Proposal for an INRIA Project-Team
PULSAR
Perception Understanding and Learning Systems for Activity Recognition

30 August 2007

Theme: CogC

Multimedia data: interpretation and man-machine interaction.

Abstract

This research project-team will be focused on Activity Recognition. More precisely we are interested in the **real-time semantic interpretation of dynamic scenes** observed by sensors. We will thus study long-term spatio-temporal activities performed by human beings, animals or vehicles in the physical world. The major issue in semantic interpretation of dynamic scenes is the gap between the subjective interpretation of data and the objective measures provided by the sensors. Our approach in order to address this problem is to keep a clear boundary between the application dependent subjective interpretations and the objective general analysis of the videos. Pulsar will propose new techniques in the field of **cognitive vision** and **cognitive systems** for *physical object recognition, activity understanding, activity learning, system design and evaluation* and will focus on two main application domains: **safety/security** and **healthcare**.

Keywords

Activity Recognition, Cognitive Vision, Cognitive Systems, Scene Understanding, Software Architecture, Computer Vision, Knowledge-based Systems, Software Engineering

A pulsar is a rapidly rotating neutron star. The radiation from such a star appears to come in a series of regular pulses (one per revolution), which explains the name.

Contents

1 The Team

The permanent scientific members are INRIA scientists with different background and skills. François Brémond and Monique Thonnat are cognitive vision scientists with a large experience in artificial intelligence and computer vision. Sabine Moisan and Annie Ressouche are computer scientists with respectively a large experience in artificial intelligence software design and in formal techniques for software verification. Guillaume Charpiat is a new junior scientist specialized in computer vision (PhD with O. Faugeras) and statistics (Post-doc with B. Scholkopf) http://www.kyb.mpg.de/∼charpiat/.

Permanent scientific members

- **Team Leader**: Monique Thonnat DR1, HDR
- **Research scientists INRIA**:

  François Brémond CR1, HDR
  Guillaume Charpiat CR2 (from December 2007)
  Sabine Moisan CR1, HDR
  Annie Ressouche CR1

- **Team Assistant**: Catherine Martin (part time) INRIA

Non permanent team members

- **External collaborator**:

  Prof Jean-Paul Rigault UNSA (University Nice Sophia Antipolis)

- **Post-doc**: Sundaram Suresh, PhD
- **PhD Students**:

  Bui Binh (Cifre RATP UNSA) [2nd year]
  Mohamed Becha Kaaniche UNSA [1st year]
  Naoufel Khayati (STIC Tunisie Tunis/UNSA) [3rd year]
  Vincent Martin UNSA [3rd year]
  Anh Tuan Nghiem UNSA [1st year]
  Lan Le Thi (AUF Hanoi/UNSA) [2nd year]
  Nadia Zouba UNSA [1st year]
  Marcos Zúniga (Conicyt/UNSA) [2nd year]

- **Research engineers INRIA**:

  Bernard Boulay, PhD (PACAlab STMicroelectronics)
  Etienne Corvée, PhD (European project CARETAKER)
  Ruihua Ma, PhD (European project SERKET)
  Luis Patino Vilchis, PhD (European project CARETAKER)
  Valery Valentin (System@tic project SIC)

HDR stands for "Habilité à Diriger des Recherches" which is the diploma enabling to supervise PhD theses. CR2 stands for Chargé de Recherches second class (junior scientist), CR1 stands for Chargé de Recherches first class and DR1 stands for "Directeur de Recherches" (senior scientist).

2 Context and Overall Goal

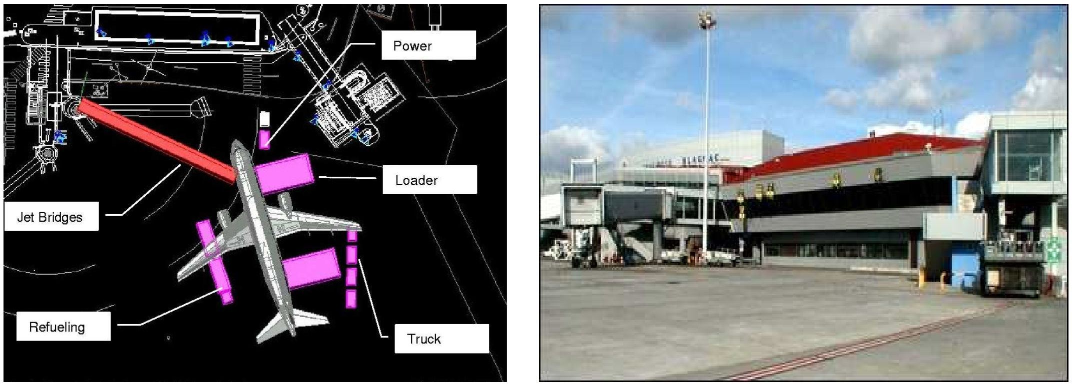
PULSAR (Perception Understanding and Learning Systems for Activity Recognition) is in the continuation of the Orion INRIA project-team in the Cognitive Systems theme "Multimedia data: interpretation and man-machine interaction",CogC in short.

This new research project-team will be focused on Activity Recognition. More precisely we are interested in the **real-time semantic interpretation of dynamic scenes** observed by sensors. We will thus study long-term spatio-temporal activities performed by several interacting human beings, animals or vehicles in the physical world. A typical example of complex activities in which we are interested is shown in figure 1 for aircraft monitoring in apron areas. In this example the duration of the servicing activities around the aircraft is about twenty minutes and the activities involve interactions between several ground vehicles and human operators. The goal is to recognize these activities through formal activity models (as shown in figure 2) and data captured by a network of video cameras (such as the ones shown in figure 3). For more details, refer to the related European project website http://www.avitrack.net/. Activity recognition should not be confused with action recognition. Action recognition seeks to detect short-term sequences of configurations of an individual articulated form (typically human) that are charactristic to specific human actions such as grasping or touching an object. In the computer vision community the term video event is often used as a general term for both a short-term action and a long-term activity.

The major issue in semantic interpretation of dynamic scenes is the gap between the subjective interpretation of data and the objective measures provided by the sensors. The interpretation of a video sequence is not unique; it depends on the a priori knowledge of the observer and on his/her goal. For instance a video showing apron scenes will have different interpretations from a security officer (looking for dangerous situations) or from an airport manager (looking at servicing operations scheduling) point of views.

Our approach to address this problem is to keep a clear boundary between the application dependent subjective interpretations and the objective analysis of the videos. We thus define a set of objective measures which can be extracted in real-time from the videos, we propose formal models to enable users to express their activities of interest and we build matching techniques to bridge the gap between the objective measures and the activity models.

Pulsar will propose new techniques in the field of **cognitive vision** and **cognitive systems** for *physical object recognition, activity understanding and learning, dependable activity recognition system design and evaluation*. We will have a multidisciplinary and a pragmatic approach. Alike Orion, Pulsar will be a combination of research in computer vision, software engineering and knowledge engineering. More precisely we will propose new computer vision techniques for mobile object perception with a focus on real-time algorithms and 4D analysis (e.g. 3D models and long-term tracking). We will deepen our research work on knowledge representation and symbolic reasoning for activity recognition. We will study how statistical techniques and machine learning can in general complement our a priori knowledge techniques for activity recognition. Pulsar will deepen our research work on software engineering for the design and the evaluation of effective and efficient activity recognition systems. For instance, we will continue our work done in ETISEO [217](http://www.etiseo.net) on designing new evaluation techniques taking into account the diversity of video sequences. Pulsar will also take advantage of a pragmatic approach working on concrete problems of activity recognition to propose new cognitive systems techniques inspired by and validated on applications, in a virtuous cycle. Thus Pulsar will provide both theoretical results for activity recognition (e.g. formal models



a b

Figure 1: Activity recognition problem in airport: a) overview of the main servicing operations and of the location of the video cameras (in blue) b) front view of the apron area showing the video cameras

```
CompositeEvent(Unloading_Operation,
  PhysicalObjects( (p1 : Person), (v1 : Vehicle), (v2 : Vehicle), (v3 : Vehicle),
                   (z1 : Zone), (z2 : Zone),(z3 :Zone), (z4 : Zone))
  Components(  (c1 : CompositeEvent Loader_Arrival(v1, z1, z2))
               (c2 : CompositeEvent Transporter_Arrival(v2, z1, z3))
               (c3 : CompositeState  Worker_Manipulating_Container(p1, v3, v2, z3, z4)))
  Constraints(  (v1->SubType = LOADER)
               (v2->SubType = TRANSPORTER)
               (z1->Name = ERA)
               (z2->Name = Front_Loading_Area)
               (z3->Name = Transporter_Area)
               (z4->Name = Container_Worker_Area)
               (c1 before_meet c2)
               (c2 before_meet c3)))
```

Figure 2: Activity recognition problem in airport: example of an activity model enabling to describe an unloading operation with a high-level language

of activity, methodology for activity recognition system design, etc.) and concrete results (e.g. real-time algorithms, performance evaluation metrics, prototypes of integrated activity recognition systems, etc.).

Activity recognition is a hot topic in the academic field not only due to scientific motivations but also due to strong demands coming from the industry and the society; in particular for videosurveillance and homecare. In fact, there is an increasing need to automate the recognition of activities observed by visual sensors (usually CCD cameras, omni directional cameras, infrared cameras), but also by microphones and other sensors (e.g. optical cells, physiological sensors)[237]. Pulsar will focus its work on two main application domains: **safety/security** and **healthcare**. Our experience in videosurveillance (see figure 4) for safety/security [149, 139, 132, 133, 134] is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluations and funding. Pulsar will develop new research activities for activity monitoring for healthcare (in particular the assistance of elderly). In this domain global trajectory analysis and short term temporal analysis are not sufficient. An issue will be to develop detailed human shape and long term temporal analyses. We have initiated new collaborations with several hospitals (CHU Nice, Prof Chauvel in Marseille La Timone) to understand their needs.

Figure 3: different views of the apron area captured by the video cameras

We have set-up an experimental laboratory (Ger'home) at Sophia Antipolis together with CSTB (the French scientific and technical centre for building,http://international.cstb.fr/) to test new sensors and new activity recognition techniques (see figure 5).

3 From Orion to Pulsar

Among the five permanent research scientists who are members of Pulsar, four belonged to the Orion INRIA project team; G. Charpiat has recently been hired and he will join the team in December 2007. Orion was created in July 1995 in the theme Data bases, Knowledge bases and Cognitive Systems (3A in short). Orion was, in the same way as Pulsar is now, a multi-disciplinary team at the frontier of computer vision, knowledge-based systems, and software engineering. In 1995 the two research axes were Image Understanding and Program Supervision. There have been several evolutions of these research axes during the last years. The image understanding axis had two main evolutions: first the increasing importance of research in video understanding and second the emergence of a cognitive vision approach mixing a priori knowledge [145, 238] with ontology [153], learning [152] and even program supervision [150]. The program supervision axis has evolved from a knowledge based approach for reusing a library of programs to a research axis on object-oriented framework for knowledge-based systems which led to the LAMA platform [1].

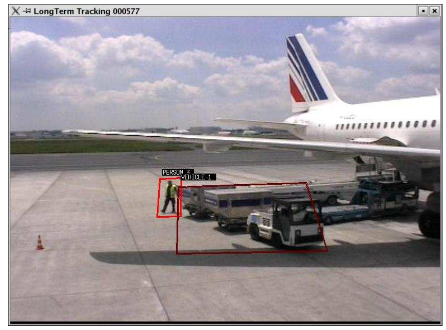
The two main research directions of Pulsar are respectively scene understanding for activity recognition and software architecture for activity recognition. The **major differences between Orion and Pulsar** are described below.

First, a focus on spatio-temporal activities of physical objects which will be studied not only from the scene understanding point of view but also from the software engineering point of view.

Second, the new application domain in healthcare, especially for assistance to the elderly, which will bring interesting challenging new problems.

[1]LAMA [254] is a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. It offers toolkits to build and to adapt all the software elements that compose a knowledge based system or a cognitive system (for more details see Annex A).

unruly behaviors detection on train



aprons monitoring on airport

bank agency access control

lock chamber monitoring

Figure 4: Typical videosurveillance problems addressed by Orion

Elderly assistance in Ger'home project at CSTB premises

Epileptic patient monitoring
(courtesy to Pr Chauvel from Inserm/La Timone hospital)

Figure 5: Examples of activity monitoring for healthcare

Third, a new research area in machine learning with the objective to propose complete cognitive systems mixing perception, understanding and learning techniques.

Finally to enrich the visual data with other sensors.

Pulsar can rely on several results obtained by Orion in the last years. Among them we can cite the following ones:

- an original 4D semantic approach for video understanding including mobile object detection, classification, tracking and complex scenario recognition [144, 146, 147, 149, 255, 256, 258],
- an original program supervision approach for software reuse including knowledge description languages for operators and planning techniques [154, 158, 159, 207, 210, 231, 256],
- a visual concept ontology [152, 153, 252] and a video event ontology [258],
- the VSIP[2] platform for real-time video understanding [139, 150, 250]; VSIP led to several products and to the creation of Keeneo (www.keeneo.com) in July 2005 for its industrialisation (see Annex B for more details),
- the LAMA platform that provides a unified environment to design knowledge bases and inference engines [156, 157, 254] (see Annex A for more details).

Annex C summarizes the performances obtained by Orion in videosurveillance using our 4D semantic approach and our VSIP real-time video understanding library. This annex presents the results of three evaluation campaigns led by industrial partners and end-users in the European project ADVISOR, the industrial project Cassiopee and the industrial project Videa. In conclusion these performance evaluations show that our current systems are very robust for simple activities mainly based on the 3D localisation and tracking of a small number of persons inside a metro station, a bank agency or a building. Globally there are more than 90% of true positive (or correct) detections, less than 2 false alarms over several hours. The results for more complex activities such as fighting are promising (95% true positives in the case of isolated persons inside a metro station and about 60% in the case of far field of view of a urban busy environment) but these results are fragile and need to be improved. **In particular new research need to be launched** for *improving the precision of object description* and *combining a priori knowledge and learning techniques* for setting the various models of an activity recognition system.

4 Research Directions

We propose to follow two main research directions: scene understanding for activity recognition and software architecture for activity recognition. **These two research directions are interleaved**: the software architecture direction will provide new methodologies for building safe activity recognition systems and the scene understanding direction will provide new activity recognition techniques which will be designed and validated for concrete videosurveillance and healthcare applications. Conversely these concrete systems will raise new software issues that will enrich the software architecture research direction.

[2]VSIP is a real-time Intelligent Videosurveillance Software Platform composed of (1) image processing algorithms in charge of mobile object detection, classification and frame to frame tracking; (2) multi-camera algorithms for the spatial and temporal analysis (4D) of the detected mobile objects; (3) a high level scenario recognition algorithms. VSIP is oriented to help developers building activity monitoring systems and describing their own scenarios dedicated to specific applications.

4.1 Scene Understanding for Activity Recognition

Participants[3] : **François Brémond, Guillaume Charpiat, Monique Thonnat** and Sabine Moisan

Despite few success stories, such as traffic monitoring (e.g. http://www.citilog.fr), swimming pool monitoring (e.g. http://www.poseidon-tech.com) and intrusion detection (e.g. http://www.objectvideo.com), scene understanding systems remain brittle and can function only under restrictive conditions (e.g. daylight, diffuse lighting conditions, no shadows), with poor performance over time [139]. Moreover, these systems are very specific and need to be developed from scratch for a new application. To address these issues, most researchers have tried to develop new vision algorithms, robust enough to handle real life conditions. Up to now no vision algorithm has been able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives.

4.1.1 Related Work

The state-of-the-art in scene understanding embraces several decades of productive work. It recently led to a new research theme: *Cognitive Vision*.

Scene understanding is an ambitious research topic which aims at solving the complete interpretation problem ranging from low level signal analysis to semantic description of what is happening in the scene viewed by video cameras and possibly other sensors. This issue implies solving several problems grouped in three major categories: perception, understanding and learning.

Video Processing for Activity Recognition Several teams (e.g. University of Leeds, University of Hamburg, University of Southern California, Georgia Tech, University of Central Florida, University of Kingston upon Thames, National University of Cheng Kung and Prima Team at Inria Grenoble) work on video processing for activity recognition, and a huge literature describes video processing techniques. Among the classical approaches, we can cite motion estimation [126, 115, 36], object detection from a background image [82, 106], feature-based object detection [65, 85], template-based object detection [48], crowd detection [130, 6], human-body detection [116, 59, 114, 29, 38, 2], gait analysis [96, 76], vision systems [60, 34, 44] and video processing performance evaluation [13, 64, 131]. Several surveys [3, 118, 90] categorize these techniques in a more or less exhaustive way.

