyCoalition* intention.

- *Accept-Coalition.*
  $< A_j,\ accept - proposal\ (Fc < Fc, MC >, \phi(A_j, O_k^t, C_l)) >$
  $\equiv < A_j, inform(Fc, I_{Fc}\ Done(< Fc, MC >,\ \phi(A_j, O_k^t, C_l))) >$

  where MC stands for *MakeCoalition* action. $accept - proposal$ is a general purpose acceptance of a proposal that was previously submitted (typically through a propose act), in this case the *MakeCoalition* action. The agent sending the acceptance (A_j) informs the receiver (Fc) that it intends that (at some point in the future) the receiving agent will perform the action (MC), once the given precondition is $(\phi(A_j, O_k^t, C_l))$, or becomes, true.

- *Reject-Coalition.*
  $< A_j,\ reject - proposal\ (Fc,\ < Fc, MC >,\ \phi(A_j, O_k^t, C_l)\ ,\ \Gamma) >$
  $\equiv < A_j, inform(Fc,\ \neg I_{A_j} Done(< A_j,\ MC >, \phi(A_j, O_k^t, C_l))\ \wedge \Gamma) >$

  Agent A_j informs agent Fc that, because of proposition Γ, A_j does not have the intention for Fc to perform *MakeCoalition* action with precondition $\phi(A_j, O_k^t, C_l)$.

A Corollary of the fact that agents are autonomous is that Coalition Formation processes may fail. This could happen if the surveillance-sensor agents involved sends a Reject-Coalition message (we must need at least two agents for make a coalition). If the Coalition Formation process fail, the fusion center agent (who is the initiator of the process) wait some random time and then restart the process again. The restart process fails until it reaches some threshold (number of re-attempts). If the coalition formation was successfully completed the agents belonging to the same coalition must interchange messages about the same target, by sending Inform-Coalition messages (performed by the $SendTargetInfo$ intention):

- Inform-Coalition
  $< A_j, inform(Fc, \phi(A_j, O_k^t, C_l)) >$

  With this message the sender (A_j) informs the receiver (Fc) the value of the given preposition ($\phi(A_j, O_k^t, C_l)$).

Any agent could leave the coalition, it only need to send a message to the fusion center agent that manages the coalition, which is performed by the $LeaveCoalition$ intention:

- Cancel-Coalition
  $< A_j, cancel(Fc, MC) >\equiv$
  $< A_j, disconfirm(Fc, I_{A_j} Done(MC)) >$

  Agent A_j informs agent Fc that no longer intends to perform the action MC.

If the number of surveillance-sensor agents involved in the coalition reach one the local fusion agent informs to the surveillance-sensor agent and the fusion center agent that the coalition is broken. Also, the coalition could be broken by the local fusion agents if the consistency checking between the tracks is not satisfied. This action is performed by $InformBrokenCoalition$ intention which trigger the next messages:

- Inform-Broken-Coalition
  $< F_i, inform(A_j, Broken - Coalition(\Psi_i)) > < F_i, inform(Fc, Broken - Coalition(\Psi_i)) >$
  Agent F_i informs agent A_j and Fc that the coaliton Ψ_i is broken.

When the surveillance-sensor agents finished the coalition for a specific target, the information of that target is send to the fusion center agent instead of the local fusion agent. On other hand, while when they where involved in the coalition, the information is sent to the local fusion agent instead to the fusion center agent. In the case of the fusion center agent the Inform-Broken-Coalition message makes the intention $DestroyCoalition$ possible. In Fig. 4 we shown an example of the coalition formation protocol.

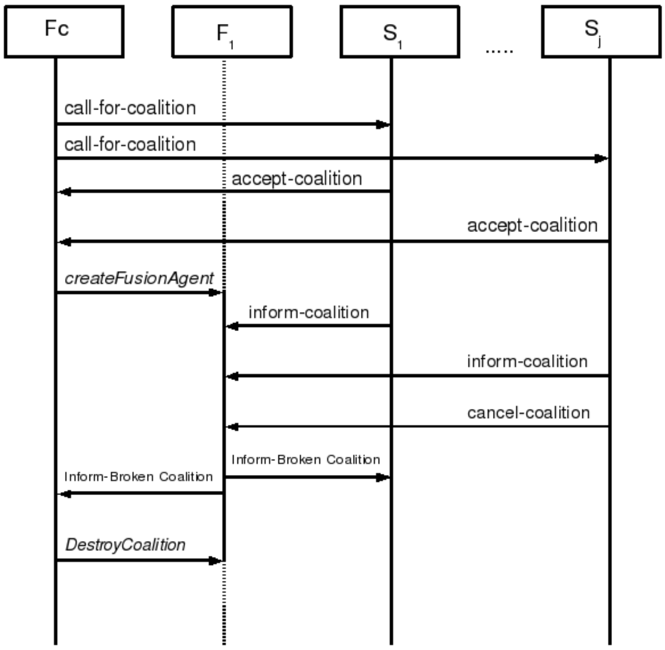

Fig. 4. A temporal example of messages exchanged during a coalition formation process. The words in italic corresponds to intentions. We ommited the parameters due to provide a better illustration.

