Towards a Semi-automatic Situation Diagnosis System in Surveillance Tasks

José Mira[1], Rafael Martínez[1], Mariano Rincón[1], Margarita Bachiller[1], and Antonio Fernández-Caballero[2,⋆]

[1] E.T.S.I. Informática - Univ. Nacional de Educación a Distancia, Madrid, Spain
{jmira,rmtomas,mrincon,marga}@dia.uned.es
[2] Escuela Politécnica Superior de Albacete & Instituto de Investigación en Informática de Albacete, Universidad Castilla-La Mancha, Albacete, Spain
caballer@dsi.uclm.es

Abstract. This paper describes an ongoing project that develops a set of generic components to help humans (semi-automatic system) in surveillance and security tasks in several scenarios. These components are based in the computational model of a set of selective and Active VISual Attention mechanisms with learning capacity (*AVISA*) and in the superposition of an "intelligence" layer that incorporates the knowledge of human experts in security tasks. The project described integrates the responses of these alert mechanisms in the synthesis of the three basic subtasks present in any surveillance and security activity: real-time monitoring, situation diagnosing, and action planning and control. In order to augment the diversity of environments and situations where *AVISA* system may be used, as well as its efficiency as support to surveillance tasks, knowledge components derived from situating cameras on mobile platforms are also developed.

1 Introduction

Surveillance is a multidisciplinary task affecting an increasing number of scenarios, services and customers. It aims to detect threats by continually observing large and vulnerable areas of a scenario considered to be of economic, social or strategic value because it can suffer theft, fire, vandalism or attacks. The range of scenarios is very wide and of very different complexity, going from the mere detection of movement that sets off an alarm to an integral control system that monitors the scene with different sensors, diagnoses the situation and plans a series of consistent actions. In any case, it always implies the observation of mobile objects (people, vehicles, etc.) in a predetermined environment to provide a description of their actions and interactions. Hence, this implies the detection of moving objects and their tracking, the recognition of objects, the analysis of the

⋆ The contribution to this paper from the other members of the AVISA project (A. Delgado, E. Carmona, J.R. Álvarez, F. de la Paz, M.T. López and M.A. Fernández) has been equal to the other authors. All of them should appear on the author list, but the rules of the congress stipulate a maximum of five recognized authors.

J. Mira and J.R. Álvarez (Eds.): IWINAC 2007, Part II, LNCS 4528, pp. 90–98, 2007.

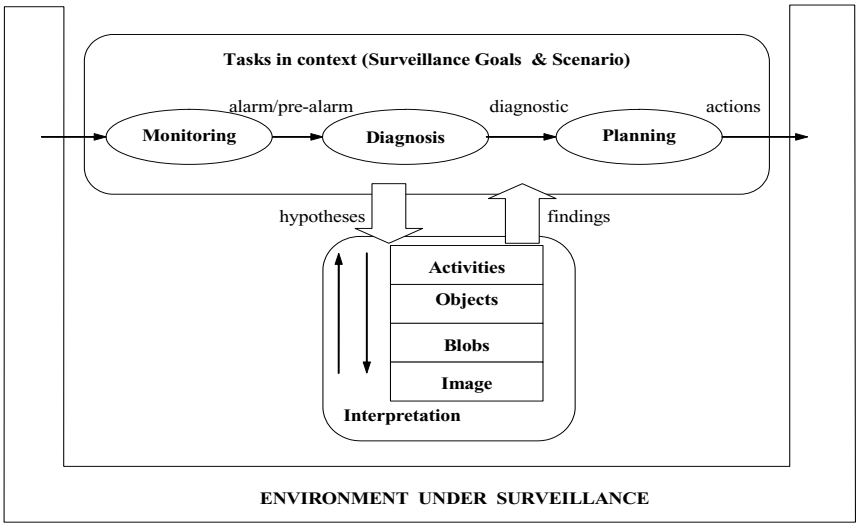

Fig. 1. Graphical Description of the Surveillance Task

specific movement for each type of object and the interpretation of the activity. Most approximations for activity analysis activity [4] [6] [15] [18] are posed as a classification problem, i.e., they are based on defining models for each type of activity in a specific application domain, and they are highly dependent on the results obtained during tracking. As you may observe in Fig. 1 the surveillance task follows a common structure with general control tasks: (1) monitoring of a number of critical variables, whose deviation from normality is a sign of some disfunction, (2) diagnosis of the problem, and (3) planning of relevant actions for solving the problem.