We highlight one major issue related to perception for activity recognition: object tracking and multi-sensor analysis. Much work studied tracking algorithms (mostly based on feature-tracking [54, 86] or template-tracking [18, 28] using for instance probabilistic techniques [43, 55, 24], 3D techniques [37, 113], active-vision techniques [49], long-term techniques [128, 10] or graph-based techniques [112, 58]) and fusion of information techniques (mostly with multiple-cameras [103, 42, 17] and, less frequently by combining video cameras with other sensors [40, 31, 51]) to improve results and robustness. In particular, an interesting work [89] has addressed the problem of multi-view tracking using synchronized cameras. This system is able to combine information coming from multiple camera pairs (i.e., up to 16 synchronized cameras were used in the experiments) in order to handle occlusions and correctly track densely located objects in a cluttered scene. However, due to complexity, the system is not yet able to work in real-time (i.e., it currently takes

[3]The main permanent research scientists working in this research direction are in bold font, permanent research scientist working only partly in this research direction are in regular font

approximately 5 seconds per processing step). In the same way, recent work has been done to combine different types of media (video, audio and others) [237] to move from video understanding to scene understanding. However, a general framework is needed to perform this information fusion seamless. In comparison, the Pulsar team will design a global framework for 4D spatio-temporal reasoning.

Understanding Concerning related work in high level understanding for activity recognition we will focus on two aspects: event recognition and ontology-based knowledge acquisition.

An increasing effort is done to recognize events from videos [9, 19, 122], in particular using motion-based techniques [27, 99, 80, 97]. For instance several studies have been done on home-care applications [71, 30]. Along the work achieved in event recognition, two main categories of approaches are used to recognize temporal scenarios based on (1) a probabilistic network [110, 73, 124] (especially using Bayesian techniques [109, 83] or Hidden Markov models HMMs [52, 7]) or a neural network [16, 72] combining potentially recognized scenarios, (2) a symbolic network that stores all previously recognized scenarios [94, 101, 41]. For instance, for the probabilistic techniques, Hongeng *et al.* [70] propose a scenario recognition method that uses concurrent Bayesian threads to estimate the likelihood of potential scenarios. Most of these techniques allow an efficient recognition of scenarios, but there are still some temporal constraints which cannot be processed. Moreover, most of these approaches require that the scenarios are bounded in time. To address this issue and to handle uncertainty, the Pulsar team will extend the Orion's original approach on scenario recognition based on temporal constraint checking [238, 258]. Knowledge acquisition is still a complex and time-consuming task but it is required to obtain an effective system to recognize complex objects. Most approaches use dedicated knowledge dependent on the application domain. For instance, in [117] object description is based on domain related annotations. We propose to use a perceptual concept ontology to make easier the knowledge acquisition task in the same way as we have done in [153, 152, 252] for static 2D visual concepts.

Learning Concerning related work in learning for activity recognition, we focus on two aspects: learning techniques for object and event recognition and learning techniques for program parameter tuning.

Regarding event recognition, most approaches learn event patterns using statistical techniques [125, 79, 68, 50] and several approaches can deduce abnormalities from the learnt events [121, 127]. Regarding object recognition, a great effort has also been devoted to representing objects via their 2D characteristics computed with statistical techniques [92, 33, 81]. For instance, the Lear project of INRIA is working on object recognition and scene interpretation from static images and video sequences. They use shape and affine interest point information to describe objects in images. This numerical content is learned to build visual models of object classes. Therefore, visual description and object models are mainly based on low-level features. They do not use symbolic information to describe the objects.

Little work has been done to design learning techniques for system tuning and for setting program parameters [63, 105, 88]. For instance, in the Prima Team at INRIA Grenoble, Hall [62] proposes an automatic parameter regulation scheme for a tracking system. An auto-critical function is able to detect a drop in system performances with respect to a scene reference model. In such a situation, an automatic regulation module is triggered to provide a new parameter setting with better performances. The comparison with a manually tuned system shows that the controlled system is not able to reach the same performances. This is due to the fact that the controlled system does not use a priori knowledge of programs whereas the expert takes advantage of this kind of knowledge and is rapidly able to understand which parameters are related to which tracking problems. This work shows that this research domain is new and that more work is still needed. To tackle this problem, the Pulsar team proposes to use program supervision techniques to configure and control system execution, following the promising results we have obtained in [150, 250]. This involves adding a reasoning level which combines a priori knowledge of the program to tune and learned parameter values in function of the contextual class.

4.1.2 Proposed Approach

Our goal is to propose innovative techniques for autonomous and effective scene understanding systems for long-term activity recognition [150, 250]. This objective is very ambitious; however the current state-of-the-art techniques in cognitive vision already led to partial solutions [23, 95, 26, 123].

The major issue in semantic interpretation of dynamic scenes is the gap between the subjective interpretation of data and the objective measures provided by the sensors. Our approach to solve this problem is to keep a clear boundary between the application dependent subjective interpretations and the objective analysis of the videos. We thus define a set of objective measures which can be extracted in real-time from the videos, we also propose formal models to enable users to define their activities of interest and we build matching techniques to bridge the gap between the objective measures and the activity models.

Scene understanding is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and activities. The signal and the feature levels are characterized by strong noise, ambiguous, corrupted and missing data. The whole process of scene understanding consists in analysing this information to bring forth pertinent insights of the scene and its dynamics. Moreover, to reach a semantic abstraction level, activity models are the crucial points. A still open issue consists in determining whether these models should be given a priori or learned. The whole challenge consists in organizing this knowledge in order to capitalize experience, share it with others and update it along with experimentation. To face this challenge, tools in knowledge engineering such as ontologies are needed.

More precisely we plan to work along the following research axes: perception (how to detect physical objects), understanding (how to recognize their activities based on high level activity models) and learning (how to learn the models needed for activity recognition). The results of this work will be integrated as perception, understanding and learning components in the PULSAR platform for activity recognition which will be described in section 4.2.3.

Perception for Activity Recognition Our contribution in perception will be twofolds: first to propose new techniques for physical object recognition, and second to deepen our work on the management of a library of video processing programs.

A first issue is to recognize in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. We will not work on hardware solutions (as GPU Graphics Processing Unit or FPGA Field-Programmable Gate Array) for getting video frame rate performance but we will design software solutions either by adapting existing algorithms or by proposing new ones. We plan to work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). Currently we are working with CSTB (Centre Scientific et Technique du Batiment) to setup Ger'home (http://gerhome.cstb.fr/), an experimental laboratory to study different sensor configurations. For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). Despite all the work done in object detection and tracking, state of the art algorithms remain brittle. To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modeling and managing the uncertainty of these processes is also an open issue. In Pulsar we propose to add a reasoning layer to a classical bayesian framework modelling the uncertainty of the tracked objects. This reasoning layer will take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking. Moreover a lot of research work is needed to provide fine and accurate models of human shape and gesture. Currently the VSIP library is limited to coarse motion detection and global 3D features. We plan to deepen the work we have done on human posture recognition with B. Boulay [144, 176, 177] to match 3D models and 2D silhouettes. We also plan to work on gesture recognition based on feature tracking; for instance M. B. Kaaniche is starting a PhD on this topic. These kinds of detailed 3D features are required, for instance, for patient monitoring in healthcare.

A second issue is to manage a library of video processing programs. One contribution in this domain will be to build a perception library in the same way VSIP was developed in Orion. More precisely we will not only select robust algorithms for feature extraction but also ensure they work efficiently with real-time constraints and formalize their conditions of use within the program supervision models. In the case of video cameras at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues we plan to use supervised learning techniques (see the following paragraph on learning).

Understanding for Activity Recognition A second research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and on the objective perceptual measures (e.g. robust and coherent object tracks).

We propose to define new activity recognition algorithms and activity models. Activity recognition includes the computation of relationships between physical objects. The real challenge is to explore efficiently all the possible spatio-temporal relationships of the objects that may correspond to activities of interest. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event). One challenge is to explore and analyse this large event space without getting lost in combinatorial searches during the on-line processing of the video streams.

Concerning the modeling of activities, there are two kinds of open problems: the introduction of uncertainty in the formalism for expressing a priori knowledge and the development of knowledge acquisition facilities based on ontological engineering techniques. For the first problem we will investigate either classical statistical techniques and fuzzy logics or new approaches mixing the two kinds of techniques such as Markov Logic Networks [39]. We need to extend our current visual concept ontology (currently limited to color, texture and spatial concepts) with temporal concepts (motion, trajectories, events, ...) and other perceptual concepts (such as audio concepts or physiological sensor concepts). A short term goal will be to add audio concepts respectively in the framework of the European SERKET project for security and in the framework of the European project CARETAKER [179] for activity monitoring.

Learning for Activity Recognition The third research axis is to study which machine learning techniques could help to learn the models needed for activity recognition. Since it is still very difficult to build an activity recognition system for a new application, we propose to study how machine learning techniques can automate model building or model enrichment at the perception level and at the understanding level.

At the perception level, a remaining open problem is the low level detection and in particular image segmentation. A possible approach to improve image segmentation is to use learning techniques for image segmentation method selection and for image segmentation parameter optimization. A PhD student (V. Martin[208]) is currently working on this topic. For an image sampling set a user provides ground truth data (manual region boundary and semantic labels). An evaluation metric together with an optimization scheme based on the simplex algorithm or a genetic algorithm are applied for finding the best parameters for each segmentation algorithm. Another work we have done with B. Georis [150, 250]) has shown in a specific example of illumination change how clustering techniques applied to intensity histograms could help in learning the different classes of illumination context for dynamic parameter setting. This research work can be generalized, for example, to learn parameter initialization criteria in the program supervision formalism. More generally there is a need for learning techniques for program supervision knowledge base refinement.

At the understanding level two main research topics are planned: a first topic is the learning of primitive event detectors. This can be done in the same way as in Nicolas Maillot's PhD [252, 153, 152], by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual features samples. This work was restricted to color, texture and simple shapes. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. lowering the human annotation task)? A second topic is the learning of typical composite event models for frequent activities [233] using data mining techniques. We name composite event a particular combination of several primitive events. This topic is very important for healthcare activity monitoring where large amounts of data are available but little a priori knowledge is formalized.

Coupling learning techniques with a priori knowledge techniques is very promising but still at a very preliminary stage for activity recognition and, in particular, to recognize meaningful semantic activities.

Pulsar being interested in proposing new techniques for activity recognition systems, the results of the research performed in scene understanding (first research direction) will contribute to specifying the needs for new software architectures (second research direction).

4.2 Software Architecture for Activity Recognition

Participants[4]: **Sabine Moisan, Annie Ressouche**, Jean-Paul Rigault, François Brémond, Monique Thonnat.

In Pulsar, our first aim is to design a new platform dedicated to activity recognition that integrates and extends the functionalities of both LAMA and VSIP; the architecture of this platform will follow the same Software Engineering approach that led to LAMA. Second, we wish to contribute to set up, in the long term, a complete software engineering methodology for the development of activity recognition systems. This methodology should be based on Model Engineering and formal techniques.

[4]The main permanent research scientists working in this research direction are in bold font, other permanent research scientist working only partly in this research direction are in regular font

4.2.1 Software Engineering Aspects of Activity Recognition

Activity recognition systems are complex to design due to a number of characteristic properties. First, their *algorithmic structure* is complex: these systems encompass *many abstraction levels*, from pixel-based processing to high level knowledge and data handling; each level as well as their interactions impact the quality of the overall system. Second, their *architecture* is inherently *distributed* and often *heterogeneous* at almost all levels: many sensors of various kinds, local processing, distributed ressources, distributed knowledge and data, etc. Third, these systems have to be *efficient* since they have to work more and more in *real time*. Fourth, the *target applications* are demanding in terms of *dependability*, *safety* and *privacy* (e.g., transport, medical domains). Fifth and finally, and for all the previous reasons, these systems have *requirements* that are *difficult to express and validate.*

A single characteristic property in the previous list makes software development a toil. Combining them constitutes a real challenge which calls for sophisticated Software Engineering (SE) techniques to set up a clear methodology and develop adequate supporting tools. It is not sufficient, however, to adopt well proven traditional SE methods; one needs to invent new ones, possibly picking, adapting, and developing state of the art research in this domain. Hence, our goal is not only to provide operational software tools, but also to draw new methodologies to build these tools.

In particular two bodies of SE techniques, both of which have drawn a considerable interest in the recent years, may be fruitfully applied to the design of activity recognition systems: *genericity* for organizing design and code, and *model engineering* for methodological guidance.

Developing generic entities has always been a major task of software engineering, since its creation in 1969. Indeed such entities can be highly reusable, sparing both development and testing/validation time. Consequently, they also promote extensibility, dependability, and maintainability. Subprograms, modules, components, objects, frameworks, all of these contribute to the domain. There is a danger, though: genericity implies raising the abstraction level, possibly at the expense of specific application requirements or demanded performance. But software engineering has always striven to produce techniques and tools that try to reconcile reusability with adaptability to a specific task. Nowadays, the key studies in genericity involve the use of patterns [53] and object-oriented or component-based frameworks [45] as well as the numerous techniques hidden behind the general term of *generative programming* [32] (metaprogramming, aspect-oriented programming, etc.) [1, 77].

A few years ago, the Object Management Group (OMG) launched an initiative to develop a so-called *Model-driven Architecture* (MDA) [12, 87]. In the strictest sense, MDA is an attempt to favor a seamless development of "business" software using models (based on the UML, another production of OMG) and automatic transformations of these models, but establishing a clear separation between the high level application requirements and the middle and low level implementations, which are platform dependent. The current research broadens the scope of MDA, envisioning to design systems by an automatic (or semi-automatic, but controlled and validated) sequence of *model transformations.* With this meaning the approach is best designated as *Model Engineering* or *Model Driven Development* (MDD). Also, we do not underestimate the important work of the *Software Architecture* community [5] and think that both approaches are worth being merged.

The Software Architecture research direction proposed by Pulsar—as a continuation of previous Orion work—concerns the study of generic systems for activity recognition as

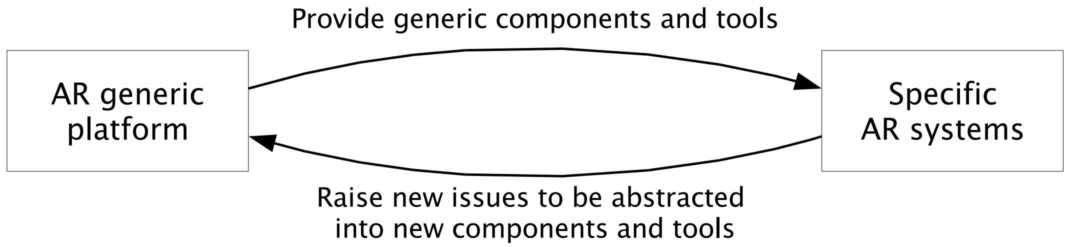


Figure 6: The two Pulsar research axes consolidate each for producing activity recognition (AR) platmform and systems.

well as the definition of a methodology for their design. We wish to ensure genericity, modularity, reusability, extensibility, dependability, and maintainability, but still ensuring adequate (real time) performance. To tackle this challenge, we rely on state of the art Software Engineering practice. In particular we shall explore the new techniques mentioned above, those related to genericity on the architecture side as well as model engineering approaches on the methodological side.

It is almost impossible to achieve genericity from scratch, hence our approach is pragmatic: it is indeed fed by the problems that Pulsar has to face in image and video understanding and more generally in activity recognition. Now that we have reached some maturity in image and scene understanding, we can consider building more general solutions and tools. Therefore, we need to abstract from our experience with operational systems to provide more generic solutions and user-friendly tools. Once the corresponding generic entities are available, they can be applied to new problems. This new experience will in turn be abstracted, enriching generic models and platform architecture. Therefore, the two Pulsar research axes (scene understanding and software architecture) consolidate each other in a “virtuous cycle”: as shown on figure 6, an operationalization flow produces effective systems from the generic platform whereas an abstraction flow enriches the platform from the lessons learned with concrete applications.

Another important characteristic of this research direction is that we strive to develop models with formally correct foundations. This makes it possible to build dependable systems relying on formal verification techniques, since traditional testing is not sufficient in these domains.