4 Data Fusion for object tracking across multiple sensors in a Coalition

In this section we describe the fusion process to track the objects of interest while the coalition is active. The coalition includes several surveillance-sensor agents and a local fusion agent, whose results are reported to the fusion center agent. Other possibilities of coalitions in this context are explained in [13].

As mentioned, the visual sensors (surveillance-sensor agents) are deployed with their fields of view (FoVs) partially overlapped. This fact allows redundancy for smooth transitions across overlapped areas and continuity of targets along the whole area covered by the sensor network. The inter-sectional regions between cameras are very important here, they provide the data to compute corrections to refine the time-space alignment (inter-sensor registration). The basic aspects of the distributed fusion process carried out within the multi-agent coalition are the following ones:

- A local tracking process is performed (using the $Tracking()$ capability) by every surveillance-sensor agent S_i, whose result is expressed in global coordinates and sent to the local fusion agent F_j.
- The local fusion agent F_j performs the fusion algorithms (using $TrackTo$ $TrackAssociation()$, $Registration()$ and $FusionTargetInfo()$ intentions)

in the coalition: track-to-track association, consistency checking, inter-sensor registration and vector combination.
- The result of the fusion process is sent to the fusion center agent and, simultaneously, inter-sensor biases are estimated using available inter-sectional data through registration algorithm. These corrections are sent backward to surveillance-sensor agents (using $SendFeedbackInfo()$ intention) to keep global coherence of information shared within the coalition.
- In the case of consistency fail ($ConsistencyChecking()$ intention), it is informed the fusion center agent and surveillance-sensor agents to dissolve the coalition (using $InformBrokenCoalition$ messages) or remove the affected targets.

The first aspect corresponds to the first intention carried out by every surveillance-sensor agent in the cooperative architecture, the *tracking* capability. Each surveillance-sensor agent is assumed to measure the location of mobile targets within its field of view with respect to a common reference system (applying a specific *projection* capability). This is a mandatory step in visual sensor, since they must share common coordinates during the cooperative process.

Once a surveillance-sensor agent detect a new target in its field of view, it starts to perform the *Tracking* capability. The agent which starts the coalition formation, in this case the fusion center agent, is called as *agent initiator*. The agent initiator looks for the cooperation to track the new target through the mechanism described in section 3.2 Protocol for Coalition Formation. After the coalition is formed, data fusion techniques are needed to combine the local target information among the surveillance-sensor agents in the coalition.

Let S_i be a surveillance-sensor agent in the coalition, so that $Apply(S_i, C_k, O_j^t)$ *is* `true`, where C_k is the capability of tracking the new target O_j at time t. The agent S_i acquires images $I(i, j)$ at a certain frame rate, V_i. The interesting target O_j is represented with a track vector $\hat{x}_j^i[n]$, containing the numeric description of their attributes and state: location, velocity, dimensions, and associated error covariance matrix, $R_j^i[n]$. In an internal process, target location and tracking are expressed in pixel coordinates, which are local to each i-th camera agent view, S_i, and n is the temporal index associated to the time moment: t_n. More details of this local video tracking process are in [32,33].

Then, these local estimates (or track vectors) are projected to global coordinates as result of applying the *projection* capability. A frequent selection in the calibration process is the use of geodetic coordinates (those of GPS: latitude, longitude, altitude), known as Geo-location [16,34], with both objects location and cinematic descriptions expressed in a common framework. Regarding time reference and synchronization, video frame grabbers usually provide a sequence of frames, f_n, which must be time stamped in a common reference time basis. A possibility, knowing the initialization time, t_0, and the grabbing rate, V_i (frames per second), is computing $t[n]$ as $t_0 + f_n/V_i$. Alternatively, an external clock signal can be broadcast for time stamping, with synchronization mechanisms between different camera grabbers, this can be refined in the registration process described below.

As a result, the message $InformCoalition()$ is sent from surveillance-sensot agents to the local fusion agent in the coalition, containing the fields $\langle t[n], \hat{x}_l^i[n], R_l^i[n] \rangle$

4.1 Check Consistency

The $ConsistencyCheckingIntention()$ is applied over all tracks received by calculating the Mahanalobis Distance (MD) between all surveillance-sensor agent pairs (S_i, S_j) track features of all transformed vectors:

$$MD_{S_i,S_j} = \left(\hat{x}_l^i[n] - \hat{x}_l^j[n]\right)^t \left(R_l^i[n] + R_l^j[n]\right)^{-1} \left(\hat{x}_l^i[n] - \hat{x}_l^j[n]\right) \leq \lambda \tag{2}$$

If the MD exceeds the λ threshold, the track pair is labeled as inconsistent, raising the alarm to discard one of them from the fusion process. And the surveillance-sensor agent is warned receiving a message triggered by the $SendFeedbackInfo()$ intention.