2 The *AVISA* Approach to Surveillance

In our surveillance approach a semi-automatic approximation is considered; at the end it is the operator of an alarms power station who takes the decisions. However, one of the fundamental problems in this kind of projects is the enormous semantic gap between the physical signal level and the knowledge level. One of the ways of overcoming this gap is by partitioning it. There is consensus in the area, although with some variety and dispersion in the nomenclature, in accepting different description levels with an increasing degree of semantics [13] [14], which facilitates the injection of domain knowledge. In our work, four description levels are considered: image, blobs, objects and activities/behaviours (see again Fig. 1). Each of these levels is modular and independent, and the information handled comes from the ontology of the proper level and from the lower level.

In the blob level, the entities are associated with the visible part of objects of interest or with parts of them, the blobs. In the object level the information

associated with blobs for producing a description of the objects of interest on the scene is reorganised. We move from a frame-oriented description to an object-oriented description. The models of the objects of interest are described here, which contain: 1) the visual characterisation of the object and its spatial-temporal evolution; 2) the composition relations used to describe complex objects; and 3) the relations between objects for generating the geometric description of the task-focused scene. At object level, it is necessary to inject additional domain knowledge, in terms of models of the objects in any description level where they are used. At activity level, the identification of a set pattern of elementary activities (events), each of them basically normal, can be interpreted as an alarming activity together.

The aims of the *"selective and Active VISual Attention mechanisms with learning capacity"* (*AVISA*) project correspond to modelling and formalization of the tasks associated to each level and their integration. This purpose is divided in *AVISA* two subprojects: (a) "Distributed monitoring of scenarios with different kinds of moving non-rigid objects", with focus on the processes associated to pixels, blobs and static and dynamic objects considered for monitoring the scene. The monitoring output is a warning in the sense that *AVISA* system has detected some anomaly. Thus it is the alarm that activates a diagnosis process. (b) "A set of diagnosis, planning and control agents with capacity of learning and cooperation with humans", which shares the object level but focuses on the situation diagnosis task, which starts from the results of the monitoring task. The following sections introduce the tasks related to *AVISA* project in detail.

3 Monitoring and Pre-diagnosis

The aim of the visual monitoring and scene pre-diagnosis tasks is to collect visual information from diverse coordinates of the scene and to pre-process them according to the selective visual attention (*SVA)* mechanisms: *Accumulative computation* and *algorithmic lateral inhibition* (*ALI*). The global purpose is to focus attention of the cameras (static and mobile) on those elements of the scene that better fulfil the criteria specified by the human expert as "objects of interest". The processes corresponding to these tasks are related to the image, blob and object levels. The decision on the proper monitoring message is taken in terms of the value of a set of parameters of motion, size and shape, firstly obtaining the blobs, and then injecting the semantic needed to get the objects of interest. It has been possible to work on a real outdoor scenario monitored by the cooperating company *SECISA* (see image at Fig. 2). This process has been developed in a sequential and incremental manner, starting with *AC* and *ALI* [9], formulated firstly at physical level and then in terms of multi-agent systems, where a totalizing process on the individual opinions of a set of working memories as agent coordination mechanism has been used. Next both mechanism have been combined [5] to face the dynamic *SVA* problem [7], that is to say, "where to look to" and "what to look to". Our strategy has been the integration of the bottom-up (connectionist) and top-down (symbolic) organizations usually accepted in Neuroscience.

Fig. 2. Outdoor scenario. Monitoring of a test case. SECISA video sequence.

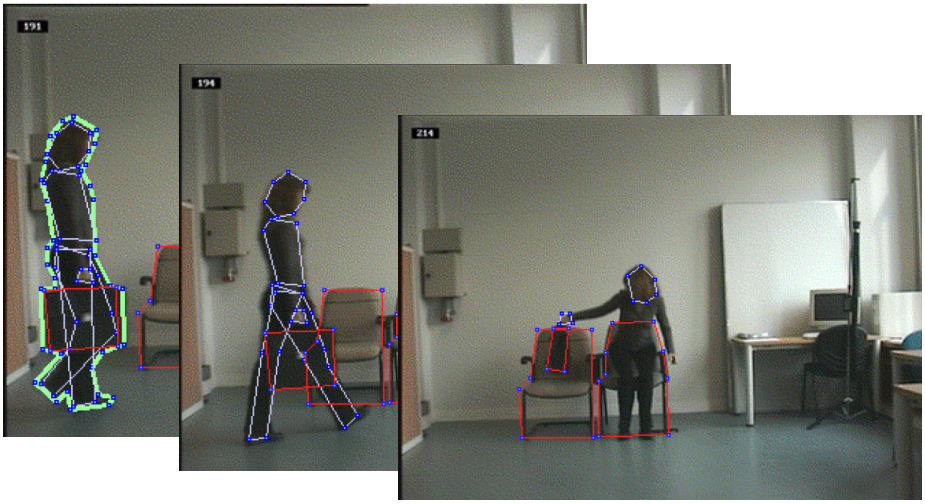

Fig. 3. Use of an annotation tool for video sequences (the indoor scenario in the figure), which allows the representation of time, position, attributes, relations, states, and events as visual objects