This approach proved feasible with the development of the Lama platform (see Appendix A) in the Orion project-team. This set of object-oriented component frameworks provides generic components (as C++ classes) for generating knowledge-based systems generators. Hence, it is a meta-meta-platform (thus rather high on the generic level scale!). Its genericity has been put to the test by using Lama to develop systems requiring very different types of tasks: model calibration, image classification, or video understanding. Yet, its high generic level is not at the expense of efficiency: for instance in a video understanding application, the planning engine generated with Lama provides high versatility but accounts for less that 4% of the overall execution time.

4.2.2 Related Work

Models Historically the first Artificial Intelligence task that we have addressed in Orion was Program Supervision. Program Supervision has two major aspects: setting up models to represent knowledge about the use of programs and controlling their use through inference engines. Hence, it requires knowledge representation [210] as well as (task) planning techniques. Thus it draws on previous work in these two domains [102, 93]. However most of these studies were generally motivated by specific application domain needs (for instance, image processing, signal processing, or scientific computing). By contrast our approach relies on knowledge-based system techniques.

Similar work has been developed, especially in the knowledge acquisition community: their main objective is not to reuse code and programs (as we do) but knowledge itself [108, 46]. For instance we can cite Protege from Stanford University [56], which is oriented toward ontology editing, knowledge representation, and reuse of abstract methods for problem solving. Program supervision suits well the image processing area: we can cite the ExTI system (IRIT, Toulouse) or the Hermes and Panthéon projects [104]. In other domains, the work of Chien and his collaborators at the Jet Propulsion Laboratory (NASA) [20] has an approach close to LAMA, although the task and domain are different. They propose two generic component systems based on artificial intelligence techniques, ASPEN and CASPER. Close to program supervision, the AROM platform also relies on a generic approach [57].

Some issues in the Semantic Web and Web Services research [8, 35] are also related to this topic as far as knowledge representation is concerned. On the other hand there exists, in the Semantic Web, a dynamic control aspect that could be a matter for program supervision techniques (see for instance the efforts towards OWL-S and Jena[5]).

Platform There are several types of cooperation between artificial intelligence and software engineering. Much work deals with applying AI techniques for better software development. We subscribe to the opposite viewpoint: we rather use SE techniques to improve AI system design, a less common line of work.

LAMA is composed of object-oriented *frameworks* (written in C++ and Java) [156]. Roughly speaking a *framework* is a generic program (and associated libraries) bringing solutions to recurrent problems in a particular (but usually wide) domain [74]. Because of its generic nature, a framework has to be adapted to a specific application in its problem domain. For a long time now, frameworks have been considered as a powerful tool for designing and building complex software systems in a rather economical way. As of today, many domains have dedicated frameworks (e.g., graphic interfaces, concurrent systems, Web applications, compilation, etc.) whereas we focus on KBS software (i.e., inference engines or knowledge managers).

In artificial intelligence, many frameworks have been developed for, e.g., agents and learning. These approaches rely on component engineering, sharing the same ideas about reusability and genericity [98]. Yet, in the domain of knowledge-based systems, LAMA appears rather original. Indeed one may find “expert systems generators” dedicated to particular tasks, but LAMA raises the genericity level one step higher, since it is a *generator* of knowledge based system *generators*. In particular, the range of LAMA extends beyond Program Supervision itself and can encompass other artificial intelligence tasks, such as model calibration in scientific computing [160] or classification [251, 252] (to name just the ones with which we had a direct experience). Obviously, LAMA must evolve to support distributed components. Some aspects of the previously mentioned Semantic Web clearly relate to this issue. We will study how the corresponding techniques can apply to our case. Moreover, the multi-agent community has also been active in developing frameworks dedicated to the distributed aspect of agents (the “middleware”): for instance Aglets [78], Jade[6], the Oasis team in Paris [15], or Proactive developed in the Oasis project at INRIA Sophia Antipolis[7]. We have already started to use agent techniques (Aglets) for distributing

[5] http://www.daml.org/services/owl-s and http://jena.sourceforge.net
[6] http://jade.tilab.com/
[7] http://www-sop.inria.fr/oasis/ProActive/

Program Supervision systems [151].

Safeness In this matter, our goal is not to develop original research as INRIA projects such as Mimosa or Aoste do. Rather we intend to rely on well known formal techniques (synchronous models[11, 61], model-checking [22, 75]) in the line of our previous work with the Blocks framework [157, 213].

4.2.3 Proposed Approach

As indicated before (see 4.2.1), the research direction we proposed here consists in studying generic systems for activity recognition and in elaborating a methodology for their design. We wish to ensure genericity, modularity, reusability, extensibility, dependability, and maintainability.

We plan to work in the following three research axes: models (adapted to the activity recognition domain), platform architecture (to cope with deployment constraints such as real time or distribution), and system safeness (to generate dependable systems). For all these tasks we shall follow state of the art Software Engineering practices and, if needed, we shall attempt to set up new ones.

Models for Activity Recognition This evolution toward activity recognition requires theoretical studies combined with validation based on the concrete experiments conducted together with the scene understanding research direction. Thus the evolution will certainly not be instantaneous but rather incremental over a long period. Altogether it is a long term goal although some of its aspects can be addressed in the near future.

First, the incorporation of a *model of time* [107, 4], both physical and logical, is needed to deal with temporal activity recognition, especially scenario recognition. We first intend to consider models of logical time based on the Synchronous Paradigm. Indeed this paradign is well suited to deterministic systems and has many nice properties such as being liable to formal verification.

Second, handling uncertainty is a major theme of Pulsar and we want to introduce it into the platform [8]. This might well shake up the synchronous paradigm and requires deep theoretical studies; thus it is a long term goal.

Third, the incorporation of an abstract *model of scenarios* is not only mandatory to describe and recognize activities. It also allows to use scenarios as a means to define new complex operators in Program Supervision. Indeed scenarios could constitute both a specification and a sort of "scripting" mechanism (this second point appears rather speculative and is clearly a long term goal) [164, 66].

Platform Architecture for Activity Recognition Figure 7 details the intended architecture of the AR platform.

It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for Perception, Understanding, and Learning.
  - Perception components contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
  - Understanding components provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.

[8] Lama already provides a fuzziness module, but probabilistic approaches are also needed.

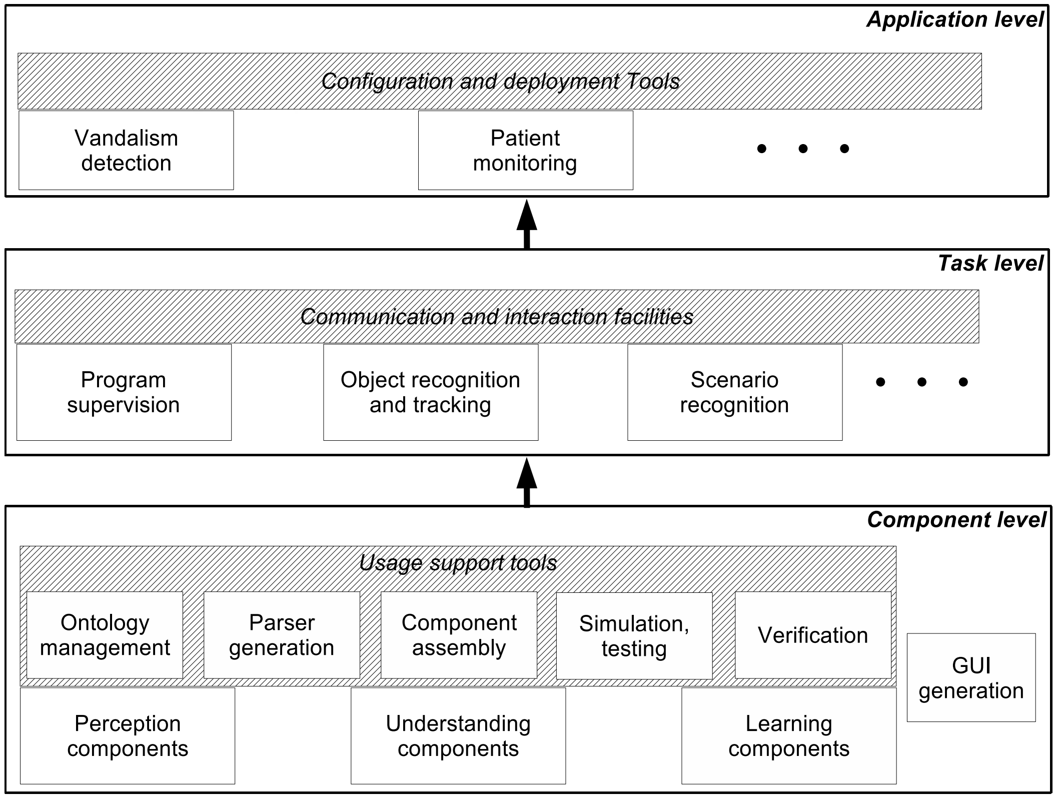


Figure 7: Architecture of the Pulsar activity recognition platform.

– Learning components implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

An Activity Recognition system is likely to pick components from these three packages. Hence, tools must be provided to select, assemble, simulate, verify the resulting component combination. Other support tools help to generate task or application dedicated languages or graphic interfaces.

- The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate for a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Program Supervision to control data processing, Object Recognition and Tracking to extract interesting features from images, Scenario Recognition to identify interesting activities. In the future, other tasks will probably enrich this level.

  For these tasks to nicely collaborate, communication and interaction facilities are needed, providing a sort of “software bus” to exchange data and control.

- The **Application Level** integrates several of these tasks to build a system for a particular type of application, e.g., vandalism detection, patient monitoring, aircraft loading/unloading follow up, etc. Each system is parameterized to adapt to the local context of its instances (number, type, location of sensors, scene geometry, visual parameters, number of objects of interest...). Thus configuration and deployment facilities are required.

The philosophy of this architecture is to offer, at each level, a balance between the widest possible genericity and the maximum effective reusability, in particular at the code level.

An intrinsic feature of most activity recognition applications is to rely on distributed hardware/software and to work in real time. Therefore, the new platform should support distributed architecture and ensure real time performance. Another issue is that the end users are generally not computer experts; thus user friendliness and graphical supports are important. That is why the platform will comply with the following design orientations:

- Concurrent implementation (multi-threading) is known to favor both real time response and distribution. First it allows to take advantage of parallel architectures (multiprocessor, Grid computing...), hence improving real time performance. Second, by enforcing modularity, it makes distribution easier. In particular, we have already designed distributed knowledge-based systems that demonstrated the advantages of concurrency.

- Real time is not just a matter of performance. The new platform should also comply with real time modeling. We shall rely on our experience in this domain [163, 165]. Moreover, real time also impacts the scene understanding algorithms as well as the reasoning mechanisms. In particular, for Program Supervision, the causality of time requires revisiting the *repair strategy*, since real time hardly allows backtracking.

- Supporting distributed systems also requires a *model of distribution*. We chiefly consider multi-agent systems since they appear to be well-suited to artificial intelligence tasks [119, 120, 47]. We shall also work on the possibility of embedding "light" knowledge-based systems (together with associated knowledge) into mobile agents.

- Last, but not least, the new platform should be easier to use than the current LAMA and VSIP. We should thus define and implement tools to support modeling, design and verification. Another important issue deals with graphical user interfaces. It should be possible to plug existing (domain or application dependent) graphical interfaces into the platform. This requires defining an extra generic layer to accommodate various sorts of interfaces. This is clearly a medium/long term goal, in its full generality at least.

Concerning implementation issues, we use when possible existing standard tools such as bison and flex for parser generators, NuSMV for model-checking, Protege for ontology management, Eclipse for graphic interfaces or model engineering support, etc. Note that, in figure 7, some of the boxes can be naturally adapted from LAMA and VSIP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

Safeness of Systems for Activity Recognition In most activity recognition systems, safeness is a crucial issue. It is a very general notion dealing with person and goods protection, respect of privacy, or even legal constraints. However, when designing software systems it will end up with software security. Here we only consider software security issues. Our previous work on verification and validation has to be pursued; in particular, we need to check its scalability and to develop associated tools.

Generation of dependable systems is an important challenge. System validation is a crucial phase in any development cycle. Testing, although required in the first phase of validation, appears to be too weak for the system to be completely trusted. An exhaustive approach of validation using formal methods is clearly needed. Model-checking is an appealing technique since it can be automatized [84, 69, 100, 67, 21] and helps to produce a code that has been formally proved.

We base verification on the component structure of our frameworks. This makes it possible to follow a compositional approach [213], a well-know technique to cope with scalability problems in model-checking.

In the long term we wish to take into account time and uncertainty. To formally describe probabilistic timed systems, the most "popular" approach proposes probabilistic extensions of timed automata (such as real-time probabilistic processes, probabilistic timed automata and continuous probabilistic timed automata). The semantics of these formalisms relies on infinite-state structures but equivalence relations can be applied to obtain finite-state systems. Then these systems can be used for model-checking [111], provided that suitable dedicated logics are defined (such as branching time temporal logic [129]).

Nevertheless, software dependability cannot be proved by relying on a single technique. Indeed, model-checking works on a *model* of the system and allows to automate the verification of decidable properties. By contrast some undecidable properties require to work directly on the *code*, for instance by applying non exhaustive methods such as abstract interpretation [25]. Thus, verification methods must recurse to several complementary techniques. This is indeed a major issue and a rather challenging one. Altogether, this is a long term goal, with an incremental evolution.

5 Objectives for the next 4 years

The general objectives presented in the previous sections are long term goals. Depending on the human resources and funding we will focus our activity on the following priorities:

5.1 Scene Understanding for Activity Recognition

1. Perception:

   - Gesture recognition: currently the VSIP library is limited to coarse motion detection and global 3D features. We plan to refine people description based on recent work we have achieved with B. Boulay [144] on real time posture detection. Moreover a PhD student currently works on gesture recognition (M. B.Kaaniche). These kinds of detailed 3D features are required for instance for patient monitoring.
   - Multi sensors perception: we plan to extend our 4D semantic approach to heterogeneous sensor data interpretation. A PhD thesis (N. Zouba) is starting on this topic in the context of healthcare monitoring.

2. Understanding:

   - Ontology-based activity recognition: we need to extend our current visual concept ontology [152, 153, 252] (currently limited to color, texture and geometry) with temporal concepts (motion, trajectories, events, ...) and other perceptual concepts (such as audio events or physiological sensor events). A short term goal will be to add audio concepts respectively in the context of the European SERKET project for security and in the context of the European project CARETAKER for activity monitoring.
   - Uncertainty management: Concerning the modeling of activity we plan to extend our formal activity model [238] to take into account the uncertainty of the detected physical objects. An example of use of the current formalism is shown

in figure 2. A major issue is to keep the realtime performances of the recognition process.

3. Learning:

   - Primitive event detectors: an important research topic is the learning of primitive event detectors. This can be done by learning visual concept detectors with perceptual features samples. An open question is for each type of perceptual concept how far we can go in weakly supervised learning. Two PhD theses (M. Zúniga and L. Le Thi [227]) are starting on this topic.

5.2 Software Architecture for Activity Recognition

1. Models and methods:

   - Model of time and scenarios: a *model of time*, both physical and logical, is needed to deal with temporal activity recognition. We shall propose a PhD to explore this topic. Moreover, the extension with an abstract *model of scenarios* is unavoidable for activity recognition and might also constitute a PhD work.
   - Real time support: we intend to make software generated with the Pulsar platform able to work in *real time*. In particular, for Program Supervision [150, 250], this requires to revisit the *repair strategy*, since real time hardly allows backtracking. A PhD is planned to start in 2008.
   - Methodology: we shall start to explore the applicability of model driven development (MDD) to activity recognition systems. Of course, setting up a complete methodology is a long term objective and within a period of 4 years we only expect feasibility results.

2. Platform:

   - We intend to propose in the next 4 years a first version of the Pulsar platform. This platform will integrate most of the perception and understanding components belonging to VSIP and LAMA. Moreover we plan to develop new support tools mainly for component assembly.
   - Concurrent implementation (multi-threading) is underway. It will ensure better modularity and evolving as well as faster real time response. It will also make it easier to distribute knowledge-based systems.
   - Distributed architecture: a PhD (N. Khayati) is already underway on this topic; it uses multi-agents technology. The objectives are to provide broader accesses and collaboration facilities, and to take advantage of scattered computing resources or data. Distribution also demands incorporating mechanisms to enforce privacy and security within the systems generated with the platform. Beyond the well-known system issues, we have to cope with semantic and knowledge representations. Hence we will propose ontologies for describing, deploying, and using distributed knowledge-based systems.
   - User interface and tools: to support the methodological aspect, tools and user-friendly interfaces are necessary. This is indeed a time-consuming activity and development in these domains will depend on the available work power.