4.2 Multi-camera registration

The inter-sensor registration is triggered by the $Registration()$ intention of the local fusion agent, in parallel with the combination of tracks received from local sensors. As mentioned previously, the local-to-global transformations are the result of a static calibration process, adjusted off-line in the system configuration phase, to have the output of diverse cameras systems expressed in the common extrinsic Cartesian reference frame. An additional on-line refinement must be carried out with the fusion data while the coalition is active, to remove residual systematic errors of slow time variation among the sources of information, and guarantee the stability of fusion. This process of dynamic multi-sensor alignment is an important component for sensor fusion is referred in sensor fusion literature to as multi-sensor registration [35–39]. As mentioned above, on-line solutions are needed to estimate on-line the potentially time-variant systematic errors in parallel with tracking, using the same available data. Both camera calibration and registration processes may be learned from observation data, being the difference that calibration is performed with static data (reference landmarks) while registration uses the dynamic trajectories to be fused.

Let $\hat{x}_j^i[n]$, $R_j^i[m]$ be the result of the projection and on-line registration in global coordinates for i-th surveillance-sensor agent at m-th frame, target j. In this on-line registration process, the bias estimation vectors, $\hat{b}^i[n]$, are computed to refine their alignment with respect to fusion solution, $\hat{x}_j^F[n]$, $R_j^F[n]$. The correction includes the time stamps to remove clock shifts among agents local processors.

A batch registration technique [37] is applied in the fusion agent to on-line estimate and cancel the systematic camera errors. Multi-camera data corresponding to the interval segments where local surveillance-sensor agent are consistently referred to common targets are stored in data blocks. The corrections to the transformation are estimated with a weighted least squares approach: it is computed the bias vector that, once applied to correct each camera agent output, minimize the weighted magnitude of the measurement difference vector, expressed in the central (fusion) coordinates.

Let $\Delta x_c^i[m]$ represent the difference vector between vector estimates provided by surveillance-sensor agent S_i, and the central track delivered by fusion agent, once applied the bias correction and coordinate transformations $f_i(.)$. This difference is weighted with $R_c^i[m]$, the error covariance matrix of the difference vector between measures provided by the surveillance-sensor agent S_i and local fusion agent. The goal function to be minimized is the squared module of the corrected difference vector, weighted with their error covariance matrices in order to consider the relative accuracy of the vector components: $(\Delta x_c^i)^T(R_c^i)^{-1}\Delta x_c^i$.

The correction vector, $\hat{b}_i$, contains the absolute correction in location $[b_x\ b_y]^t$, the parameters of the homography between 2D camera projection and fusion coordinates (3D assuming ground motion), $[a_{xx}\ a_{xy}\ a_{yx}\ a_{yy}]$, and local clock correction, b_t. With previous assumptions, the correction vector $\hat{b}$ has a linear relation with the fused location, $\left[x^F[k]\ y^F[k]\right]$ and local estimation at sensor S_i, $\left[x^i[k]\ y^i[k]\right]$, considering also velocity $\left[v_x^F[k]\ v_y^F[k]\right]$ to to make observable the time misalignment:

$$\begin{bmatrix} \Delta x_c^i[m] \\ \Delta y_c^i[m] \end{bmatrix} = \begin{bmatrix} x^F[m] \\ y^F[m] \end{bmatrix} - \begin{bmatrix} 1 & 0 & x_l^i[m] & y_l^i[m] & 0 & 0 & v_x^F[m] \\ 0 & 1 & 0 & 0 & x_l^i[m] & y_l^i[m] & v_y^F[m] \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ a_{xx} \\ a_{xy} \\ a_{yx} \\ a_{yy} \\ b_t \end{bmatrix} \quad (3)$$

$$= \overline{x}^F[m] - F^i[m]\overline{b}$$

Considering a set of N difference measures $\Delta x_i^c[m], m = 1, \ldots, N$, stored in the block corresponding to temporal interval with overlapped transition, and taking into account the matrix covariances of the estimated differences, $R_i[k]$, the least squares solution results in [40]:

$$\left(\hat{b}\right)_*^i = \left(\sum_{m=1}^{N} (F^i[m])^t (R^i[m])^{-1} (F^i[m])\right)^{-1} \left(\sum_{m=1}^{N} (F^i[m])^t (R^i)^{-1} \overline{x}^F[m]\right) \quad (4)$$

The resulting corrections are applied to local tracks for each surveillance-sensor agent S_i, and then sent back (through the $SendFeedbackInfo()$ intention) to them so they are included in their corresponding *projection* capability applied to future images. The message $SendFeedbackInfo()$ is sent from the local fusion agent to surveillance-sensor agents, containing the computed vector $\hat{b}_*^i$

4.3 Track fusion between consistent tracks

Once consistent tracks are selected, the data fusion is performed according to the reliability of each track. We take a simple federated fusion approach [34], based on weighting each source of information according to the covariance error matrix, modified with an additional score function assessing the level of confidence assigned to the tracking process [41]. For each j-th object being tracked in the coalition, combination is given by:

$$\left(R_j^F\right)^{-1} = \sum_{k \in C} \left(\alpha_j^i R_j^i\right)^{-1}; \ \hat{x}_j^F = R_j^F \sum_{k \in C} \left(\alpha_j^i R_j^i\right)^{-1} \hat{x}_j^i \quad (5)$$

The level of confidence for each consistent camera and for each common target is based on the inverse covariance value of each sensor and target multiplied by the heuristic score function α_j^i. The score function $\alpha_j^i \in [1, \inf)$ is an scalar which characterizes the performance of the camera of k-th Surveillance sensor based on a combination of image tracking performance metrics (combination of color, spatial regularity, shape uniformity, motion stability, etc.)