To exemplify the methodological proposal developed, initially a very simple scenario was defined (Fig. 3). It is an indoor space that is a pass-through area for humans. Humans can move freely, come in and go out of the observation area, sit down, carry a briefcase, leave it, pick it up, etc. This scenario simplifies the problem (there are no dogs, trees or cars, etc.) and it helps us specify the other two components of a context, the aims and sensors. An alarm would go off when someone leaves a briefcase (a package) in the area under surveillance.

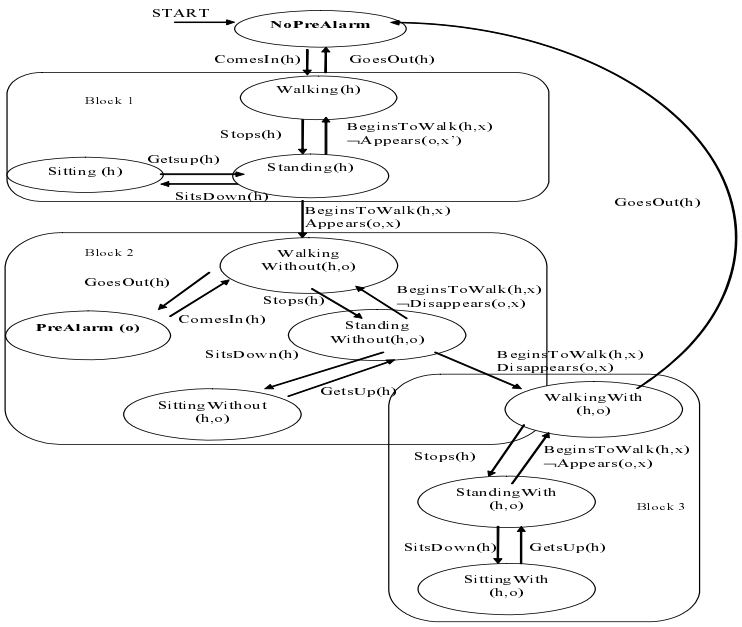

Fig. 4. State transition diagram used to model the activity level in the monitoring and pre-diagnosis steps according with the reconocibles events in the video sequence

If the person returns to his initial position and picks up the briefcase, the alarm will stop. This is a situation that forms part of the most thorough surveillance of a situation, with no going back, where it is identified that a person who leaves a larger site (an airport, hospital, etc.), first causing an alarm of this sort because he has left a package, which would require a response action. Fig. 4 shows the state transition diagram for the automaton describing the scene at activity leve in the context of the monitoring task.

The main difficulties found are related to the interaction between moving objects (e.g., a human) and static ones (for instance, a suitcase) and to the complexity of working in real-time in real scenarios (camera movement, noise, 2D vs. 3D, calibration, stereovision, and so on).

4 Diagnosis of the Activity Level

The aim of the diagnosis task is to model the knowledge of the human expert in surveillance tasks. The main problem faced in this task is that, when interpreting a scene the external knowledge necessary to fill the enormous semantic gap between the physical signal level and the knowledge level has to be injected, i.e., the AVISA "intelligence" level needs to adapt the sensory information to the abstraction level.