Short term (4 years) concrete goals in terms of activity recognition systems are :

- to design manually a real-time recognition system for complex activities involving interactions between a person and his/her environment based on postures with fixed large field of view cameras.
- to automate the set-up (parametrization) of an activity recognition system (such as an intrusion detection system) for a new site.

Long term concrete goals in terms of activity recognition systems are :

- to generate semi-automatically new activity recognition systems using a tool box and high level specifications (in terms of activity to recognize and constraints).
- to quantify the performances of a new activity recognition system based on benchmarks.

6 Evaluation

We are aware that Pulsar research directions are very ambitious. Our concrete goal in terms of academic impact is to publish both in computer vision (in particular CVIU and IJCV international journals) and in artificial intelligence (IJCAI conference, AI Journal). We also consider conferences and journals establishing links between Software Engineering and Artificial Intelligence (such as SEKE or Data and Knowledge Engineering) as well as the ones more specifically devoted to Software engineering (ICSE) with a particular emphasis on those promoting the model-driven approaches (MoDeLs conference, SoSym journal). We also intend to be actors in the major conferences in vision and cognitive systems (such as ICVS) and in visual surveillance (such as AVSS).

Activity recognition systems are difficult to evaluate, since their performance depends directly on the test videos and on the system configuration phase. A robust algorithm tested in new environmental conditions or with inadequate parameters can get poor performance. Thus, we will continue our work performed in the evaluation program ETISEO [217] (http://www.etiseo.net) to design novel evaluation metrics taking into account the dependencies with video characteristics (e.g. contrasted shadows, illumination changes) and configuration phases (e.g. parameter tuning, object model learning). We will also continue to participate through industrial cooperations to performance evaluation campaigns made with data taken with the existing operational video cameras on real sites (see annex A for some recent experimentations).

7 Positioning and Collaborations

7.1 Positioning within INRIA

PULSAR has an original multidisciplinary approach mixing techniques from the domain of perception, cognitive systems and software engineering for activity recognition.

Perception through video streams

Several INRIA teams are working on video analysis for human activity description and thus are close to PULSAR objectives. Currently we do not have effective collaborations with these teams, but we intend to develop new collaborations on certains topics.

- **PRIMA** The PRIMA research team, which is also in CogC is very close to PULSAR as it shares with PULSAR a cognitive system approach. They mainly work on 2D appearance-based approaches whereas Pulsar will design 4D spatio-temporal video understanding approaches.

- **VISTA** The VISTA research team has a strong background in 2D motion analysis. More recently they have been interested in motion event recognition (cf Y. Laptev's work on repetitive motion detectors). Pulsar's work is complementary to these approaches since we are using 3D models and human expertise. A new cooperation with VISTA for learning new motion event detectors for instance between Y. Laptev and G. Charpiat would be certainly fruitful.

- **LEAR** The LEAR research team is working on object recognition for static images and video sequences. They focus on learning low level visual features. In particular they use shape and affine interest point information to describe objects in images. This numerical content is learned to build visual models of object classes. A new cooperation with LEAR is certainly interesting for learning new shape and texture descriptors.

- **MOVI/Perception** The MOVI research team is currently creating a new project named Perception. Their background is oriented towards 3D geometric models. Depending on their new orientations we could have fruitful cooperations.

Perception through other sensors

For audio sensor understanding new cooperations with the projects of CogC are possible and interesting, in particular with PAROLE, METISS and/or CORDIAL. A new contact is starting with DEMAR (Bio) for the study of wearable physiological sensors for healthcare monitoring.

Cognitive Systems

We share common interests in knowledge representation, knowledge acquisition, ontologies, machine learning and reasoning with several INRIA project-teams mainly in CogA. Among them we can cite the following ones:

- **DREAM** This project works on diagnosis and surveillance systems using artificial intelligence techniques. We will continue our unformal cooperations on temporal activity modeling and reasoning.

- **MAIA** This project is focused on autonomous real-time robotic systems . We share common objectives on real-time artificial intelligence techniques for signal understanding and on healthcare applications.

- **ACACIA/EDELWEISS** This project develops knowledge management and ontologies techniques mainly for Semantic Web applications. We had unformal cooperations with ACACIA on ontology languages which are interesting to continue.

- **HELIX** This is a project with which we have already collaborated [91]; it works on components, platforms, and problem solving environments, all concerns which are close to ours.

- **AXIS** This multi-disciplinary project-team (Artificial Intelligence, Data Analysis and Software Engineering) studies the area of Information Systems, particularly Web sites. They develop data mining techniques for modeling web user activity profiles. Unformal cooperations with AXIS are interesting on data mining techniques for modeling long term people activities observed by sensors in particular for healthcare applications.

- **CORTEX** This project studies neuronal connectionist models developed through two sources of inspiration, namely computational neurosciences and machine learning. Their goal is to build intelligent systems, able to extract knowledge from data and to manipulate that knowledge to solve problems. A cooperation in machine learning techniques and more precisely connectionist ones would be useful in the future for learning event detectors.

Software engineering

Several INRIA project-teams within ComA or ComC deal with component-based programming. We share the synchronous foundations with AOSTE (a project with which a long collaboration has existed) and ESPRESSO and the concern about formal proof for component assembly with TRISKELL and the two former projects. The techniques and tools developed within OBASCO and JACQUARD could also be interesting for us in the future.

In the domain of distributed and multi-agent systems, the models and tools that OASIS is developing are certainly interesting for distributing knowledge-based systems.

7.2 Academic collaborations and positioning

7.2.1 National

We share common interests with several research teams in France, outside INRIA.

Video Analysis

- We cooperate with the Mathematical Morphology laboratory of Ecoles des Mines de Paris at Fontainebleau (Michel Bilodeau) together with StMicroelectronics in a project for intelligent video cameras. In particular we combine their robust image segmentation algorithms with our 3D human posture recognition algorithm [177, 144].
- We regularly cooperate with the research team directed by Patrik Sayd at CEA in Saclay in the context of videosurveillance [237]. Currently, we work together in the context of the SIC project. They focus in 2D video analysis for people tracking over networks of video cameras while we are in charge of high level scenario understanding.

Software engineering

- We plan to continue our cooperation with the Rainbow team (I3S-UNSA-CNRS) at Sophia Antipolis on the component-based approach and in healthcare monitoring which has produced joint papers [189, 148, 221, 222].
- We already cooperate with the Mosarts team (I3S-UNSA-CNRS) and with the CMA team (Ecole des Mines) at Sophia Antipolis to begin the development of a special purpose language dedicated to activity recognition applications.

Healthcare and Life Sciences

We have initiated a series of new cooperations in the domain of healthcare. These cooperations are a first step which need to be enriched and supported.

- We have started a collaboration with CSTB (Centre Scientifique et Technique du Batiment) and the Nice City Hospital (Groupe de Recherche sur la Tophicité et le Viellissement) in the framework of the GER'HOME project, funded by the PACA region. The GER'HOME project is devoted to experiment and develop techniques that allow long-term people monitoring at home. In this project an experimental home has been built in Sophia Antipolis nearby Pulsar premises [244].

- We have begun a new collaboration with the INSERM neurophysist group headed by Prof Chauvel at La Timone hospital in Marseille for epileptic patient observation with videocameras and EEG signal.

We have a long standing collaboration with INRA at Sophia Antipolis, in the domain of life sciences for the early detection of insects in crops or for the theoretical study of their behavior. This biological domain raises interesting issues in image and video interpretation [199].

7.2.2 International

We have a precise view of the main groups working in our domains of interest through international networks gathering our main competitors and collaborators.

Europe We are in competition with several European research teams (e.g. Prof A Cohn and Prof D. Hoggs from University of Leeds, UK, Prof Neumann from University of Hamburg) on activity recognition and (e.g. Prof. R. Cucchiara from University of Modena, Italy, Prof. G. Jones from University of Kingston, UK) on videosurveillance. We are also involved with some of these European research teams (e.g. University of Leeds, University of Hamburg) in a new network of excellence EuCognition (see http://www.eucognition.org) for the Advancement of Artificial Cognitive Systems. This network is a follow-up and an extension of the Cognitive Vision network of excellence ECVision (for more details see http://www.ECVision.info/home/Home.htm). We also cooperate with several European research teams through European Projects (see next section on industrial collaborations and funding).

US We are competing with US academic research teams (e.g., Prof. Nevatia from University of Southern California, Prof. Bobick for Georgia Tech, Prof. Larry Davis from University of Maryland) on the topic of video event recognition. We also cooperate with these teams for defining standard video event ontology (e.g., Prof. Nevatia from University of Southern California, Prof. R. Chelappa from University of Maryland) and for comparing video understanding algorithm performances (e.g. Prof. M. Shah from University of Florida and Prof. Larry Davis from University of Maryland).

Asia Our team is a member of the STIC-Asie Inter-media Semantic Extraction and Reasoning (ISERE) action. ISERE action gathers six research centers from Asia (I2R A-STAR and NUS for Singapore, MICA for Vietnam, NII Tokyo for Japan and the NCKU and the NTU for Taiwan) and three French teams. It concerns both the development of research on semantics analysis, reasoning and multimedia data, and the application of these results in the domains of e-learning, automatic surveillance and medical issues.

Africa Pulsar will continue to cooperate with ENSI Tunis (Tunisia) in the framework of PAI cooperations; this work deals with software reuse for distributed applications, more particularly in medecine; we co-direct several PhD and Master theses.

7.2.3 General

Our team has launched ETISEO (http://www.etiseo.net), an international academic and industrial competition for the evaluation of videosurveillance techniques with currently 25 international teams.

7.3 Industrial Collaborations and Funding

Pulsar will rely on a strong network of industrial partners in the domain of videosurveillance (see below). Several other industrials are interested in healthcare monitoring such as France Telecom, Accenture and STMicroelectronic.

National Initiatives

- **SYSTEM@TIC SIC project**: Pulsar is strongly involved in the new "pole de competivité" SYSTEM@TIC which is a strategic initiative in security. More precisely a new project (SIC) is accepted for funding for 42 months on the theme of perimeter security. The industrial partners include Thales, EADS, BULL, SAGEM, Bertin, Trusted Logic.

European Projects

- **SERKET** is a European ITEA project in collaboration with Thales R&T FR, Thales Security Syst, CEA, EADS and Bull (France); Atos Origin, INDRA and Universidad de Murcia (Spanish); XT-I, Capvidia, Multitel ABSL, FPMs, ACIC, BARCO, VUB-STRO and VUB-ETRO (Belgium). It has begun at the end of November 2005 and will last 2 years. The main objective of this project is to develop techniques to analyze crowd behaviors and to help in terrorist prevention.

- **CARETAKER** is a new STREP European project that began in march 2006. Its duration is planned for thirty months. The main objective of this project is to discover information in multimedia data. The prime partner is Thales Communications (France) and others partners are: Multitel (Belgium), Kingston University (UK), IDIAP (Switzerland), Roma ATAC Transport Agency (Italy), SOLID software editor for multimedia data basis (Finland) and Brno University of Technology (Czechia). Our team has in charge modeling, recognizing and learning scenarios for frequent or unexpected human activities using both video and audio events.

Industrial Cooperations

- **Bull**: we have a research collaboration with Bull since 1998 through a sequence of contracts. Within three successive industrial contracts (TELESCOPE 1, 2 and 3) from 1998 to march 2006 we have developed together a toolkit in the domain of cognitive video interpretation for videosurveillance applications. We are currently continuing this strategic cooperation through 2 projects in safety/security: the European Project SERKET and the SIC project within the security "pole de compétivité" SYSTEM@TIC.

- **RATP**: We have had, since 2001, a close cooperation with RATP for multisensors access control systems. A PhD is currently working on learning techniques for passenger classification.

- **STmicroelectronic** Orion also cooperates with STmicroelectronics and Ecole des Mines de Paris at Fontainebleau for the design of intelligent cameras including image analysis and interpretation capabilities. In particular one post-doc and one PhD student are currently working on new algorithms for 3D human posture recognition in real-time for video cameras.

- **Keeneo**: the start-up Keeneo was created in July 2005 for the industrialization and exploitation of Orion results in videosurveillance (VSIP library). Pulsar will continue to maintain a close cooperation with Keeneo for impact analysis of VSIP and for exploitation of new results.

8 References

[1] D. Abarahams and A. Gurtovoy. *Template Metaprogramming: Concepts, Tools, and Techniques for Bost and Beyond.* Addison Wesley, 2005.

[2] A. Agarwal and B. Triggs. A local basis representation for estimating human pose from cluttered images. In *Asian Conference on Computer Vision*, January 2006.

[3] J. Aggarwal and Q. Cai. Human motion analysis: a review. In *Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 90–102, Puerto Rico, USA, June 1997.

[4] J.F Allen. Time and time again: The many ways to represent time. *Intern.. Jr. of Intelligent Systems*, 6(4):341–356, July 1991.

[5] R. Allen and D. Garlan. A formal basis for architectural connection. *ACM Transactions on Software Engineering and Methodology*, 6(3):213–249, 1997.

[6] E. Andrade, S. Blunsden, and R. Fisher. Performance analysis of event detection models in crowded scenes. In *Workshop on Towards Robust Visual Surveillance Techniques and Systems at Visual Information Engineering 2006*, pages 427–432, Bangalore, India, September 2006.

[7] E. L. Andrade, S. Blunsden, and R. B. Fisher. Hidden markov models for optical flow analysis in crowds. In *International Conference on Pattern Recognition*, Hong Kong, August 2006.

[8] G. Antoniou and F. van Harmelen. *A Semantic Web Primer.* MIT Press, 2004.

[9] D. Ayers and M. Shah. Monitoring human behavior from video taken in an office environment. *Image and Vision Computing*, 19:833–846, October 2001.

[10] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In *Conference on Computer Vision and Pattern Recognition*, 2006.

[11] G. Berry. The foundations of esterel. In G. Plotkin, C. Stearling, and M. Tofte, editors, *Proof, Language, and Interaction, Essays in Honor of Robin Milner.* MIT Press, 2000.

[12] Sami Beydeda and Matthias Book, editors. *Model-Driven Software Development.* Springer, 2005.

[13] M. Borg, D. J. Thirde, J. M. Ferryman, K. D. Baker, J. Aguilera, and M. Kampel. Evaluation of object tracking for aircraft activity surveillance. In *The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2005)*, Beijing, China, October 2005.

[14] B. Boulay, F. Bremond, and M. Thonnat. Applying 3d human model in a posture recognition system. *Pattern Recognition Letter, Special Issue on vision for Crime Detection and Prevention*, 27(15):1788–1796, 2006.

[15] J.-P. Briot. Actalk: A framework for object-oriented concurrent programming - design and experience. In Jean-Paul Bahsoun, Takanobu Baba, Jean-Pierre Briot, and Akinori Yonezawa, editors, *Object-Based Parallel and Distributed Computing II - Proceedings of the 2nd OBPDC'97 France-Japan Workshop*. Hermès Science Publications, Paris, 1999.

[16] H. Buxton. Learning and understanding dynamic scene activity. In *ECCV Generative Model Based Vision Workshop*, Copenhagen, 2002.

[17] A. Cavallaro. Event detection in underground stations using multiple heterogeneous surveillance cameras. In *International Symposium on Visual Computing, Intelligent Vehicles and Autonomous Navigation*, Lake Tahoe, Nevada, December 2005.

[18] A. Cavallaro, O. Steiger, and T. Ebrahimi. Tracking video objects in cluttered background. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):575– 584, April 2005.

[19] R. Chellappa, A. R. Chowdhury, and S. Zhou. *Recognition of Humans and Their Activities Using Video.* Morgan Claypool, 2005.

[20] S. Chien, R. Knight, A. Stechert, R. Sherwood, and G. Rabideau. Using iterative repair to improve the responsiveness of planning and scheduling. In *Proc. 5th International Conference on Artificial Intelligence Planning and Scheduling (AIPS2000)*, Breckenridge, CO, April 2000.

[21] A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV 2: an OpenSource Tool for Symbolic Model Checking. In Ed Brinksma and Kim Guldstrand Larsen, editors, *Proceeeding CAV*, number 2404 in LNCS, pages 359–364, Copenhagen, Danmark, July 2002. Springer-Verlag.

