INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Pulsar

Perception Understanding Learning Systems for Activity Recognition

Sophia Antipolis - Méditerranée

THEME COG

Activity Report

2008

Table of contents

1. Team

Research Scientist

Monique Thonnat [Team Leader, Research Director (DR) Inria, HdR]
François Brémond [Research Associate (CR) Inria, HdR]
Guillaume Charpiat [Research Associate (CR) Inria]
Sabine Moisan [Research Associate (CR) Inria, HdR]
Annie Ressouche [Research Associate (CR) Inria]

External Collaborator

Jean-Paul Rigault [Professor, Faculty Member Nice Sophia-Antipolis University]
Christophe Tornieri

Technical Staff

Bernard Boulay [European Project COFRIEND]
Etienne Corvée [European Projects CARETEKER and VIDEoID]
Erwan Demairy [IR2 INRIA for the SUP Software Platform's ADT, since April 2008]
Ruihua Ma [European project SERKET, upto July 2008]
José Luis Patino Vilchis [European Projects CARETEKER and COFRIEND]
Tomi Raty [European project SERKET, upto February 2008]
Valery Valentin [SIC Project]

PhD Student

Duc Phu Chau [Paca Grant, Nice Sophia-Antipolis University since December]
Lan Le Thi [Hanoi University and Nice Sophia-Antipolis University]
Mohamed Bécha Kaâniche [Paca Lab Grant, Nice Sophia-Antipolis University]
Anh Tuan Nghiem [Nice Sophia-Antipolis University, SIC Project]
Guido Pusiol [Nice Sophia-Antipolis University, CORDIs, since September 2008]
Nadia Zouba [Nice Sophia-Antipolis University]
Marcos Zúñiga [CONYCIT Grant, Nice Sophia-Antipolis University, upto December 2008]

Post-Doctoral Fellow

Ikhlef BECHAR [Arc Bioserre, since December 2008]
Vincent Martin [Arc Monitore, since April 2008]
Sundaram Suresh [ERICM PostDoc, upto September 2008]

Visiting Scientist

Lavee Gal [since September 2008 upto October 2008]

Administrative Assistant

Catherine Martin [Secretary (SAR) Inria]

Other

Slavomir Bak [Internship EGIDE Inria since March 2008 upto September 2008]
Duc Phu Chau [Internship EGIDE Inria since March 2008 upto September 2008]
Claudio Costaglia [Internship EGIDE Inria since January 2008 upto July 2008]
Edouardo Cruz Gonzalès [Internship Color COSIMO since May 2008 upto November 2008]
Germàn Delbianco [Internship EGIDE Inria since May 2008 upto November 2008]
Makrem Djebali [Internship EGIDE Inria upto February 2008 2008]
Cecilia Diaz [Internship EGIDE Inria since September 2008 upto end of December 2008]
Frederic Fabre [Internship Arc Bioserre since June 2008 upto December 2008]
Vivien Fighiera [Internship Color SynComp since July 2008 upto October 2008]
Adrian Hoitan [Internship Arc Bioserre since June 2008 upto December 2008]
Mohamed Moumene [Internship Arc MoniTore since April 2008 upto September 2008]
Domnica Pavel [Internship Color SynComp since April 2008 upto October 2008]
Guido Pusiol [Internship EGIDE Inria since March 2008 upto September 2008]
Rym Romdhame [since December 2008]

Mohamed Siala [Internship EGIDE Inria since April upto October 2008]
Alejandro Sanchez [Internship EGIDE Inria since April 2008 upto September 2008]

2. Overall Objectives

2.1. Presentation

2.1.1. Research Themes

Pulsar is focused on cognitive vision systems for Activity Recognition. We are particularly interested in the **real-time semantical interpretation of dynamic scenes** observed by sensors. We thus study spatio-temporal activities performed by human beings, animals and/or vehicles in the physical world.

Our objective is to propose new techniques in the field of **cognitive vision** and **cognitive systems** for *mobile object perception, behavior understanding, activity model learning, dependable activity recognition system design and evaluation*. More precisely Pulsar proposes 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). Our research work includes knowledge representation and symbolic reasoning for behavior understanding. We also study how statistical techniques and machine learning in general can complement a priori knowledge models for activity model learning. Our research work on software engineering consists in designing and evaluating effective and efficient activity recognition systems. Pulsar takes advantage of a pragmatic approach working on concrete problems of activity recognition to propose new cognitive system techniques inspired by and validated on applications, in a virtuous cycle.