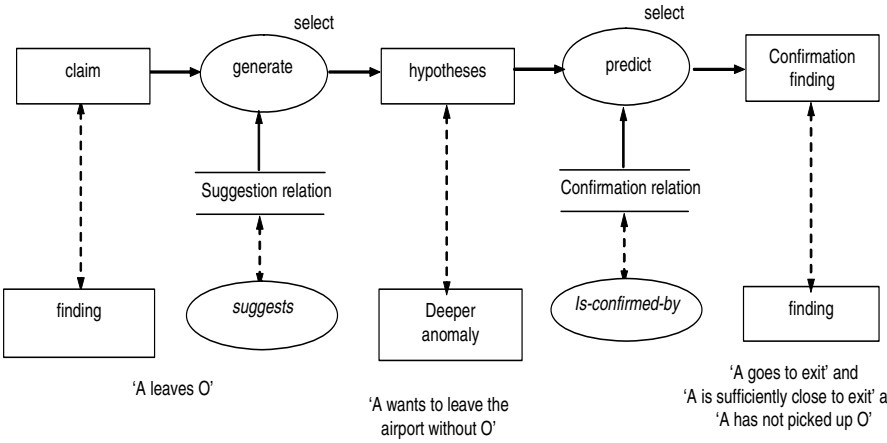

Fig. 5. The "generate and predict" inferential scheme. A first inference "generate" selects the hypotheses from a claim (from monitoring) requiring confirmation. This inference is based on the "suggestion" relation between anomalies. From this hypothesis, the inference "predict" selects the confirmation findings. In this instance the inference is based on the confirmation relation between the anomaly of a specific depth level requiring confirmation and the relevant findings.

If the monitoring step focuses on the bottom-up interpretation, the logical continuation would be to include a "diagnosis" stage. The use of models is a way of introducing domain knowledge in a top-down control which makes it possible to improve segmentation by aiding detection of the constituent parts of the object, detecting regions of potential confusion between object and background, and improving tracking with specific methods depending on the object type.

From the firing of the alarm by the monitoring module, there are two possible relations with the diagnosis which determine its meaning in the global process:

a) The alarm triggers the "diagnosis of what has happened" process. In other words, there is an analysis of what has happened from the recorded images (and data from other sensors).

b) The alarm triggers the "diagnosis of what is happening" process. This process is focused and guided by hypotheses of what could be happening. The diagnosis ends with confirmation of hypotheses that require a response that the planning phase will determine, although it could also conclude that the situation is not in fact alarming.

Figure 5 shows the inferential schema of the problem solving method (PSM) "generate-predict" used to solve the diagnosis task in AVISA. Certain behaviors suggest individual causal relations and that those relations have visual consequences that can be verified. So, for example, the event "A leaves the object O" suggests two hypotheses: "A picks up O" and "A wants to leave the airport without O". The confirmation of this last one demands the obtaining of the findings: "A goes to Exit", "A has not picked up O" and "A is sufficiently close

to Exit". They are findings that directly do not emerge from the interpretation scene, but that there is to look for them.

In the following table, the generetad hypothesis is associated to the confirmation finding and this one is associated to its decomposition in more elementary events and their space-temporary relation.

Claim	Hypothesis	Confirmation finding Decomposition (looking for)
Leaves(A, O, t_1, x)	Takes(A, O, t_2, x)	Picks(A, O, t_2, x)
	Wants-to-leave-airport (A, O, t_3)	Is-going-to(A, Exit, t_3) $\neg$Picks-up(A, O, t_4) $\forall t_4 \in [t_1, t_3]$ Is-close-to(A, Exit, t_4)

5 Learning

The objectives of learning task are (1) to guide the search process in the *SVA* mechanisms by reinforcing the descriptions of the objects that look more like the predefined patterns, (2) to guide the robots in their pathways towards the selected coordinates, and, (3) to accumulate the *AVISA* experience in its collaboration with the human operator. Of these three functions the first one has been fully accomplished [11], with the complementation of reinforcement learning with other evolutionary models. The second function is in development phase; to date an autonomous navigation system based on the calculation of the centre of area in open space is available [1]. The third function, associated to the interaction of the prototype with the human operator, is currently being developed.

Fig. 6. Use of learning by reinforcement to select those parameters that contribute to a better discrimination (diagnosis) of different situations (video sequences) in accordance with the intention of the observer

The main problems of the learning-related tasks are given by the difficulty of constructing valid training sets for supervised learning in different scenarios. The solutions explored so far are the construction of a case base for each concrete scenario and the use of evolutionary procedures to enhance the segmentation task. Those parameters that contribute to a better discrimination (diagnosis) of different situations in accordance with the intention of the observer have to be reinforced [11].

6 Conclusions

This paper has presented an ongoing project denominated *AVISA* that develops a set of generic components to help humans in surveillance and security tasks in several scenarios. The paper has introduced the tasks related to the project, highlighting the effort taken, the difficulties found and the main contributions offered so far.