[22] E.M. Clarke, O. Grumber, and D. Long. Verification tools for finite-state concurrent systems. *LNCS: A Decade of Concurrency, Proc REX School/Symp*, 803:124–175, 1994.

[23] A. G. Cohn, D. C. Hogg, B. Bennett, V. Devin, A. Galata, D. R. Magee, C. Needham, and P. Santos. Cognitive vision: Integrating symbolic qualitative representations with computer vision. *Cognitive Vision Systems: Sampling the Spectrum of Approaches, Lecture Notes in Computer Science*, pages 221–246, 2006.

[24] C. Coué, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessière. Bayesian occupancy filtering for multitarget tracking: an automotive application. *Journal of Robotics Research*, 25(1):19–30, January 2006.

[25] P. Cousot and R. Cousot. On Abstraction in Software Verification. In Ed Brinksma and Kim Guldstrand Larsen, editors, *Proceeeding CAV*, number 2404 in LNCS, pages 37,56, Copenhagen, Danmark, July 2002. Springer-Verlag.

[26] J. L. Crowley. *Cognitive Vision Systems*, chapter Things that See: Context-Aware Multi-modal Interaction. Springer Verlag, March 2006.

[27] J. L. Crowley. Situation models for observing human activity. *ACM Queue Magazine*, May 2006.

[28] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati. The sakbot system for moving object detection and tracking. *Video-based Surveillance Systems: Computer Vision and Distributed Processing (Part II - Detection and Tracking)*, pages 145–158, 2001.

[29] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Probabilistic posture classification for human behaviour analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 35(1):42–54, 2005.

[30] R. Cucchiara, A. Prati, and R. Vezzani. Domotics for disability: smart surveillance and smart video server. In *8th Conference of the Italian Association of Artificial Intelligence - Workshop on Ambient Intelligence*, pages 46–57, Pisa, Italy, September 2003.

[31] R. Cucchiara, A. Prati, R. Vezzani, L. Benini, E. Farella, and P. Zappi. An integrated multi-modal sensor network for video surveillance. *Journal of Ubiquitous Computing and Intelligence (JUCI)*, 2006.

[32] K. Czarnecki and U. W. Eisenecker. *Generative Programming: Methods, Tools, and Applications*. Addison Wesley, 2000.

[33] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *European Conference on Computer Vision*, 2006.

[34] A. C. Davies and S. A. Velastin. Progress in computational intelligence to support cctv surveillance systems. *International Scientific Journal of Computing*, 4(3):76–84, 2005.

[35] J. Davies, D. Fensel, and F. van Harmelen. *Towards the Semantic Web: Ontology-Driven Knowledge Management*. John Wiley & Sons, 2003.

[36] X. Desurmont, J. B. Hayet, J. F. Delaigle, J. Piater, and B. Macq. Trictrac video dataset: Public hdtv synthetic soccer video sequences with ground truth. In *Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE)*, pages 92–100, 2006.

[37] X. Desurmont, I. Ponte, J. Meessen, and J. F. Delaigle. Nonintrusive viewpoint tracking for 3d for perception in smart video conference. In *Three-Dimensional Image Capture and Applications VI*, San Jose, CA USA, January 2006.

[38] M. Dimitrijevic, V. Lepetit, and P. Fua. Human body pose recognition using spatio-temporal templates. In *ICCV workshop on Modeling People and Human Interaction*, Beijing, China, October 2005.

[39] P. Domingos and M. Richardson. Markov logic networks. *Machine Learning*, 62(1-2):107–136, February 2006.

[40] W. Dong and A. Pentland. Multi-sensor data fusion using the influence model. In *Body Sensor Networks Workshop*, Boston, MA, April 2006.

[41] C. Dousson, P. Gaborit, and M. Ghallab. Situation recognition: Representation and algorithms. In *The 13rd IJCAI*, pages 166–172, August 1993.

[42] W. Du, J. B. Hayet, J. Piater, and J. Verly. Collaborative multi-camera tracking of athletes in team sports. In *Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE)*, pages 2–13, 2006.

[43] W. Du and J. Piater. Multi-view tracking using sequential belief propagation. In *Asian Conference on Computer Vision*, pages 684–693, Hyderabad, India, 2006. Springer-Verlag.

[44] A. Enficiaud, B. Lienard, N. Allezard, R. Sebbe, S. Beucher, X. Desurmont, P. Sayd, and J. F. Delaigle. Clovis - a generic framework for general purpose visual surveillance applications. In *IEEE International Workshop on Visual Surveillance (VS)*, 2006.

[45] M. Fayad and D. Schmidt, editors. *Object-Oriented Application Frameworks*, volume 40 of *Special issue of the Communications of ACM*, October 1997.

[46] D. Fensel, V. R Benjamins, E. Motta, and B. Wielinga. UPML: A Framework for Knowledge System Reuse. In *IJCAI-99*, 1999.

[47] J. Ferber. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence.* Addison Wesley, 1st edition edition, 1999.

[48] J. M. Ferryman, S. Maybank, and A. Worrall. Visual surveillance for moving vehicles. *International Journal of Computer Vision*, 37(2):187–197, June 2000.

[49] G. L. Foresti and C. Micheloni. A robust feature tracker for active surveillance of outdoors scenes. *Electronic Letters on Computer Vision and Image Analysis*, 1(1):21–34, 2003.

[50] G. L. Foresti, C. Piciarelli, and L. Snidaro. Trajectory clustering and its applications for video surveillance. In *Advanced Video and Signal-Based Surveillance (AVSS)*, Como, Italy, September 2005.

[51] G. L. Foresti and C. S. Regazzoni. Multisensor data fusion for driving autonomous vehicles in risky environments. *IEEE Transactions on Vehicular Technology*, 51(5):1165 –1185, September 2002.

[52] A. Galata, N. Johnson, and D. Hogg. Learning variable length markov models of behaviour. *Computer Vision and Image Understanding*, 81(3):398–413, 2001.

[53] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison Wesley, 1995.

[54] D. M. Gavrila and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. *International Journal of Computer Vision (est: February-March 2007 issue)*, 2007.

[55] M. Gelgon, P. Bouthemy, and J. P. Le-Cadre. Recovery of the trajectories of multiple moving objects in an image sequence with a pmht approach. *Image and Vision Computing Journal*, 23(1):19–31, 2005.

[56] J. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubezy, H. Eriksson, N. F. Noy, and S. W. Tu. The evolution of Protégé: An environment for knowledge-based systems development. In *SMI-2002*, 2002.

[57] P. Genoud, V. Dupierris, M. Page, C. Bruley, D. Ziébelin, J. Gensel, and D. Bardou. From arom, a new object based knowledge representation system, to webarom, a knowledge bases server. In *Workshop Application of Advanced Information Technologies to Medecine (AAITM), AIMSA 2000*, September 2000.

[58] C. Gomila and F. Meyer. Tracking objects by graph matching of image partition sequences. In *3rd IAPR-TC15 Workshop on Graph-Based Representations in Pattern Recognition*, pages 1–11, Ischia, Italy, May 2001.

[59] N. Gourier, J. Maisonnasse, D. Hall, and J. L. Crowley. Head pose estimation on low resolution images. In *CLEAR 2006*, South Hampton, UK, April 2006.

[60] D. Greenhill, P. Remagnino, and G. A. Jones. *Video Based Surveillance Systems - Computer Vision and Distributed Processing*, chapter VIGILANT: Content-Querying of Video Surveillance Streams, pages 193–204. Kluwer Academic Publishers, 2002.

[61] N. Halbwachs. *Synchronous Programming of Reactive Systems.* Kluwer Academic, 1993.

[62] D. Hall. Automatic parameter regulation for a tracking system with an auto-critical function. In *Proceedings of the IEEE International Workshop on Computer Architecture for Machine Perception (CAMP05)*, pages 39–45, Palermo, Italy, July 2005.

[63] D. Hall, R. Emonet, and J. L. Crowley. An automatic approach for parameter selection in self-adaptive tracking. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, Setubal, Portugal, February 2006.

[64] D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R. B. Fisher, J. Santos-Victor, and J. L. Crowley. Comparison of target detection algorithms using adaptive background models. In *2nd Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 113–120, Beijing, October 2005.

[65] D. Hall, F. Pelisson, O. Riff, and J. L. Crowley. Brand identification using gaussian derivative histograms. *Machine Vision and Applications, in Machine Vision and Applications*, 16(1):41–46, 2004.

[66] D. Harel and R. Marelly. *Come, Let's Play: Scenario-Based Programming Using LSCs and the Play-Engine.* Springer, 1st edition edition, 2003.

[67] T.A. Henzinger, P.-H. Ho, and H. Wong-Toi. HyTech: A Model Checker for Hybrid System. *Software Tools for Technology Transfer*, 1:110–122, 1997.

[68] D. C. Hogg, A. G. Cohn, V. Devin, D. Magee, C. Needham, and P. Santos. Learning about objects and activities. In *24th Leeds Annual Statistical Research (LASR) Workshop*, Leeds, UK, 2005.

[69] G.J. Holzmann. The Spin Model Checker. *IEEE Trans. on Software Engineering*, 23:279–295, 1997.

[70] S. Hongeng, F. Brémond, and R. Nevatia. Bayesian framework for video surveillance application. In *Proceedings of the International Conference on Pattern Recognition (ICPR'00)*, pages 1164–1170, Barcelona, Spain, September 2000.

[71] C. L. Huang, E. L. Chen, and P. C. Chung. Fall detection using modular neural networks and back-projected optical flow. In *The 12th International Conference on Neural Information Processing (ICONIP 2005)*, 2005.

[72] C. L. Huang, E. L. Chen, and P. C. Chung. Using modular neural networks for falling detection. In *The 10th Conference on Artificial Intelligence and Application (TAAI2005)*, 2005.

[73] Y. Ivanov and A. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(8):852–872, 2000.

[74] R. E. Johnson. Frameworks = (Components + Patterns). *CACM*, 10(40):39–42, 1997.

[75] E. M. Clarke Jr, O.Grumberg, and D.Peled. *Model Checking*. MIT Press, 2000.

[76] A. Kale, N. Cuntoor, B. Yegnanarayana, A. N. Rajagoplan, and R. Chellappa. *Using Appearance Matching in Optical and Digital Techniques of Information Security*, chapter Gait-based Human Identification. Springer, 2005.

[77] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. V. Lopez, J. M. Loingtier, and J. Irwin. Aspect-oriented programming. In *Proc. ECCOP 1997*, number 1241 in LNCS. Springer Verlag, June 1997.

[78] Danny B. Lange and Mitsuru Oshima. *Programming and Deploying Java(TM) Mobile Agents with Aglets(TM)*. Addison-Wesley, 1998.

[79] I. Laptev, B. Caputo, and T. Lindeberg. Local velocity-adapted motion events for spatio-temporal recognition. *Computer Vision and Image Understanding*, 2007.

[80] I. Laptev and T. Lindeberg. Velocity-adaptation of spatio-temporal receptive fields for direct recognition of activities: An experimental study. *Image and Vision Computing*, 22:105–116, 2004.

[81] D. Larlus and F. Jurie. Latent mixture vocabularies for object categorization. In *British Machine Vision Conference*, 2006.

[82] S. N. Lim, A. Mittal, L. S. Davis, and N. Paragios. Fast illumination-invariant background subtraction using two views: Error analysis, sensor placement and applications. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, June 2005.

[83] F. Lv, X. Song, B. Wu, V. K. Singh, and R. Nevatia. Left luggage detection using bayesian inference. In *PETS*, 2006.

[84] K.L MacMillan. *The SMV Model Checker*. Available from http:www-cad.eecs.berkeley.edu/∼kenmcmil/smv/, 2001.

[85] W. Mark and D. M. Gavrila. Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):38–50, March 2006.

[86] T. Mathes and J. Piater. Robust non-rigid object tracking using point distribution manifolds. In *28th Annual Symposium of the German Association for Pattern Recognition (DAGM)*, 2006.

[87] S. J. Mellor and M. J. Balcer. *Executable UML: A Foundation for Model-Driven Architecture*. Addison-Wesley, 2002.

[88] C. Micheloni and G. L. Foresti. Fast good feature selection for wide area monitoring. In *IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 271–276, Miami Beach, Florida, July 2003.

[89] A. Mittal and L. Davis. M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, 51(3):189–203, 2003.

[90] T. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, 2006.

[91] S. Moisan and D. Ziébelin. Résolution de problèmes en pilotage de programmes. In *RFIA'2000, 12ème Congrès Reconnaissance des Formes et Intelligence Artificielle*, pages 387–395, Paris, France, février 2000.

[92] S. Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1863–1868, 2006.

[93] D. Nau, M. Ghallab, and P. Travers. *Automated Planning: Theory and Practice*. Morgan Kaufmann, 2004.

[94] B. Neumann and T. Weiss. Navigating through logic-based scene models for high-level scene interpretations. In Springer, editor, *3rd International Conference on Computer Vision Systems - ICVS 2003*, pages 212–222, 2003.

[95] R. Nevatia, J. Hobbs, and B. Bolles. An ontology for video event representation. In *IEEE Workshop on Event Detection and Recognition*, June 2004.

[96] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, December 2005.

[97] T. Ogata, W. Christmas, J. Kittler, and S. Ishikawa. Improving human activity detection by combining multi-dimensional motion descriptors with boosting. In *Proceedings of the 18th International Conference on Pattern Recognition*, volume 1, pages 295–298. IEEE Computer Society, August 2006.

[98] Mourad Oussalah and Collectif. *Ingénierie des composants : Concepts, techniques et outils*. Vuibert, 2005.

[99] V. Parameswaren and R. Chellappa. Human action-recognition using mutual invariants. *Computer Vision and Image Understanding*, 98:295–325, September 2005.

[100] P. Pettersson and K. Larsen. Uppaal2k. *Bulletin of the European Association for Theoretical Computer Science*, 70:40–44, 2000.

[101] C. Pinhanez and A. Bobick. Interval scripts: a programming paradigm for interactive environments and agents. *Pervasive and Ubiquitous Computing*, 7(1):1–21, 2003.

[102] R.Brachman and H. Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann Series in Artificial Intelligence, 2004.

[103] P. Remagnino, A. I. Shihab, and G. A. Jones. Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition, Special Issue on Agent-based Computer Vision*, 37(4):675–689, April 2004.

[104] A. Renouf, R. Clouard, and M. Revenu. Un modèle de formulation d'applications de traitement d'images. In *ORASIS'05*, Mai 2005.

[105] K. Sage and H. Buxton. Joint spatial and temporal structure learning for task based control. In *International Conference on Pattern Recognition*, 2004.

[106] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant colour features. *Computer Vision and Image Understanding*, 95(2):238–259, August 2004.

[107] F.A Schreiber. *Is Time a Real Time? An Overview of Time Ontology in informatics*, volume F127, pages 283–307. Springer Verlag NATO-ASI, 1994.

[108] G. Schreiber, B. Wielinga, R. de Hoog, H. Akkermans, and W. van de Velde. CommonKADS: A Comprehensive Methodology for KBS Development. *IEEE Expert*, 9(6):28–37, 1994.

[109] Y. Sheikh and M. Shah. Bayesian modelling of dynamic scenes for object detection. *IEEE Transactions on PAMI*, October 2005.

[110] Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa. Propagation networks for recognition of partially ordered sequential action. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR 2004)*, Washington DC, June 2004.

[111] J. Sproston. *Validation of Stochastic System*, volume 2925 of *Lecture Notes in Computer Science*, chapter Model Checking for Probabilistic Timed Systems, pages 189–229. Springer Berlin / Heidelberg, 2004.

[112] A. M. Taj, E. Maggio, and A. Cavallaro. Multi-feature graph-based object tracking. In *Classification of Events, Activities and Relationships (CLEAR) Workshop*, Southampton, UK, April 2006. Springer.

[113] R. Urtasun, D. Fleet, and P. Fua. 3d people tracking with gaussian process dynamical models. In *Conference on Computer Vision and Pattern Recognition*, June 2006.

[114] A. Veeraraghavan, A. K. RoyChowdhury, and R. Chellappa. Matching shape sequences in video with applications in human movement analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27, December 2005.

[115] S. A. Velastin, B. A. Boghossian, and M. A. Vicencio-Silva. A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. *Transportation Research Part C: Emerging Technologies*, 14(2):96–113, 2006.

[116] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[117] S. Von-Wum, L. Chen-Yu, Y. Jaw-Jium, and C. Ching-Chi. Using sharable ontology to retrieve historical images. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries*, pages 197–198, 2002.

[118] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3):585–601, 2003.