Within Pulsar we focus on two main applications domains **safety/security** and **healthcare**. There is an increasing need to automate the recognition of activities observed by sensors (usually CCD cameras, omni directional cameras, infrared cameras), but also microphones and other sensors (e.g. optical cells, physiological sensors). Safety/security application domain 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, problems and fundings. Pulsar is also interested in developing activity monitoring applications for healthcare (in particular assistance for the elderly).

2.1.2. International and Industrial Cooperation

Our work has been applied in the context of more than 7 European projects such as AVITRACK, SERKET, CARETAKER, COFRIEND. We have industrial collaborations in several domains: transportation (CCI Airport Toulouse Blagnac, SNCF, INRETS, ALSTOM, RATP, Rome ATAC Transport Agency (Italy), Turin GTT (Italy)), banking (Crédit Agricole Bank Corporation, Eurotelis and Ciel), security (THALES R&T FR, THALES Security Syst, INDRA (Spain), EADS, Sagem, Bertin, Alcatel, Keeneo, ACIC, BARCO, VUB-STRO and VUB-ETRO (Belgium)), multimedia (Multitel (Belgium), Thales Communications, IDIAP (Switzerland), SOLID software editor for multimedia data basis (Finland)), civil engineering sector (Centre Scientifique et Technique du Bâtiment (CSTB)), computer industry (BULL), software industry (SOLID software editor for multimedia data basis (Finland), AKKA) and hardware industry (ST-Microelectronics). We have international cooperations with research centers such as Reading University (UK), ARC Seibersdorf research GMBHf (Wien Austria), ENSI Tunis (Tunisia), National Cheng Kung University (Taiwan), National Taiwan University (Taiwan), MICA (Vietnam), IPAL (Singapore), I2R (Singapore), NUS (Singapore), University of Southern California (USC), University of South Florida (USF), University of Maryland.

2.2. Highlights of the year

Pulsar is a new Project team created in January 2008 in the continuation of the Orion project. A new research scientist (Guillaume Charpiat) has joined the team. We have started to redesign the team software platform called SUP (Scene Understanding Platform) for activity recognition based on new software engineering paradigms. We have also started new work on learning techniques such as data mining in large multimedia database. For instance, we have been able to extract meaningful knowledge from the processing of more than one week of audio-video recordings in Rome and Turin subways.

3. Scientific Foundations

3.1. Introduction

Pulsar conducts two main research axes : scene understanding for activity recognition and software engineering for activity recognition.

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

Software engineering methods allow to ensure genericity, modularity, reusability, extensibility, dependability, and maintainability. To tackle this challenge, we rely on the correct theoretical foundations of our models, and on state-of-the art software engineering practices such as components, frameworks, (meta-)modeling, and model-driven engineering.

3.2. Scene Understanding for Activity Recognition

Keywords: *Activity Recognition*, *Computer Vision*, *Scene Understanding*.

Participants: Guillaume Charpiat, François Brémond, Sabine Moisan, Monique Thonnat.

3.2.1. Introduction

Our goal is to design a framework for the easy generation of autonomous and effective scene understanding systems for activity recognition. Scene understanding is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and events. The signal level is characterized by strong noise, ambiguous, corrupted and missing data. Thus to reach a semantic abstraction level, models and invariants are the crucial points. A still open issue consists in determining whether these models and invariants are given a priori or are learned. The whole challenge consists in organizing all this knowledge in order to capitalize experience, share it with others and update it along with experimentation. More precisely we work in the following research axes: perception (how to extract perceptual features from signal), understanding (how to recognize a priori models of physical object activities from perceptual features) and learning (how to learn models for activity recognition).

3.2.2. Perception for Activity Recognition

We are proposing computer vision techniques for physical object detection and control techniques for supervisation of a library of video processing programs.