5 Experiments

In [13], authors describe how the use of CS-MAS allows more robust and decentralized system to be designed, where management is distributed between the different surveillance-sensor agents. The coordination among surveillance-sensor agents is proved and justified in order to achieve together a surveillance task. This surveillance system designed is a prototype for a distributed surveillance system at the university campus, deployed both in outdoor and indoor areas.

The illustrative scenario analyzed in this work (depicted in Fig. 5) is an outdoor scene in which video cameras with overlapped fields of view cover the pedestrians walking along a footpath. Both surveillance-sensor agents and a local fusion agent establish a coalition in order to track the same object. In the shared area, the agents are simultaneously tracking the object, which is used for aligning time-space coordinates and fusing their local tracks during the coalition maintenance by the local fusion agent.

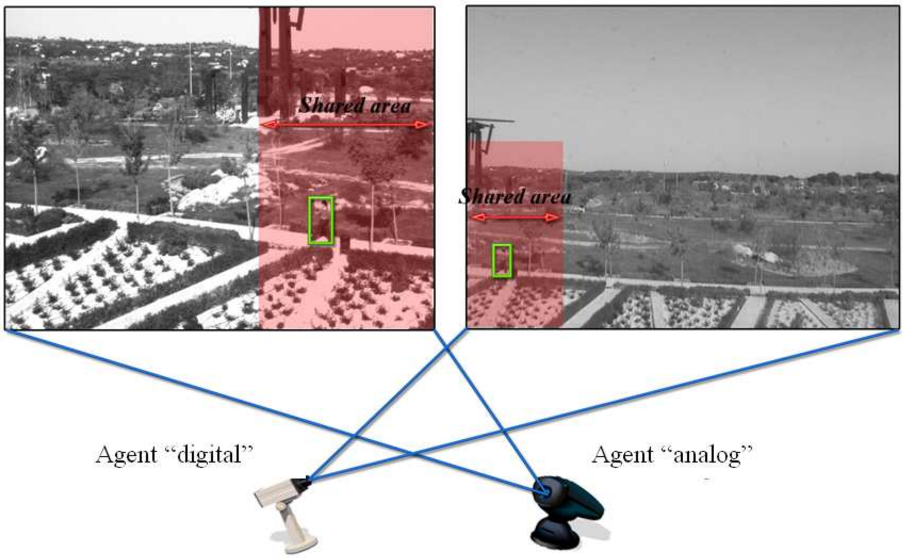

Fig. 5. Scenario of coalition for tracking, with an overlapping zone

The overlapped regions are marked in Fig. 6, and the reference ground-truth lines to identify the footpath are shown in Fig. 7.

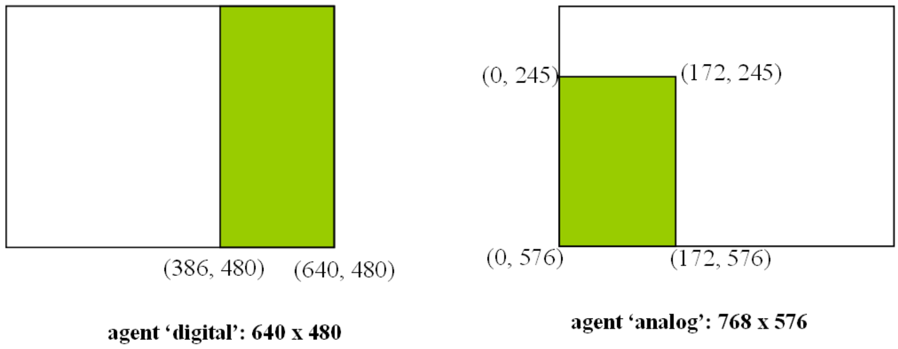

Fig. 6. Sizes and overlapped area of images provided by both agents of the coalition

Ground truth in agent "analog" Ground truth in agent "digital"

Fig. 7. Ground truth lines in both agents of the coalition

This illustrative configuration was enough to run experiments in which surveillance sensor agents provide different views of a common scenario (the footpath), and their local tracks can be fused in a global representation.

The two surveillance-sensor agents are monitoring the same scene and sends tracks information by applying the $SendTargetInfo()$ intention. These two surveillance-sensor agents, referred from now on as "analog-agent" and "digital-agent", use different acquisition technologies. The first camera agent it is based on analogical technology, connected to a frame grabber card, and the second one uses digital firewire camera technology, whose main features are indicated in Table 1. When the fusion center receives the information of the same target from the two surveillance-sensor agents it start the coalition process applying the $MakeCoalition()$ intention. Then, it sends $Call{-}for{-}Coalition$ messages for establish the coalition. This is an example of the $Call - for - Coalition$ message content in the FIPA Semantic Language (SL):

```
(cfp
  :sender (agent-identifier :name fusion-center)
  :receiver (set (agent-identifier :name analog-agent))
  :content
```

```
    "((action (agent-identifier :name analog-agent)
      (MakeCoalition local-fusion-agent-1 ObjectId-1 Tracking))
     )"
  :language fipa-sl)
```