[119] G. Weiss. *Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence.* MIT Press, 1999.

[120] M. Wooldridge. *An Introduction to MultiAgent Systems.* John Wiley & Sons, 2001.

[121] T. Xiang and S. Gong. Video behaviour profiling and abnormality detection without manual labelling. In *IEEE International Conference on Computer Vision*, pages 1238–1245, Beijing, October 2005.

[122] T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67(1):21–51, 2006.

[123] T. Xiang and S. Gong. Incremental visual behaviour modelling. In *IEEE Visual Surveillance Workshop*, pages 65–72, Graz, May 2006.

[124] T. Xiang and S. Gong. Optimal dynamic graphs for video content analysis. In *British Machine Vision Conference*, volume 1, pages 177–186, Edinburgh, September 2006.

[125] S. Zaidenberg, O. Brdiczka, P. Reignier, and J. L. Crowley. Learning context models for the recognition of scenarios. In *In 3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI) 2006*, Athens, Greece, June 2006.

[126] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM MM*, Santa Barbara, CA, USA, 2006.

[127] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan. Semi-supervised adapted hmms for unusual event detection. In *EEE CVPR*, 2005.

[128] J. Zhang and S. Gong. Beyond static detectors: A bayesian approach to fusing long-term motion with appearance for robust people detection in highly cluttered scenes. In *IEEE Visual Surveillance Workshop*, pages 121–128, Graz, May 2006.

[129] L. Zhang, H. Hermanns, and D. Jansen. Logic and model checking for hidden markov models. In F. Wang, editor, *25th International Conference on Formal Techniques for Networked and Distributed Systems*, volume Lecture Note in Conputer Science, pages 98–112, National Tawain University, October 2005.

[130] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. *Computer Vision and Pattern Recognition*, 2:406–413, 2004.

[131] F. Ziliani, S. Velastin, F. Porikli, L. Marcenaro, T. Kelliher, A. Cavallaro, and P. Bruneaut. Performance evaluation of event detection solutions: the creds experience. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS 2005)*, Como, Italy, September 2005.

9 Bibliography of the Orion Team (from 1998)

Book Chapters

[132] N. Chleq, F. Bremond, and M. Thonnat. *Advanced Video-based Surveillance Systems*, chapter Image Understanding for Prevention of Vandalism in Metro Stations, pages 108–118. Kluwer A.P. , Hangham, MA, USA, C.S. Regazzoni, G. Fabri, and G. Vernazza edition, November 1998.

[133] N. Chleq, C. Regazzoni, A. Teschioni, and M. Thonnat. *Advances in Information Technologies: the business challenge*, chapter A visual surveillance system for the prevention of vandalism in metro stations, pages 287– 294. IOS Press, J.Y. Roger, B. Stanford-Smith and P.T. Kidd edition, 1998.

[134] F. Cupillard, F. Brémond, and M. Thonnat. Tracking groups of people for video surveillance. In P. Remagnino, GA. Jones, N. Paragios, and C. Regazzoni, editors, *Video-Based Surveillance Systems*, pages 89–100. Kluwer Academic, 2002.

[135] J. Malenfant, S. Moisan, and A. Moreira, editors. *Object-Oriented Technology - ECOOP'2000 Workshop Reader*, number 1964 in Lecture Notes in Computer Science, Sophia Antipolis and Cannes, France, juin 2000. Springer.

[136] V. Martin and M. Thonnat. *Scene Reconstruction, Pose Estimation and Tracking*, chapter A Learning Apporach for Adaptive Image Segmentation, pages 431– 454. ISBN 978-3-902613-06-6, Rustam Stolkin edition, 2007.

[137] S. Moisan and J.-L. Ermine. *Ingénierie des connaissances*, chapter Gestion opérationnelle des connaissances sur les codes, pages 297–319. L'Harmattan, 2005.

[138] Monique Thonnat. Knowledge-based techniques for image processing and for image understanding. In A. Heck and F. Murtagh, editors, *Voies nouvelles pour l'Analyse de Donnees en Sciences de l'Univers*, volume 12 of *Journal Phys. IV FRance*, pages 189–236. EDP Sciences, Les Ulis, 2002.

Articles in Referred Journals

[139] A. Avanzi, F. Bremond, C. Tornieri, and M. Thonnat. Design and assessment of an intelligent activity monitoring platform. *EURASIP Journal on Applied Signal Processing, special issue in "Advances in Intelligent Vision Systems: Methods and Applications"*, 2005(14):2359–2374, August 2005.

[140] P. Boissard, C. Hudelot, G. Pérez, P. Béarez, P. Pyrrha, M. Thonnat, and J. Lévy. La détection précoce in situ des bioagresseurs : application à l'oidium du rosier de serre. *Revue Horticole*, septembre 2004.

[141] P. Bonton, A. Boucher, M. Thonnat, R. Tomczak, P.J. Hidalgo, J. Belmonte, and C. Galan. Colour image in 2d and 3d microscopy for the automation of pollen rate measurement. *Image Analysis and Stereology*, 20(Suppl 1):527–532, 2001. also Proc. 8th European Congress for Stereology and Image Analysis (ECSIA).

[142] A. Boucher, P.J. Hidalgo, M. Thonnat, J. Belmonte, C. Galan, P. Bonton, and R. Tomczak. Development of a Semi-Automatic System for Pollen Recognition. *Aerobiologia, International Journal of Aerobiology,*, 2001.

[143] A. Boucher, P.J. Hidalgo, M. Thonnat, J. Belmonte, C. Galan, P. Bonton, and R. Tomczak. Development of a Semi-Automatic System for Pollen Recognition. *Aerobiologia, International Journal of Aerobiology,*, 18(3-4), 2002.

[144] B. Boulay, F. Bremond, and M. Thonnat. Applying 3d human model in a posture recognition system. *Pattern Recognition Letter, Special Issue on vision for Crime Detection and Prevention*, 27(15):1788–1796, 2006.

[145] F. Brémond and M. Thonnat. Issues of representing context illustrated by video-surveillance applications. *International Journal of Human-Computer Studies Special Issue on Context*, 48:375–391, 1998.

[146] F. Brémond and M. Thonnat. Tracking multiple non-rigid objects in video sequences. *IEEE Transaction on Circuits and Systems for Video Technology Journal*, 8(5), September 1998.

[147] F. Bremond, M. Thonnat, and M. Zúniga. Video Understanding Framework For Automatic Behavior Recognition. *Behavior Research Methods Journal*, 38(3):416–426, 2006.

[148] A-M. Dery, M. Blay-Fornarino, and S. Moisan. Apport des interactions pour la distribution des connaissances. *L'OBJET, special issue « Systèmes distribués et Connaissances »*, 8(4), 2002.

[149] Florent Fusier, Valery Valentin, François Bremond, Monique Thonnat, Mark Borg, David Thirde, and James Ferryman. Video understanding for complex activity recognition. *Machine Vision and Applications Journal*, 18:167–188, 2007.

[150] Benoit Georis, François Bremond, and Monique Thonnat. Real-time control of video surveillance systems with program supervision techniques. *Machine Vision and Applications Journal*, 18:189–205, 2007.

[151] N. Khayati, W. Lejouad-Chaari, S. Moisan, and J.P. Rigault. Distributing knowledge-based systems using mobile agents. *WSEAS Transactions on Computers*, 5(1):22–29, January 2006.

[152] N. Maillot and M. Thonnat. Ontology based complex object recognition. *Image and Vision Computing Journal, Special Issue on Cognitive Computer Vision*, 2006. To appear, available on line July 2006.

[153] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology based cognitive vision. *Machine Vision and Applications (MVA)*, 16(1):33–40, December 2004.

[154] M. Marcos, S. Moisan, and A. P. del Pobil. Knowledge modeling of program supervision task and its application to knowledge base verification. *Applied Intelligence*, 10(2/3):181–192, March/April 1999.

[155] G. Medioni, I. Cohen, F. Brémond, S. Hongeng, and G. Nevatia. Activity Analysis in Video. *Pattern Analysis and Machine Intelligence PAMI*, 23(8):873–889, 2001.

[156] S. Moisan. Réutilisation et générateurs de systèmes à base de connaissances : le *framework* BLOCKS. *TSI*, 20(4):529–553, 2001.

[157] S. Moisan, A. Ressouche, and J-P. Rigault. Blocks, a component framework with checking facilities for knowledge-based systems. *Informatica, Special Issue on Component Based Software Development*, 25(4):501–507, November 2001.

[158] C. Shekhar, S. Moisan, R. Vincent, P. Burlina, and R. Chellappa. Knowledge-based control of vision systems. *Image and Vision Computing*, 17(8):667–683, May 1999.

[159] M. Thonnat and S. Moisan. What can Program Supervision do for Software Reuse? *IEE Proceedings - Software Special Issue on Knowledge Modelling for software components reuse*, 147(5), 2000.

[160] J-P. Vidal, S. Moisan, J-B. Faure, and D. Dartus. Towards a reasoned 1D river model calibration. *J. Hydroinformatics*, 7(2):79–90, April 2005.

[161] J.P. Vidal, S. Moisan, and D. Faure, J.B.and Dartus. River model calibration, from guidelines to operational support tools. *Environmental Modelling and Software*, 22(11):1628–1640, November 2007.

[162] T. Vu, F. Brémond, and M. Thonnat. Human behaviour visualisation and simulation for automatic video understandin. *Journal of WSCG*, 10(1-3):485–492, 2002.

Publications in International Referred Conferences

[163] C. André and J-P. Rigault. Variations on the semantics of graphical models for reactive systems. In *SMC'02, ISBN: 2-9512309-4-x*, Hammamet, Tunisia, October 2002. IEEE Press.

[164] Charles André, Marie-Agnès Peraldi-Frati, and Jean-Paul Rigault. Scenario and Property Checking of Real-Time Systems Using a Synchronous Approach. In *4th International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2001)*, Magdeburg, Germany, May 2001. IEEE.

[165] Charles André, Marie-Agnès Peraldi, and Jean-Paul Rigault. Integrating the synchronous paradigm into UML: Application to control-dominated systems. In *5th UML Conference («UML 2002»*, number 2460 in LNCS, Dresden, Germany, 2002. Springer Verlag.

[166] A. Avanzi, F. Brémond, and M. Thonnat. Tracking multiple individuals for video communication. In *International Conference on Image Processing ICIP'01*, Thessaloniki, Greece, October 2001. IEEE Signal Processing Society.

[167] J. Bannour, B. Georis, F. Bremond, and M. Thonnat. Generation of 3d animations from scenario models. In *Proceedings of VIIP'04- 4th IASTED International Conference on Visualization, Imaging and Image Processing*, September 6-8th, 2004.

[168] H. Benhadda, J.L. Patino, E. Corvee, F. Bremond, and M. Thonnat. Data mining on large video recordings. In *Veille Strategique Scientifique et Technologique VSST 2007*, Marrakech, Morocco, 21st -25th October 2007.

[169] P. Boissard, C. Hudelot, M. Thonnat, G. Perez, P. Pyrrha, and F. Bertaux. Automated approach to monitor the sanitary status and to detect early biological attacks on plants in greenhouse. In IEEE, editor, *International Symposium on Greenhouse Tomato, Avignon*. CTIFL, 2003.

[170] Mark Borg, David Thirde, James Ferryman, Florent Fusier, Francois Bremond, and Monique Thonnat. An integrated vision system for aircraft activity monitoring. In *Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, Breckenridge, Colorado, USA, January 2005. IEEE Computer Society.

[171] Mark Borg, David Thirde, James Ferryman, Florent Fusier, Valery Valentin, Francois Bremond, and Monique Thonnat. Video surveillance for aircraft activity monitoring. In *Proceedings of IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Como, Italy, September 2005. IEEE Computer Society.

[172] Mark Borg, David Thirde, James Ferryman, Florent Fusier, Valery Valentin, Francois Bremond, Monique Thonnat, Josep Aguilera, and Martin Kampel. Visual surveillance for aircraft activity monitoring. In *Proceedings of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, Beijing, China, October 2005. IEEE Computer Society.

[173] A. Boucher, P. Hidalgo, M. Thonnat, J. Belmote, and C. Galan. 3D Pollen Image Recognition based on Palynological Knowledge. In *Second European Symposium on Aerobiology*, Vienne, Autriche, September 2000.

[174] A. Boucher and M. Thonnat. Object recognition from 3d blurred images. In *International Conference on Pattern Recognition, ICPR2002*, Quebec, Canada, August 2002.

[175] A. Boucher, P. Tomczack, P. Hidalgo, M. Thonnat, J. Belmote, C. Galan, and P. Bonton. A Semi-Automatic System for Pollen Recognition. In *Second European Symposium on Aerobiology*, Vienne, Autriche, September 2000.

[176] B. Boulay, F. Bremond, and M. Thonnat. Human posture recognition in video sequence. In *Proceedings Joint IEEE International Workshop on VS-PETS, Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 23–29, OCT 11-12 2003.

[177] Bernard Boulay, Francois Bremond, and Monique Thonnat. Posture recognition with a 3d human model. In *Proceedings of IEE International Symposium on Imaging for Crime Detection and Prevention*, London, UK, 2005. Institution of Electrical Engineers.

[178] Binh Bui, Francois Brémond, Monique Thonnat, and JC. Faure. Shape recognition based on a video and multi-sensor system. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2005)*, Como, Italie, September 2005.

[179] C. Carincotte, X. Desurmont, B. Ravera, F. Brémond, J. Orwell, S. Velastin, J-M. Odobez, B. Corbucci, J. Palo, and J. Cernocky. Toward generic intelligent knowledge extraction from video and audio: the eu-funded caretaker project. In *IET conference on Imaging for Crime Detection and Prevention (ICDP 2006)*, London, Great Britain, June 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[180] J. Cavarroc, S. Moisan, and J-P. Rigault. Simplifying an Extensible Class Libray Interface with OpenC++. In *OOPSLA'98, Workshop on Reflective Programming in C++ and Java (UTCCP Report 98-4)*. University of Tsukuba, October 1998.

[181] M. Crubézy, M. Marcos, and S. Moisan. Experiments in Building Program Supervision Engines from Reusable Components. In *13th European Conference on Artificial Intelligence-ECAI98. Workshop on Applications of Ontologies and Problem-Solving Methods.* John Wiley & Sons, Ltd., August 1998.

[182] F. Cupillard, A. Avanzi F. Brémond, and M. Thonnat. Video understanding for metro surveillance. In *IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems*, Tapei(Taiwan), March 2004.

[183] F. Cupillard, F. Brémond, and M. Thonnat. Tracking groups of people for video-surveillance. In *2nd European Workshop n Advanced Video-based surveillance Systems AVBS'01*, Kingston, UK, September 2001.

[184] F. Cupillard, F. Brémond, and M. Thonnat. Group behavior recognition with multiple cameras. In *IEEE Proceedings of the Workshop on Applications of Computer Vision - ACV2002*, Orlando, USA, December 2002.

[185] F. Cupillard, F. Brémond, and M. Thonnat. Behaviour recognition for individuals, groups of people and crowd. In *IEE Proceedings of the Intelligent Distributed Surveillance Systems Symposium - IDSS2003*, London, UK, February 2003.

[186] Frederic Cupillard, Francois Brémond, and Monique Thonnat. Automatic visual recognition for metro surveillance. In *Measuring Behavior*, Wageningen The Netherlands, September 2005.

[187] S. Delaitre, A. Giboin, and S. Moisan. The AEX Method and its Instrumentation. In Kluwer, editor, *ICEIS, International Conference on Enterprise Information Systems*, Ciudad Real - Spain, April 2002.

[188] S. Delaître and S. Moisan. Knowledge Management by Reusing Experience. In Rose Dieng and Olivier Corby, editor, *EKAW'2000, 12th International Conference on Knowledge Engineering and Knowledge Management*, Lecture Notes in Artificial Intelligence 1935, pages 304–311, Juan les Pins, France, octobre 2000. Springer.

[189] M. Dery, A-M.and Blay-Fornarino, S. Moisan, B. Arcier, and L. Mule. Distributed access knowledge-based system: Reified interaction service for trace and control. In *3rd International Symposium on Distributed Objects & Applications, DOA'01*, pages 76–84, Rome, Italy, September 2001.

[190] N. Dey, A. Boucher, and M. Thonnat. Image formation model of a 3d translucent object observed in light microscopy. In *International Conference on Image Processing ICIP2002*, Rochester, USA, September 2002.