First for the real time detection of physical objects from perceptual features, we design methods either by adapting existing algorithms or proposing new ones. In particular, we 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). Also to garantee the long-term coherence of tracked objects, we are adding a reasoning layer to a classical bayesian framework, modeling the uncertainty of the tracked objects. This reasoning layer is taking into account the a priori knowledge of the scene for outlier elimination and long term coherency checking. Morover we are working on providing fine and accurate models for human shape and gesture, extending the work we have done on human posture recognition matching 3D models and 2D silhouettes. We are also working on gesture recognition based on 2D feature point tracking and clustering.

A second research direction is to manage a library of video processing programs. We are building a perception library by selecting robust algorithms for feature extraction, by insuring they work efficiently with real time constraints and by formalising their conditions of use within a program supervision model. In the case of video cameras, at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues, we are developing new learning techniques.

3.2.3. 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 the objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. 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).

Concerning the modeling of activities, we are working towards two directions: the uncertainty management for expressing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. For example, we have built a language for video event modeling and a visual concept ontology (including color, texture and spatial concepts) to be extended with temporal concepts (motion, trajectories, events ...) and other perceptual concepts (physiological sensor concepts ...).

3.2.4. Learning for Activity Recognition

Given the difficulty of building an activity recognition system with a priori knowledge for a new application, we study how machine learning techniques can automate building or completing models at the perception level and at the understanding level.

At the perception level, to improve image segmentation, we are using program supervision techniques combined with learning techniques. For instance, given an image sampling set associated with ground truth data (manual region boundaries and semantic labels), an evaluation metric together with an optimisation scheme (e.g. simplex algorithm or genetic algorithm) are applied to select an image segmentation method and to tune image segmentation parameters. Another example, for handling illumination changes, consists in clustering techniques applied to intensity histograms to learn the different classes of illumination context for dynamic parameter setting.

At the understanding level, we are learning primitive event detectors. This can be done for example by learning visual concept detectors using SVMs (Support Vector Machine) with perceptual feature samples. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. leveraging the human annotation task). A second direction is the learning of typical composite event models for frequent activities using trajectory clustering or data mining techniques. We name composite event a particular combination of several primitive events.

Coupling learning techniques with a priori knowledge techniques is promising to recognise meaningful semantical activities.

The new proposed techniques for activity recognition systems (first research axis) are then contributing to specify the needs for new software architectures (second research axis).

3.3. Software Engineering for Activity Recognition

Keywords: *Generic Components*, *Knowledge-based Systems*, *Model-driven Engineering*, *Object-oriented Frameworks*, *Software Component Platform*, *Software Engineering*, *Software Reuse*.

Participants: François Brémond, Erwan Demairy, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Monique Thonnat.

3.3.1. Introduction

The aim of this research axis is to build general solutions and tools to develop systems dedicated to activity recognition. For this, we rely on state-of-the art Software Engineering practices to ensure both sound design and easy use, providing genericity, modularity, adaptability, reusability, extensibility, dependability, and maintainability.

This year we focused on four aspects: the definition of a joint software platform with KEENEO, the study of model-driven engineering approaches to facilitate platform usage, the extension of behavioral models, and formal verification techniques to design dependable systems.

3.3.2. Platform for Activity Recognition

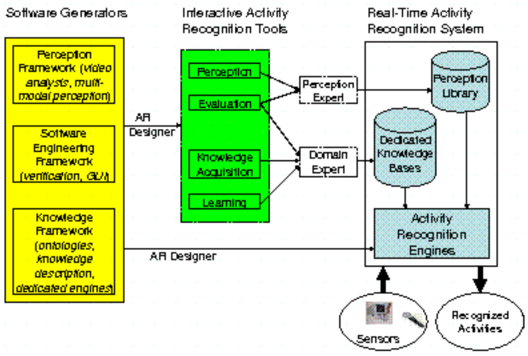

Figure 1. Global Architecture of an Activity Recognition Platform.

In the former project team Orion, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [12], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system or a cognitive system.

Pulsar will continue to study generic systems and object-oriented frameworks to elaborate a methodology for the design of activity recognition systems. We want to broaden the approach that led to LAMA and to apply it to the other components of the activity recognition platform, in particular to the image processing ones. We also wish to contribute to set up, in the long term, a complete software engineering methodology to develop activity recognition systems. This methodology should be based on model engineering and formal techniques.