Then each of the agents asked for the coalition answer to the fusion center with an $Accept - Coalition$ message:

```
(accept-proposal
  :sender (agent-identifier :name analog-agent)
  :receiver (set (agent-identifier :name fusion-center))
  :in-reply-to MakeCoalition
  :content
    "((action (agent-identifier :name fusion-center)
      (MakeCoalition local-fusion-agent-1 ObjectId-1 Tracking))
     )"
  :language fipa-sl)
```

Now we have a specific coalition $\Psi_1 = \{analog - agent,\ digital - agent\}$ for the common objective of tracking the same object. At this moment, a local fusion agent is created dynamically by the fusion center by performing the $CreateFusionAgent$ $(\Psi_1,\ ObjectId1,\ FusionCapability)$ intention. At this stage tracks information are send it to the local fusion agent (F_1) by the $Inform{-}Coalition$ message triggered by the $SendTargetInfo$ $(F_1,\ ObjectId{-}1,\ Tracking)$ intention, as we shown in this example:

```
 (inform
  :sender (agent-identifier :name analog-agent)
  :receiver (set (agent-identifier :name local-fusion-agent-1))
  :content
    (ObjectId-1)
  )
```

The content of the previous message $(ObjectId{-}1)$ is an instance of a java class which contains the numeric description of the location, velocity, dimensions and the associated error covariance matrix.

Previously to this experimentation, an off-line calibration process was performed over each scene. Therefore, each surveillance-sensor agent is assumed to measure the location of a moving target within its field of view with respect to a common reference system, this was done by the $Projection()$ intention. We have chosen the GPS (Global Position System) reference to represent the objects' location, using a portable equipment to take the measurements (GarminTM GPS-18 USB). Due to the calibration process, the correspondences between 2D image coordinates (pixels) and their respective GPS world position can be set up.

The overlapped area allows the two surveillance-sensor agents to track the targets simultaneously. Once the right agent has detected a pedestrian, it calculates its size, location and velocity. Based on these data from the overlapped area the delivered tracks may be used by local fusion agent to align and correct the tracks of the other side of the coalition.

We have analyzed ten videos of pedestrians walking at different speeds from right to left through both scenes. In Fig. 8 we can see the pedestrian's tracked positions, expressed in local image coordinates (without the *projection* intention), for the first recorded video. Every point is within of the calibrated region described by the calibration markers, at the two sides of the footpath (asterisks).

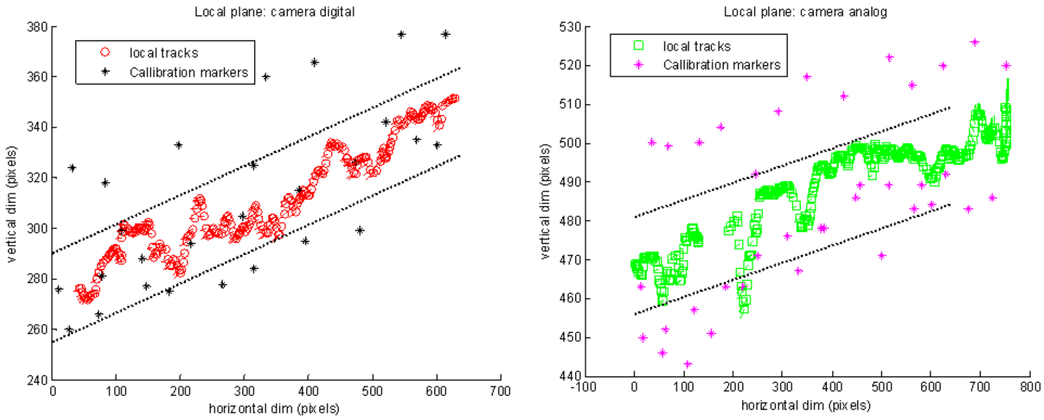

Fig. 8. Local tracking with both agents of the coalition

After the surveillance-sensor agent projection (*Projection* intention), we were able to map the image coordinates toward global coordinates. The results of this transformation for both tracks are depicted in Fig. 9, in the geodetic coordinates of GPS after calibration: latitude, longitude, altitude. In fact, they are expressed as a relative shift, in thousandths of minute, over a reference point at $North$ 40 32 min, $West$ 4 0 min. Then, a direct projection from geodetic to Cartesian coordinates is carried out, using the stereographic transformation with the reference mentioned above used as tangent point ($coordinate$ 0,0). A detail is depicted in Fig. 10, where the initialization of track from the digital-agent with noisy velocity can be appreciated, compared with the track coming from the analog-agent.

The fused output, carried out by the local fusion agent after alignment is depicted in Fig. 11, where the transition is more smooth than a direct switch between both tracks. Besides, the alignment correction fixed to tracks from digital agent for the rest of time, allowed achieving a more coherent fused track. The tracks of all videos after fusion are depicted in Fig. 12.