[191] James Ferryman, Mark Borg, David Thirde, Florent Fusier, Valery Valentin, Francois Bremond, Monique Thonnat, Josep Aguilera, and Martin Kampel. Automated scene understanding for airport aprons. In *Proceedings of 18th Australian Joint Conference on Artificial Intelligence*, Sidney, Australia, December 2005. Springer-Verlag.

[192] B. Georis, F. Bremond, M. Thonnat, and B. Macq. Use of an evaluation and diagnosis method to improve tracking performances. In *Proceedings of VIIP'03- 3rd IASTED International Conference on Visualization, Imaging and Image Processing*, Benalmadera, SPAIN, SEP 8-10 2003.

[193] B. Georis, M. Maziere, F. Bremond, and M. Thonnat. A video interpretation platform applied to bank agency monitoring. In *Proceedings of IDSS'04 - 2nd Workshop on Intelligent Distributed Surveillance Systems*, February 23th, 2004.

[194] B. Georis, M. Mazière, F. Brémond, and M. Thonnat. Evaluation and knowledge representation formalisms to improve video understanding. In *Proceedings of the*

International Conference on Computer Vision Systems (ICVS'06), New-York, NY, USA, January 2006.

[195] F. Hongeng, S. Brémond and R. Nevatia. Bayesian Framework for Video Surveillance Application. In *15th International Conference on Pattern Recognition*, Barcelone, Espagne , September 2000.

[196] F. Hongeng, S. Brémond and R. Nevatia. Representation and Optimal Recognition of Human Activities. In *IEEE Conference on Computer Vision and Pattern Recognition*, USA , June 2000.

[197] C. Hudelot and M. Thonnat. An architecture for knowledge based image interpretation. In *Workshop on Computer Vision System Control Architectures, VSCA, Graz*, 2003.

[198] C. Hudelot and M. Thonnat. A cognitive vision platform for automatic recognition of natural complex objects. In IEEE, editor, *International Conference on Tools with Artificial Intelligence, ICTAI, Sacramento*, 2003.

[199] M. B. Kaâniche, F. Brémond, and M. Thonnat. Monitoring trichogramma activities from videos : An adaptation of cognitive vision system to biology field. In *International Cognitive Vision Workshop, (ICVW'2006) in conjunction with the 9th European Conference on computer Vision (ECCV'2006)*, Graz, Austria, 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICVW06.html.

[200] Naoufel Khayati, Wided Lejouad-Chaari, and Sabine Moisan. Systèmes distribués de pilotage de programmes : application à l'imagerie médicale. In *Proc. Méthodes et Heuristiques pour l'Optimisation des Systèmes industriels (MHOSI'2005)*, Hammamet, Tunisie, April 2005.

[201] Naoufel Khayati, Wided Lejouad-Chaari, Sabine Moisan, and Jean-Paul Rigault. A mobile agent-based architecture for distributed program supervision systems. In *Proc. 4th WSEAS Int. Conf. on Information Security, Communications and Computers*, Tenerife, Spain, December 2005.

[202] A. Lux. Tools for automatic interface generation in scheme 2nd. In *Workshop on Scheme and Functional Programming, Colloquium principles, Logics, and Implementations of high-level programming languages"*, Florence, Italy, September 2001.

[203] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology based cognitive vision. In James L. Crowley, Justus H. Piater, Markus Vincze, and Lucas Paletta, editors, *Computer Vision Systems, Third International Conference, ICVS*, volume 2626 of *Lecture Notes in Computer Science*. Springer, 2003.

[204] N. Maillot, M. Thonnat, and A. Boucher. A visual concept ontology for automatic image understanding. In *Poster Session of the 2nd International Semantic Web Conference, ISWC*, 2003.

[205] Nicolas Maillot and Monique Thonnat. A weakly supervised approach for semantic image indexing and retrieval. In *International Conference on Image and Video Retrieval, CIVR*, pages 629–638, 2005.

[206] Nicolas Maillot, Monique Thonnat, and Celine Hudelot. Ontology based object learning and recognition : Application to image retrieval. In *Proceedings of 16th IEEE International Conference on Tools For Artificial Intelligence*, Boca Raton, USA, 2004. IEEE Computer Society Press.

[207] M. Marcos, S. Moisan, and A. Del Pobil. Knowledge modeling of program supervision task. In Jose Mira, Angel Pasqual del Pobil, and Moonis Ali, editors, *Proceeding of the 11th International Conference on Industrial Applications of Artificial Intelligence and Expert Systems*, Lecture Notes in Computer Science no 1415, Benicassim, Spain, June 1998. Springer-Verlag.

[208] Vincent Martin, Nicolas Maillot, and Monique Thonnat. A learning approach for adaptive image segmentation. In *Proceedings of the International Conference on Computer Vision Systems (ICVS'06)*, New-York, NY, USA, January 2006.

[209] N. Moenne-Loccoz, F. Brémond, and M. Thonnat. Recurrent bayesian network for the recognation of human behaviors from video. In *Third International Conference On Computer Vision Systems (ICVS 2003)*, volume LNCS 2626, pages 44–53, Graz, Austria, 2003. Springer.

[210] S. Moisan. Knowledge representation for program reuse. In *Proc. European Conference of Artificial intelligence (ECAI 2002)*, Lyon, France, July 2002.

[211] S Moisan, M. Boër, C. Thiebaut, F. Tricoire, and M. Thonnat. A versatile scheduler for automatic telescopes. In *SPIE Conference on Astronomical Telescopes and Instrumentation*, Hawaii, USA, August 2002.

[212] S. Moisan and W. Lejouad-Chaari. Réutilisation intelligente de programmes de vision en environnement distribué. In *2émes Ateliers Traitement et Analyse d'Images Méthodes et Applications, TAIMA01*, Hammamet, Tunisia, October 2001.

[213] S. Moisan, A. Ressouche, and J.-P. Rigault. Towards Formalizing Behavorial Substitutability in Component Frameworks. In *2nd International Conference on Software Engineering and Formal Methods*, pages 122–131, Beijing, China, 2004. IEEE Computer Society Press.

[214] S. Moisan, A. Ressouche, and J-P. Rigault. Synchronous Formalism and Behavioral Substitutability in Component Frameworks. In F. Maraninchi, editor, *Synchronous Languages Applications and Programming*, ENCS. Elsevier Science, 2005.

[215] S. Moisan and M. Thonnat. Program Supervision under Resource Constraints. In *13th European Conference on Artificial Intelligence-ECAI98. Workshop on Monitoring and Control of Real-Time Intelligent Systems.* John Wiley & Sons, Ltd., August 1998.

[216] Sabine Moisan, Annie Ressouche, and Jean-Paul Rigault. Behavioral substitutability in component frameworks: a formal approach. In *Proc. Specification and Verification of Component-based Systems (SAVCBS'2003)*, pages 22–28, Helsinki, Finland, 2003. IEEE Computer Society Press.

[217] A. T. Nghiem, F. Bremond, M. Thonnat, and R. Ma. A new evaluation approach for video processing algorithm. In *Proceedings of WMVC 2007 IEEE Workshop on Motion and Video Computing*, Austin, Texas, USA, February 2007.

[218] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin. Etiseo: an evaluation project for video surveillance systems. In *Proceedings of AVSS2007, IEEE International Conference on Advanced Video and Singal based Surveillance*, London, UK, September 2007.

[219] J.L. Patino, H. Benhadda, E. Corvee, F. Bremond, and M. Thonnat. Video-data modelling and discovery. In *International Conference on Visual Information Engineering VIE 2007*, London, UK, 25th -27th July 2007.

[220] J.L. Patino, E. Corvee, F. Bremond, and M. Thonnat. Management of large video recordings. In *Ambient Intelligence Developments AmI.d 2007*, Sophia Antipolis, France, 17th -19th September 2007.

[221] Pascal Rapicault and Jean-Paul Rigault. Open implementation of UML meta-model(s): Making meta-modeling and meta-programming meet. In *Reflection 2001*, number 2192 in LNCS, pages 276–277, Kyoto, Japan, 2001. Springer Verlag.

[222] Pascal Rapicault, Jean-Paul Rigault, and Luc Bourlier. Model, notation, and tools for verification of protocol-based components assembly. In Judith Bishop, editor, *Component Deployment*, LNCS 2370, pages 257–268, Berlin, Germany, 2002. Springer-Verlag.

[223] A. Ressouche, V. Roy, J.Y. Tigli, and D. Cheung. Sas architecture: Verification oriented formal modeling of concrete critical systems. In *2003 IEEE International Conference on Systems, Man and Cybernetics, SMC'03*, Washington, USA, October 2003.

[224] N. Rota, R. Stahr, and M. Thonnat. Tracking for Visual Surveillance in VSIS. In *First IEEE International Workshop on Performance Evaluation of Tracking and Surveillance PETS2000*, Grenoble, France, March 2000.

[225] N. Rota and M. Thonnat. Activity Recognition from Video Sequences using Declarative Models. In *14th European Conference on Artificial Intelligence. ECAI2000.*, Berlin, Allemagne, August 2000.

[226] N. Rota and M. Thonnat. Video Sequence Interpretation for Visual Surveillance. In *Workshop on Visual Surveillance. VS'00.*, pages 59–67, Dublin,Irlande , July 2000.

[227] Lan Le Thi, Alain Boucher, and Monique Thonnat. Subtrajectory-based video indexing and retrieval. In *13th International MultiMedia Modeling Conference (MMM)*, Singapore, 2007.

[228] David Thirde, Mark Borg, James Ferryman, Florent Fusier, Valery Valentin, Francois Bremond, and Monique Thonnat. Video event recognition for aircraft activity monitoring. In *Proceedings of 8th IEEE International Conference on Intelligent Transportation Systems*, Vienna, Austria, September 2005. IEEE Computer Society.

[229] M. Thonnat, A. Avanzi, and F. Brémond. Analyse de scènes de bureaux pour la communication video. In *2èmes Ateliers Traitement et Analyse d'Images Méthodes et Applications, TAIMA01*, Hammamet, Tunisia, October 2001.

[230] M. Thonnat and S. Moisan. What can program supervision do for program re-use? In Angel Pasqual del Pobil, Jose Mira, and Moonis Ali, editors, *Proceeding of the 11th International Conference on Industrial Applications of Artificial Intelligence and Expert Systems*, Lecture Notes in Computer Science no 1416, Benicassim, Spain, June 1998. Springer-Verlag.

[231] M. Thonnat, S. Moisan, and M. Crubézy. Experience in Integrating Image Processing Programs. In *Internatianal Conference on Computer Vision Systems*, Las Palmas, Canary Islands, January 1999.

[232] M. Thonnat and N. Rota. Image Understanding for Visual Surveillance Applications. In *Third International Workshop on Cooperative Distributed Vision*, Kyoto, Japan, November 1999.

[233] A. Toshev, F. Brémond, and M. Thonnat. An a priori-based method for frequent composite event discovery in videos. In *Proceedings of 2006 IEEE International Conference on Computer Vision Systems*, New York USA, January 2006.

[234] J.-P. Vidal, J.-B. Faure, S. Moisan, and D. Dartus. Decision Support System for Calibration of 1-D River Models. In S. Y. Liong, K.-K. Phoon, and Vladan Babovic, editors, *6th International Conference on Hydroinformatics*, volume 2, pages 1067–1074, Singapore, June 2004. World Scientific Publishing.

[235] J.-P. Vidal and S. Moisan. Opérationalisation de connaissances pour le calage de modèles numériques – Application en hydraulique fluviale. In *IC'2004, 15èmes Journées Francophones d'Ingénierie des Connaissances*, pages 261–272, Lyon, France, 2004. Presses Universitaires de Grenoble.

[236] J-P. Vidal, S. Moisan, and J-B. Faure. Knowledge-based hydraulic model calibration. In L. Jain V. Pallade, R.J. Howlett, editor, *Knowledge-Based Intelligent Information and Engineering Systems (KES)*, volume I of *LNAI*, pages 676–683, Oxford, UK, September 2003. Springer.

[237] T. Vu, F. Brémond, G. Davini, M. Thonnat, QC. Pham, N. Allezard, P. Sayd, JL. Rouas, S. Ambellouis, and A. Flancquart. Audio video event recognition system for public transport security. In *IET conference on Imaging for Crime Detection and Prevention (ICDP 2006)*, London, Great Britain, June 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[238] V-T. Vu, F. Brémond, and M. Thonnat. Automatic video interpretation: A novel algorithm for temporal scenario recognition. In *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, 2003.

[239] V-T. Vu, F. Brémond, and M. Thonnat. Automatic video interpretation: A recognition algorithm for temporal scenarios based on pre-compiled scenario models. In *The 3rd International Conference on Vision System (ICVS'03)*, Graz, Austria, 2003.

[240] V.T. Vu, F. Brémond, and M. Thonnat. Human behaviour visualisation and simulation for automatic video understanding. In *10-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002 WSCG 2002*, Plzen, Czech Republic, February 2002.

[241] V.T. Vu, F. Brémond, and M. Thonnat. Temporal constraints for video interpretation. In *15th European Conference on Artificial Intelligence, Workshop on Modelling and Solving Problems with Constraints*, Lyon, FRANCE, July 2002.

[242] V.T. Vu, F. Brémond, and M. Thonnat. Video surveillance: human behaviour representation and on-line recognition. In *Sixth International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES2002*, Crema, Italy, September 2002.

[243] M. Zhan, P. Remagnino, S. Velastin, F. Brémond, and M. Thonnat. Matching gradient descriptors with topological constraints to characterise the crowd dynamics. In *The 3rd International Conference on Visual Information Engineering (VIE2006)*, pages 441–446, Bangalore, India, September 26-28 2006.

[244] N. Zouba, F. Bremond, M. Thonnat, and V. T. Vu. Multi-sensors analysis for everyday elderly activity monitoring. In *Proceedings of SETIT 2007 the 4th International Conference Sciences of Electronics, Technologies of Information and Telecommunications*, Tunis, Tunisia, March 2007.

[245] M. Zúniga, F. Brémond, and M. Thonnat. Fast and reliable object classification in video based on a 3d generic model. In *The 3rd International Conference on Visual Information Engineering (VIE2006)*, pages 433–441, Bangalore, India, September 26-28 2006.

HDR and PhD Theses

[246] Bernard Boulay. *Human Posture Recognition for behavior Understanding.* PhD thesis, Nice-Sophia Antipolis University, January 2007.

[247] Francois Bremond. *Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition.* Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, juillet 2007.

[248] Monica Crubézy. *Pilotage de programmes pour le traitement d'images médicales.* PhD thesis, Université de Nice-Sophia Antipolis, February 1999.

[249] Nicolas Dey. *Etude de la formation de l'image d'un objet microscopique 3D translucide - Application à la microscopie optique.* PhD thesis, Université de Nice-Sophia Antipolis, December 2002.

[250] Benoit Georis. *Program Supervision Techniques for Easy Configuration of Video Understanding Systems.* PhD thesis, Louvain Catholic University, January 2006.

[251] Céline Hudelot. *Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms.* PhD thesis, Nice-Sophia Antipolis University, 2005.

[252] Nicolas Maillot. *Ontology Based Object Learning and Recogntion.* PhD thesis, Nice-Sophia Antipolis University, December 2005.

[253] Maria Marcos. *Verification and Validation of Program Supervision Systems.* PhD thesis, université de Jaume I (Castellón, Espagne), February 1999.

[254] Sabine Moisan. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances.* Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, avril 1998.

[255] Nathanael Rota. *Contribution à la reconnaissance de comportements humains à partir de séquences videos.* PhD thesis, Université de Nice-Sophia Antipolis, October 2001.

[256] Monique Thonnat. *Vers une vision cognitive: mise en oeuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images.* Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, octobre 2003.

[257] Jean-Philippe Vidal. *Assistance au calage de modèles numériques en hydraulique fluviale : apports de l'intelligence artificielle.* PhD thesis, INP Toulouse, March 2005.

[258] Van-Thinh Vu. *Temporal Scenario for Automatic Video Interpretation*. PhD thesis, Nice-Sophia Antipolis University, 2004.

A The LAMA Platform

LAMA is a software platform for knowledge-based system design. It is written in Java and C++ and contains toolkits to build and to adapt all the software elements that compose a knowledge-based system, such as knowledge bases, inference engines and verification tools.