To this end, Pulsar plans to develop a new platform (see figure 1) which integrates all the necessary modules for the creation of real-time activity recognition systems. Software generators provide designers with perception, software engineering and knowledge frameworks. Designers will use these frameworks to create both dedicated activity recognition engines and interactive tools. The perception and evaluation interactive tools enable a perception expert to create a dedicated perception library. The knowledge acquisition, learning and evaluation tools enable a domain expert to create a new dedicated knowledge base.

This platform will rely on the LAMA experiment, but necessitates some architectural changes and model extensions. We plan to work in the following three research directions: 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 practice and, if needed, we shall attempt to set up new ones.

The new platform should be easy to use. We should thus define and implement tools to support modeling, design, verification inside the framework. Another important issue deals with user graphical interfaces. It should be possible to plug existing (domain or application dependent) graphical interfaces into the platform. This requires defining a generic layer to accommodate various sorts of interfaces. This is clearly a medium/long term goal, in its full generality at least.

3.3.3. Software Modeling for Activity Recognition

Developing integrated platforms such as SUP is a current trend in video surveillance. It is also a challenge since these platforms are complex and difficult to understand, to use, to validate, and to maintain. The situation gets worse when considering the huge number of choices and options, both at the application and platform levels. Dealing with such a *variability* requires formal modeling approaches for the task specification as well as for the software component description.

Model Driven Engineering (MDE) [50] is a recent line of research that appears as an excellent candidate to support this modeling effort while providing means to make models operational and even executable. Our goal is to explore and enrich MDE techniques and model transformations to support the development of product lines for domains presenting multiple variability factors such as video surveillance.

The long term scientific objective concerns research activities both on Model Driven Engineering and video surveillance. On the MDE side, we wish to identify the limits of current techniques when applied to real scale complex tasks. On the video surveillance side, the trend is toward integrated software platforms, which requires formal modeling approaches for the task specification as well as for the software component description.

This MDE approach is complementary to the Program Supervision one, which has been studied by Orion for a long time [21]. Program Supervision focuses on programs, their models and the control of their execution. MDE also covers task specification and transformations to a design and implementation.

3.3.4. Behavioral Models for Activity Recognition

Pursuing the work done in Orion, we need to consider other models to express knowledge about activities, their actors, their relations, and their behaviors.

The evolution toward activity recognition requires various theoretical studies. We need to complement the knowledge representation model with a first-class notion of relations. The incorporation of a *model of time*, both physical and logical, is mandatory to deal with temporal activity recognition especially in *real time*. A fundamental concern is to define an abstract *model of scenarios* to describe and recognize activities. Supporting distributed systems is unavoidable for current software systems and it requires a *model of distribution*. These conceptual models will lead to define corresponding *ontologies*.

Finally, handling uncertainty is a major theme of Pulsar and we want to introduce it into our platform; this requires deep theoretical studies and is a long term goal.

3.3.5. Safeness of Systems for Activity Recognition

Another aim is to build dependable systems. Since traditional testing is not sufficient, it is important to rely on formal verification techniques and to adapt them to our component models.

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. In Orion, we already provided toolkits to ensure validation and verification of systems built with LAMA. First, we offered a knowledge base verification toolkit, allowing to verify the consistency and the completeness of a base as well as the adequacy of the knowledge with regard to the way an engine is going to use it. Second, we also provided an engine verification toolkit that relies on *model-checking* techniques to verify that the BLOCKS library has been used in a safe way during knowledge based system engine designs.

Generation of dependable systems for activity recognition is an important challenge. System validation really is a crucial phase in any development cycle. Partial validation by tests, 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. Formal methods help to produce a code that has been formally proved and the size and frequency of which can be estimated. Consistently with our component approach, it appears natural to rely on component modeling to perform a verification phase in order to build safe systems. Thus we study how to ensure safeness for components whose models take into account time and uncertainty.

Nevertheless, software dependability cannot be proved by relying on a single technique. Some properties are decidable and they can be checked using formal methods at the model level. By contrast, some other properties are not decidable and they require non exhaustive methods such as abstract interpretation at the code level. Thus, a verification method to ensure generic component dependability must take into account several complementary verification techniques.