When the track is lost by the analog-agent, the $LeaveCoalition$ $(analog - agent,\ ObjectId - 1,\ Tracking)$ intention is triggered, and it causes the

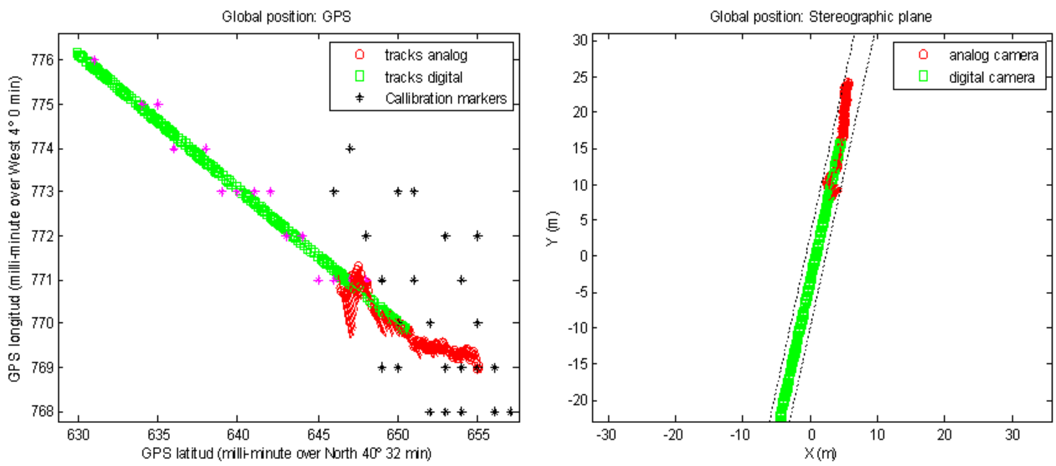

Fig. 9. Projected tracks in global coordinates: GPS and stereographic plane

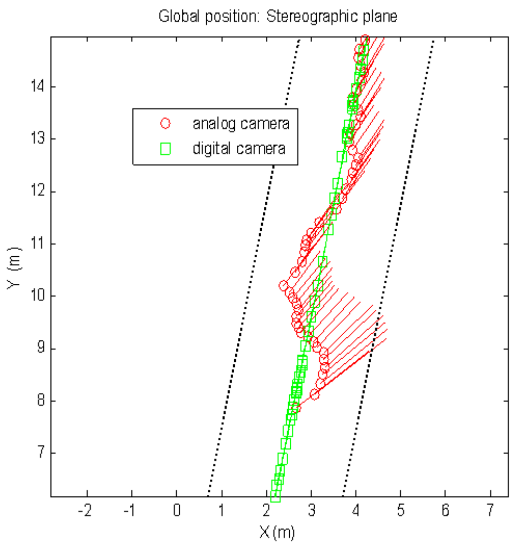

Fig. 10. Projected tracks in the Cartesian stereographic plane (detail)

$Cancel - Coalition$ message, as for example:

```
(cancel
 :sender (agent-identifier :name analog-agent)
 :receiver (set (agent-identifier :name local-fusion-agent-1))
 :content
    "((action (agent-identifier :name local-fusion-agent-1)
     (MakeCoalition local-fusion-agent1 ObjectId-1 Tracking)))"
 :langage fipa-sl
)
```

The previous process is the same for the digital agent.