A.1 Motivation and Approach

A knowledge-based system (KBS) is a *program* that performs complex processing, relying on declarative knowledge. Its architecture is composed of three main parts: a knowledge base storing expert knowledge on a particular domain, a fact base containing facts about a end-user problem in this domain, and an engine, written by a software designer, that performs inferences to solve the end-user problem based on the expert knowledge. From a software engineering point of view, building and maintaining a KBS is a difficult task because there is a huge semantic gap between the contents of the expert's mind and the capabilities of regular programming languages on which the implementation relies. The lack of software engineering support is especially noticeable when reenginering existing systems to cope with evolutions and adaptations. Designing a knowledge-based system implies to develop and to maintain several software codes, which are the ones necessary for running the KBS itself (i.e. inference engine and knowledge manager). It also implies to develop all the utilities for both experts and end-users to communicate with the system; for instance, knowledge editors and verificators are needed by the experts during the knowledge base development. These developments may be long, complex and error-prone. The notion of KBS *generators* emerged in the 80's to share, to a certain extend, a panel of common elements from which a KBS can be designed, such as an inference engine, a knowledge representation manager, verification tools, and various editors. A generator may yield many systems across many application domains (e.g. classification of pollens, of galaxies, of insects, of flowers...). However, throughout their rather long life time (typically more than 10 years), reconfigurations of generator elements are unavoidable.

A.2 LAMA Contents

To support these changes, we propose a "meta generator" approach to go one step further and to generate or adapt elements proposed by generators, i.e. inference engines, interfaces, knowledge base description languages, knowledge verification tools, etc. This approach is implemented in a software platform, named LAMA, (see figure 8) which gathers several generic extensible toolkits to design, test, and modify these software elements. The platform relies on up-to-date software engineering approaches, mainly reusable components and frameworks. For instance, in order to (re)engineer engines, we propose a C++ component framework (library and usage configuration facilities) named BLOCKS. The platform also includes a C++ library for verifying knowledge bases, composed of functions to perform basic knowledge verifications on usual knowledge representation schemes (structured objects, or frames, and rules). Next, a framework for parsers and compilers of knowledge description languages allows to create expert-oriented languages and to translate them into executable (C++) code. Finally, to customize graphic interfaces, a Java framework supports the development of editors for creating and browsing knowledge bases.

A typical example of LAMA toolkit is the BLOCKS framework that implements generic data structures and methods for designing knowledge-based system engines without extensive code rewriting. It is rooted in our extensive experience of the design of various KBS generators for computer aided design, classification, or planning, and in domains as different as civil engineering, astronomy, medicine or biology. Briefly, components in BLOCKS correspond to interrelated classes, and more precisely to the first levels of class hierarchies. The structure of these (abstract) classes form patterns for describing the concepts involved in an engine algorithm; generic functions or (abstract) methods of these classes constitute a kernel of basic instructions that can be redefined to implement a reasoning strategy. Designers can thus extend both the set of concepts and the algorithmic capabilities. Blocks provides a commom layer that implements generic and abstract features useful for a large range of knowledge-based systems. Specific layers can be specialised for application purposes. The common layer supplies concrete and abstract classes that can be derived or composed, relationships among classes that can be extended, generic functions that can be parametrized by other functions, and (abstract) methods that can be redefined. It also provides a global organization of the classes, relationships and functions that must be respected by designers when developing a specialisation for a particular application. Respecting these constraints, that is enforcing the correct use of components, is an important issue addressed by model-checking techniques. Specialisations mainly lead to concrete classes, the instances of which will populate the knowledge bases, and methods or functions that will constitute reasoning steps in the algorithm of an engine strategy.

The general layer of Blocks consists of about 75 classes and 5,000 lines of C++ code, non including comments and utility classes that implement usual data structures (lists, sets, maps...). Except for these utilities, the remaining classes implement common concepts that can be found in KBSs. For instance, the classical artificial intelligence notions of *frame* and of inference *rule* are implemented as classes in the framework. Class interfaces are complete enough to cover most designer needs without modifications, but some points of flexibility (*hooks*) have been foreseen, in particular in methods. Specialization, composition and hooks allow designers to fine tune an engine behavior.

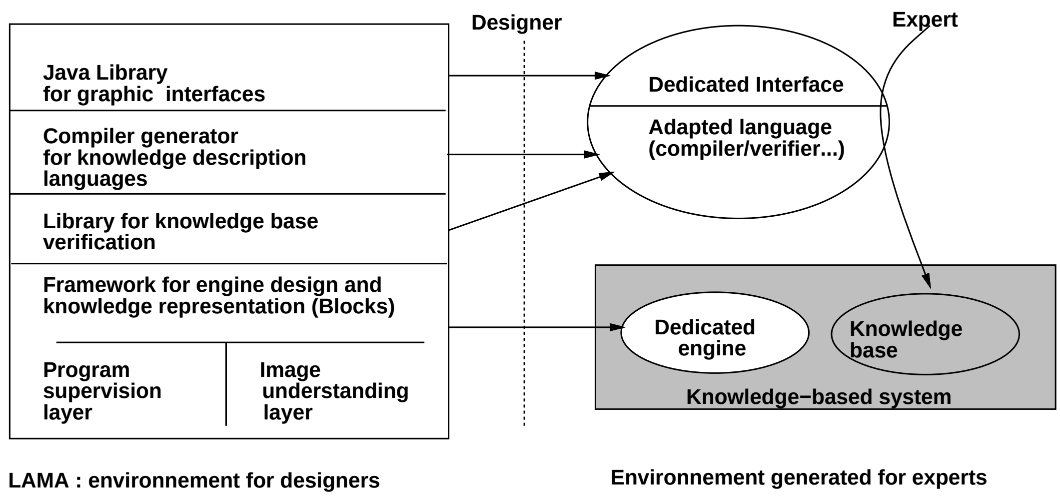


Figure 8: Lama architecture and tools for engine design, knowledge base description, verification, and visualization

A.3 Lama Status

We have used Lama to design several generators and thus several corresponding inference engines: three variants of planning engines (used e.g. at ENSI Tunis), two classification engines (used at INRA Sophia), one model calibration engine (used at Cemagref Lyon). A typical engine code introduces around 15 derived classes (specialised from BLOCKS ones), and 2 to 5 new ones. The number of extra code lines to write for the engine algorithm itself ranges from 250 to 800, which is a tractable size when it comes to modifications. This depends of course on the more or less sophisticated mechanisms that are to be implemented in the engine (e.g. an engine with history management and backtracking mechanism necessitates more code lines than an engine that does not need to memorise its past actions).

B The VSIP Platform

VSIP, VideoSurveillance Intelligent Platform, is a software platform written in C and C++ for video sequence analysis.

B.1 Motivation and Approach

There is an increasing demand in the safety and security domains for intelligent video-surveillance systems. These systems are complicated to build from scratch and difficult to configurate due to the large variety of scenes to process and to the numbers of processing steps involved in video sequence analysis. Most of these videosurveillance systems share common functionalities (as motion detection or object tracking). We thus propose a platform, named VSIP, to help developers building videosurveillance systems.

This platform is composed of reusable components which can be combined for different applications. VSIP mainly contains: (1) image processing algorithms in charge of mobile object detection, classification and frame to frame tracking; (2) multi-camera algorithms for the spatial and temporal analysis (4D) of the detected mobile objects; (3) high level scenario recognition algorithms.

Figure 9 shows the global structure of a videosurveillance system built with VSIP. First, a motion detection step followed by a frame to frame tracking is made for each video camera. Then the tracked mobile objects coming from different video cameras with overlapping fields of view are fused into a unique 4D representation for the whole scene. Depending on the chosen application, a combination of one or more of the available trackers (individuals, groups and crowd tracker) is used. Then scenario recognition is performed by a combination of one or more of the available recognition algorithms (automaton based, Bayesian-network based, AND/OR tree based and temporal constraints based). Finally the system generates the alerts corresponding to the predefined recognized scenarios.

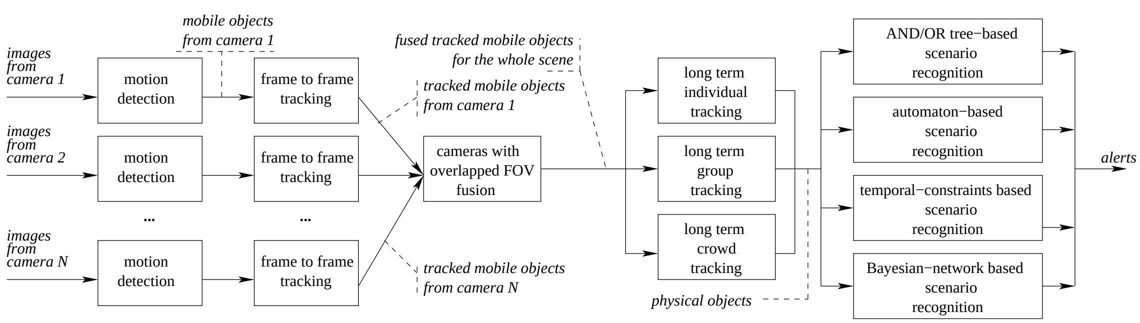


Figure 9: *Global structure of an activity monitoring system built with VSIP.*

B.2 Detailed Contents

The VSIP platform currently contains 26 modules for real-time video sequence analysis and a set of 19 tools for off-line processing and for building an application in videosurveillance.

The real-time modules are: Video Acquisition, Context Loading, Background Initialisation, Changes Adaptation, Door Detection, Image Segmentation, 2D Classification, 3D Classification, Blob Merging, Blob Separation, Blob Position Correction, Blob Filtering, Blob to Mobile Conversion, Frame to Frame Tracking, Ghost Removal, Person Detection, Noise Tracking, Noise Features, Background Updating, Camera Synchronisation, Fusion

Tracking, Long-term Tracking, Global Fusion of Distant Cameras, Global Tracking, Event Recognition based on Finite State Automata, and Temporal Event Recognition.

The off-line tools are: Datapool Manager, Application Controller, Makefiles, Video Recording, Video Header Insertion, Video Conversion, Module Parameters Tuning, Planar and Radially Distorded Camera Calibration, Scene Context Visualization, Performance Evaluation, Histogram Creation, 3D Scene Visualisation, Video Visualisation, Temporal Scenario Parser, Temporal Coherence Checking for Scenario, Trajectory Clustering, People and Event Statistics Calculation, Event Graphs Browser, and Unsupervised Scenario Learning.

B.3 VSIP status

The modules have been registered at APP (the French agency for patrinomy protection) and transfered to Keeneo (http://www.keeneo.com) in 2005 for their industrialisation. Since that date these modules have slightly evolved in particular with the adding of a 3D classification module. The main recent changes concern the tools for off-line analysis (e.g. Trajectory Clustering and Unsupervised Scenario Learning).

C Performances of Orion Videosurveillance Systems

C.1 Introduction

In the three last years Orion has obtained concrete results on intelligent videosurveillance systems [139]. In this annex we present a summary of the performances measured on real world problems through several challenging evaluations led by industrials and end-users in the framework of the European project ADVISOR and two industrial projects Cassiopee and Videa.

C.2 Evaluation Results

C.2.1 European Project ADVISOR

The objective of this European project was to show the feasability of intelligent video-surveillance in metro stations. For more details see the ADVISOR site http://www-sop.inria.fr/orion/ADVISOR. The end-users (i.e. the metro security operators) have defined five main types of activities of interest involving on person and a static object (jumping over a barrier and vandalism against a ticket vending machine), several interacting persons (fighting people), a group of persons and the physical environment (a group blocking a subway access) and a crowd (overcrowding of the metro platform).

The validation of our scenario recognition technique has been performed on **videos taken with the operational CCTV visual surveillance cameras** of two metro stations in Barcelona (Spain) and in Brussels (Belgium) with either real passengers or actors of dramatic schools acting scenarios defined by the end-users. Table 1 shows a summary of the results. One can see that the average true positive rate is really high 90% and that the false alarm rate is low i.e. 1.2%. Our approach analyses the activities in 4D and not based on 2D data. Another measure (i.e. the accurary) is given, it shows the accuracy in time of the recognition (that means what percentage of the duration of the shown activity is "covered" by the generation of the corresponding alert by the system. This value varies a lot with respect to the type of activity. In particular for fighting people, it is still very difficult to detect a precise beginning and ending of this type of behavior.

Activity type	Nb of activities	True positives	Accuracy	False alarms
fighting	21	95 %	61 %	0 (over 16 hours)
blocking	9	78 %	60 %	1 (over 16 hours)
vandalism	2	100 %	71 %	0 (over 16 hours)
jumping o.t.b.	42	88 %	100 %	0 (over 16 hours)
overcrowding	7	100 %	80 %	0 (over 16 hours)
TOTAL	81	90 %	85 %	1 (over 16 hours)

Table 1: *This table shows the results of the technical validation of our intelligent video-surveillance system in subways performed during the European project ADVISOR. For each scenario, we report in particular the percentage of recognized behaviors of this scenario (third column) and the accuracy in time of the recognition (that means what percentage of the duration of the shown behavior is "covered" by the generation of the corresponding alert by the system; this value is an average over all the scenario instances (fourth column).*

C.2.2 Industrial Project Cassiopee

The objective of this industrial project was to show the feasabilty of intelligent surveillance in bank agencies [139]. In this project a first set of activities of bank attack inside the public area of the bank agency have been defined by the security managers. As these events are rare the evaluation has been made on videos acted by bank agency employees. The videos were taken on real environments in three bank agencies to test different configurations (different sizes, geometries and illuminations). A second type of actitivity has been defined by the bank agency security managers for unauthorized access to secure zone (acces limited to one employee). In this case it has been possible to evaluate our approach **in live during a total of 10 days**. The results (see table 2) show that the false alarm rate is again very low (only 2 false alarms during 10 days) due to the 4D analysis of videos taken by synchronized cameras.

Activity type	Nb of activites	True positives	False negatives	False alarms
bank attack with 2 persons	10	10 (100 %)	0 (0 %)	0 (over 1 hour)
bank attack with 3 persons	16	15 (93.75 %)	1 (6.25 %)	0 (over 1 hour)
unauthorized access to secure zone	20 (over 10 days)	20 (100 %)	0 (0 %)	2 (over 10 days)

Table 2: *Validation results for a live installation of the bank agency monitoring system (industrial project Cassiopee).*

C.2.3 Industrial Project Videa

The objective of this cooperation was to use our videosurveillance platform VSIP to build two products: one for access control of a buiding and one for urban violence detection.

The functionality access control was defined as the count of the persons in the lock-chamber of a building and as the determination of where they are coming from and where they are going to (origin and destination) using only one video camera. The main difficulties were to locate precisely the persons because the entrances of the building have transparent doors and to estimate presicely the number of persons because lock-chambers are very narrow. The evaluation has been made **on a real site** with the existing video cameras. The results (shown in table 3) are very high between 92.9 % and 96 % of true positive rates and from 0% to 2% of false alarm rates.

The functionality urban violence has been defined as a violence behavior between several persons in a group in a street or in an open area. The evaluation has been made **on two real sites** in Namur in Belgium with data taken by the existing videosurveillance cameras. The bahaviors have been played by professionnal actors under the control of police officers. The main difficulties are the fact that the activity involves a group of persons (two or more) and that it happens outside in a busy environment along with the normal vehicle and pedestrian traffic and observed with a far field of view video camera (80 meters). The results (shown in table 3) range between 58 % and 61 % which is considered to be promising by the end-user assessment but which is far from the rate obtained when monitoring a lock-chamber of a building.

Activity type	Nb of activities	True positives	False negatives	False alarms
small lock chamber, 1 person passing alone	17	94.10 %	5.90 %	1 (over 4 hours)
small lock chamber, >= 2 persons passing together	25	96.00 %	4.00 %	1 (over 4 hours)
large lock chamber, 1 person passing alone	72	94.50 %	5.50 %	4 (over 4 hours)
large lock chamber,>= 2 persons passing together	28	92.90 %	7.10 %	2 (over 4 hours)
violence in urban site 1	26	58 %	42 %	1 (over 4 hours)
violence in urban site 2	26	61 %	39 %	2 (over 2 hours)

Table 3: *Validation results for a lock chamber access monitoring system and an urban violence videosurveillance system (industrial project Videa).*

C.3 Conclusion

In conclusion these performance evaluations show that our current systems are very robust for simple activities mainly based of the localisation in 3D of a small number of persons inside a metro station, a bank agency or a building. Globally there are more than 90% of true positive (or correct) detections and less than 2 false alarms over several hours. The results for more complex activities such as fighting are promising (95% true positives in the case of isolated persons inside a metro station and about 60% in the case of far field of view of a urban busy environment) but these results are fragile and need to be improved.