The greatest difficulty found arises when connecting blobs to activities in different real scenarios. To solve this difficulty, at least partially, the following measures have been adopted: (1) Modelling and labelling of all families of sequences of interest in the proposed real scenarios and at the levels of objects, events and behaviours (construction of the ontologies of each level for each scenario). (2) Establishment of a clear frontier between automatic and human interpretation for each context (scenario, surveillance intention, information sources, robot accessibility, and so on). (3) Emphasizing in alarm pre-diagnosis as a complement to monitoring, leaving the decision on the action at each situation to the human operator. (4) Dedicating an additional effort to the construction of a case base that enables reasoning at each scenario by analogy to previous situations, in cooperation with any learning-related task. (5) Exploring new ways of representing activities that enable limiting the combinatory explosion of the state transition diagrams of the deterministic finite automata. (6) Making explicit the difficulty of the situation diagnosis task and trying to integrate the efforts of this project with some complimentary ones (face recognition, etc.).

In spite of the difficulties found, some preliminary results has been obtained: (1) Modelling and implementation of the motion detection task [12]. (2) Modelling of the image understanding process [17]. (3) Generation of a tool for annotating images [16] and its later use in learning.

Acknowledgements

This work is supported in part by the Spanish coordinated TIN2004-07661-C02-01 and TIN2004-07661-C02-02 grants.

References

1. Álvarez J.R., De la Paz F. & Mira J. (2005). A robotics inspired method of modeling accessible open space to help blind people in the orientation and traveling tasks. Lecture Notes in Computer Science, 3561, pp. 405-415.

2. Bolles B. & Nevatia R. (2004). A hierarchical video event ontology in OWL. Final Report 2004, ARDA Project.
3. Bremond F., Maillot N., Thonnat M. & Vu V.T. (2004). Ontologies for video events. INRIA, Research Report, 5189.
4. Collins R.T., Lipton A.J. & Kanade T. (eds.) (2000). Special Issue on Video Surveillance, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:8.
5. Fernández-Caballero A., López M.T., Mira J., Delgado A.E., López-Valles J.M. & Fernández M.A. (2007). Modelling the stereovision-correspondence-analysis task by lateral inhibition in accumulative computation problem-solving method. Expert Systems with Applications, 34 (4).
6. Haritaoglu I., Harwood D. & Davis L.S. (2000). W4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), pp. 809-830.
7. López M.T., Fernández M.A., Fernández-Caballero A., Mira J. & Delgado A.E. (2007). Dynamic visual attention model in image sequences. Image and Vision Computing, 25 (5), pp. 597-613.
8. López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2007). Real-time motion detection from ALI model. Eleventh International Conference on Computer Aided Systems Theory, EUROCAST 2007.
9. López M.T., Fernández-Caballero A., Mira J., Delgado A.E. & Fernández M.A. (2006). Algorithmic lateral inhibition method in dynamic and selective visual attention task: Application to moving objects detection and labelling. Expert Systems with Applications, 31 (3), pp. 570-594.
10. López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2006). Visual surveillance by dynamic visual attention method. Pattern Recognition, 39 (11), pp. 2194-2211.
11. López M.T., Fernández-Caballero A., Fernández M.A., Mira J. & Delgado A.E. (2006). Motion features to enhance scene segmentation in active visual attention. Pattern Recognition Letters, 27 (5), pp. 469-478.
12. Mira J., Delgado A.E., Fernández-Caballero A. & Fernández M.A. (2004). Knowledge modelling for the motion detection task: The algorithmic lateral inhibition method. Expert Systems with Applications, 27 (2), pp. 169-185.
13. Nagel H.H. (2004). Steps towards a cognitive vision system. AI Magazine, 25 (2), pp. 31-50.
14. Neuman B. & Weiss T. (2003). Navigating through logic-based scene models for high-level scene interpretations. 3rd International Conference on Computer Vision Systems, ICVS-2003. Lecture Notes in Computer Science, 2626, pp. 212-222.
15. Regazzoni, C.S., Fabri, G. & Breñaza, G. (eds.) (1999). Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht.
16. Rincón M. & Martínez-Cantos J. (2007). An annotation tool for video understanding. Eleventh International Conference on Computer Aided Systems Theory, EUROCAST 2007, 12-16 February 2007, Las Palmas de Gran Canaria (Spain).
17. Rincón M., Bachiller M. & Mira J. (2005). Knowledge modeling for the image understanding task as a design task. Expert Systems with Applications, 29 (1), pp. 207-217.
18. Robertson N. & Reid I. (2006). A general method for human activity recognition in video. Computer Vision and Image Understanding, 104, pp. 232-248.
19. Taboada M., Des J., Martínez D. & Mira J. (2005). Aligning reference terminologies and knowledge bases in the health care domain. Lecture Notes in Computer Science, 3561, pp. 437-446.