The last message send it by the local-fusion-agent-1 before ends its execution

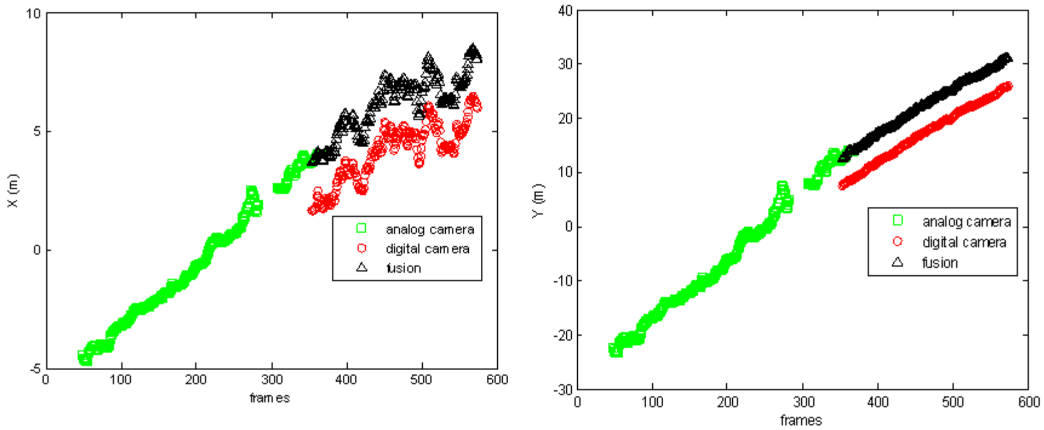

Fig. 11. Position: X, Y coordinates of original tracks and fusion result

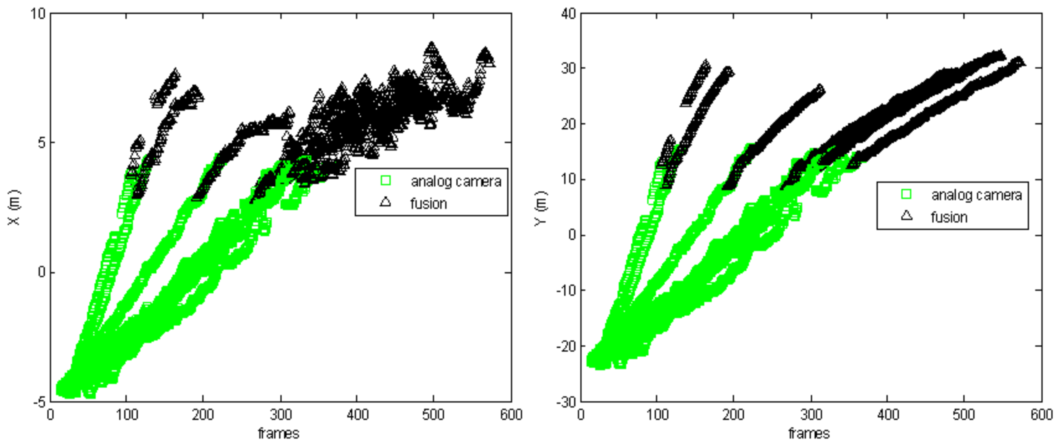

Fig. 12. Position: X, Y coordinates of fused tracks for all videos

is the $Inform - Broken - Coalition$ message, as for example:

```
(inform
:sender (agent-identifier :name local-fusion-agent-1)
:receiver (set (agent-identifier :name fusion-center))
:content
  (BrokenCoalition analog-agent digital-agent)
)
```

6 Conclusions and Future Work

Multi-agent coordination enhances the continuous and accurate tracking of objects of interest within the area covered by a visual sensor network. In this

Table 1
Cameras used in the experimentation

	Output resolution	Frames per second
Sony EVI-D100P + Data Translation 3130 Frame Grabber	768×576	24
Basler A312fc/c + Matrox Meteor2 - 1394	640×480	30

paper we present the details of the CS-MAS architecture which makes possible a global tracking in a visual sensor network. The cooperation between the agents is achieved through a coalition formation process. The coalitions of agents are created dynamically when they are necessary. The main goal is to improve the knowledge inferred from the information captured in different surveillance-sensor agents, extending surveillance functionalities through an effective management of a network interdependences to carry out the tasks. In this work, the specific process of data fusion is detailed in a higher-level layer which dynamically creates dynamic temporal coalitions associated to specific capabilities. In the experiments presented, the agents cooperate in groups with the objective of maintain the trajectories associated to specific targets moving within the guarded areas.

The data fusion algorithms performed at this logical layer are able to represent local views provided from each camera in a common time-space framework, with an accurate integration of local processes performed at different nodes. The experiments showed a continuous and seamless tracking across the transition between cameras FoVs. Time-space alignment of different data sources is performed with an on-line registration process for accurate fusion of partial outputs. The improved performance was illustrated in the cooperative tracking of two cameras with different technologies, sharing a overlapped field of view in a campus outdoor surveillance system. As an ongoing work we are considering to compare the surveillance process with other data fusion strategies.

References

[1] M E. Weber and M L. Stone. Low altitude wind shear detection using airport surveillance radars. In *Radar Conference*, pages 52–57. Record of the 1994 IEEE

National, Mar 1994.

[2] A. Pozzobon, G. Sciutto, and V. Recagno. *Security in ports: the user requirements for surveillance system.* Advanced Video-Based Surveillance Systems. Kluwer Academic, Boston, Mass, USA, 1998.

[3] P. Avis. Surveillance and canadian maritime domestic security. *Canadian Military Journal*, 4(1):9–15, 2003.

[4] B P L. Lo, J. Sun, and S A. Velastin. Fusing visual and audio information in a distributed intelligent surveillance system for public transport systems. *Acta Automatica Sinica*, 29(3):393–407, 2003.

[5] C. Nwagboso. *User focused surveillance systems integration for intelligent transport systems*, pages 8–12. Advanced Video-Based Surveillance Systems. Kluwer Academic, Boston, Mass, USA, 1998.

[6] ADVISOR. Advisor specification documents (internal classification 2001).

[7] CROMATICA. http://dilnxsvr.king.ac.uk/cromatica/.

[8] N. Ronetti and C. Dambra. *Railway station surveillance: the Italian case*, pages 13–20. Multimedia Video Based Surveillance Systems. Kluwer Academic, Boston, Mass, USA, 2000.

[9] DETECT. http://www.detec.no.

[10] GOTCHA. http://www.gotchanow.com.

[11] M. Pellegrini and P. Tonani. *Highway traffic monitoring*. Advanced Video-Based Surveillance Systems. Kluwer Academic, Boston, Mass, USA, 1998.

[12] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real-time computer vision system for measuring traffic parameters. In *Computer Vision and Pattern Recognition*, pages 495–201. IEEE Computer Society Conference, Jun 1997.

[13] M A. Patricio, J. Carbo, O. Perez, J. Garcia, and J M. Molina. Multi-agent framework in visual sensor networks. *EURASIP Journal on Advances in Signal Processing*, 2007:Article ID 98639, 21 pages, 2007. doi:10.1155/2007/98639.

[14] R T. Collins, A.J. Lipton, and T. Kanade. A system for video surveillance and monitoring, 2000. Tech. Rep. CMU-RI-TR-00-12, Robotics Institute, CarnegieMellon University, Pittsburgh, Pa, USA.

[15] C S. Regazzoni, V. Ramesh, and G L. Foresti. Special issue on video communications, processing, and understanding for third generation surveillance systems. In *Proceedings of the IEEE*, volume 89, pages 1355–1539, 2001.

[16] R T. Collins, A J. Lipton, H. Fujiyoshi, and T. Kanade. Algorithms for cooperative multisensor surveillance. In *Proceedings of the IEEE*, volume 89, pages 1456–1477. IEEE, October 2001.

[17] Xiaojing Yuan, Zehang Sun, Yaakov Varol, and George Bebis. A distributed visual surveillance system. In *AVSS '03: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, page 199, Washington, DC, USA, 2003. IEEE Computer Society.

[18] M. Wooldridge and N.R. Jennings. Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.

[19] M.E. Bratman. *Intentions, Plans and Practical Reasoning*. Harvard University Press, Cambridge, Massachusetts, 1987.

[20] F. Castanedo, M. A. Patricio, J. García, and J. M. Molina. Extending surveillance systems capabilities using bdi cooperative sensor agents. In *VSSN '06: Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pages 131–138, New York, NY, USA, 2006. ACM Press.

[21] J.P. Kahan and A. Rapoport. *Theories of Coalition Formation*. Lawrence Erlbaum Associatates Publishers, 1984.

[22] S. Ketchpel. Forming coalitions in the face of uncertain rewards. In *Proceedings of the National Conferece on Artificial Intelligence*, pages 414–419, Seattle, WA, July 1994.

[23] H. Raiffa. *The Art and Science of Negotiation*. Harvard Univ. Press, Cambridge, Massachusetts, 1984.

[24] O. Shehory and S.Kraus. Feasible formation of stable coalitions among autonomous agents in general environments. *Computational Intelligence Journal*, 1995.

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

[26] A. Rao and M. Georgeff. Bdi agents: from theory to practice. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS'95)*, pages 312–319, Cambridge, MA, USA, 1995. The MIT Press.

[27] M. D'Inverno, M. Luck, M. Georgeff, D. Kinny, and M. Wooldridge. The dMARS Architecture: A Specification of the Distributed Multi-Agent Reasoning System. *Autonomous Agents and Multi-Agent Systems*, 9(1):5–53, 2004.

[28] A. Pokahr, L. Braubach, and W. Lamersdorf. A goal deliberation strategy for bdi agent systems. In *Proceedings of the third German conference on Multi-Agent System Technologies (MATES-2005)*. Springer-Verlang. Berlin Heidelberg New York, 2005.

[29] Fabio Luigi Bellifemine, Giovanni Caire, and Dominic Greenwood. *Developing Multi-Agent Systems with JADE (Wiley Series in Agent Technology)*. John Wiley & Sons, 2007.

[30] Micheal Wooldridge. *Reasoning about Rational Agents*. The MIT Press, Cambridge, Massachusetts, 2000.

[31] http://www.fipa.org/specs/fipa00037/sc00037j.html.

[32] J.A. Besada, J. Garcia, J. Portillo, J.M. Molina, A. Varona, and G. Gonzalez. Airport surface surveillance based on video images. *Aerospace and Electronic Systems, IEEE Transactions on*, 41(3):1075–1082, July 2005.

[33] Óscar Pérez, Miguel Ángel Patricio Guisado, Jesús García, and José M. Molina. Improving the segmentation stage of a pedestrian tracking video-based system by means of evolution strategies. In *EvoWorkshops*, pages 438–449, 2006.

[34] D L. Hall and J. Llinas. *Handbook of MultiSensor Data Fusion.* CRC Press, Boca Raton, 2001.

[35] N. Nabaa and R H. Bishop. Solution to a multisensor tracking problem with sensor registration errors. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):354–363, January 1999.

[36] Y. Zhou, H. Leung, and K. Chan. A two-step extended kalman filter fusion approach for misaligned sensors. In *Proceedings of the FUSION'98 International Conference*, pages 54–58, Las Vegas, July 1998.

[37] J A. Besada, J. García, and G. Miguel. A new approach to on-line optimal estimation of multisensor biases. *IEE Proceedings. Radar, Sonar and Navigation*, 151(1), Februry 2004.

[38] J. García, J A. Besada, F J. Jiménez, and J R. Casar. A block processing method for on-line multisensor registration with its application to airport surface data fusion. In *Proceedings of the IEEE International Radar Conference*, Washington-DC, EE. UU., May 2000.

[39] H. Karniely and H T. Siegelmann. Sensor registration using neural networks. *IEEE Transaction on Aerospace and Electronic Systems*, 36(1):85–101, January 2000.

[40] J.G. Herrero, J.A.B. Portas, and J.R.C. Corredera. On-line multi-sensor registration for data fusion on airport surface. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(1):356–370, 2007.

[41] GL Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis. Active video-based surveillance system: the low-level image and video processing techniques needed for implementation. *Signal Processing Magazine, IEEE*, 22(2):25–37, 2005.