4. Application Domains

4.1. Overview

Keywords: *bioinformatics*, *environment*, *health*, *multimedia*, *transportation*, *visual surveillance*.

While in our research the focus is to develop techniques, models and platforms that are generic and reusable, we also make effort in the development of real applications. The motivation is twofold. The first is to validate the new ideas and approaches we introduced. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, the applications we achieved cover a wide variety of domains: intelligent visual surveillance of underground stations, or applications in medical domain.

4.2. Video Surveillance

The growing feeling of insecurity among the population led the private companies as well as the public authorities to deploy more and more security systems. For the safety of the public places, the video camera based surveillance techniques are commonly used, but the multiplication of the camera number leads to the saturation of transmission and analysis means (it is difficult to supervise simultaneously hundreds of screens). For example, 1000 cameras are viewed by two security operators for monitoring the subway network of Brussels. In the framework of our works on automatic video interpretation, we have studied the conception of an automatic platform which can assist the video-surveillance operators.

The aim of this platform is to act as a filter, sorting the scenes which can be interesting for a human operator. This platform is based on the cooperation between an image processing component and an interpretation component using artificial intelligent techniques. Thanks to this cooperation, this platform automatically recognize different scenarios of interest in order to alert the operators. These works have been realized with academic and industrial partners, like European projects PASSWORDS, AVS-PV, AVS-RTPW, ADVISOR and AVITRACK and more recently, European projects CARETAKER, SERKET, CANTATA and COFRIEND, national projects SIC, VIDEOID, industrial projects RATP, CASSIOPEE, ALSTOM and SNCF. A first set of very simple applications for the indoor night surveillance of supermarket (AUCHAN) showed the feasibility of this approach. A second range of applications has been to investigate the parking monitoring where the rather large viewing angle makes it possible to see many different objects (car, pedestrian, trolley) in a changing environment (illumination, parked cars, trees shacked by the wind, etc.). This set of applications allowed us to test various methods of tracking, trajectory analysis and recognition of typical cases (occlusion, creation and separation of groups, etc).

We have studied and developed video surveillance techniques in the transport domain which requires the analysis and the recognition of groups of persons observed from lateral and low position viewing angle in subway stations (subways of Nuremberg, Brussels, Charleroi, Barcelone, Rome and Turin). We have worked with industrial companies (Bull, Vigitec, Keeneo) on the conception of a video surveillance intelligent platform which is independent of a particular application. The principal constraints are the use of fixed cameras and the possibility to specify the scenarios to be recognized, which depend on the particular application, based on scenario models which are independent from the recognition system.

In parallel of the video surveillance of subway stations, projects based on the video understanding platform have started for bank agency monitoring, train car surveillance and aircraft activity monitoring to manage complex interactions between different types of objects (vehicles, persons, aircrafts). A new challenge consists in combining video understanding with learning techniques (e.g. data mining) as it is done in the CARETAKER and COFRIEND projects to infer new knowledge on observed scenes.

4.3. Detection and Behavior Recognition of Bioagressors

In the environmental domain, Pulsar is interested in the automation of early detection of bioagressor, especially in greenhouse crops, in order to reduce pesticide use. Attacks (from insects or fungi) imply almost immediate decision-taking to prevent irreversible proliferation. The goal of this work is to define innovative decision support methods for *in situ* early pest detection based on video analysis and scene interpretation from multi camera data. We promote a non-destructive and non-invasive approach to allow rapid remedial decisions from producers. The major issue is to reach a sufficient level of robustness for a continuous surveillance.

During the last decade, most studies on video applications for biological organism surveillance were limited to constrained environments where camerawork conditions are controlled. By contrast, we aim at monitoring pests in their natural environment (greenhouses). We thus intend to automate pest detection, in the same way as the management of climate, fertilization and irrigation which are carried out by a control/command computer system. To this end, vision algorithms (segmentation, classification, tracking) must be adapted to cope with illumination changes, plant movements, or insect characteristics.

Traditional manual counting is tedious, time consuming and subjective. We have developed a generic approach based on *a priori* knowledge and adaptive methods for vision tasks. This approach can be applied to insect images in order, first, to automate identification and counting of bio-aggressors, and ultimately, to analyze insect behaviors. Our work takes place within the framework of cognitive vision [48]. We propose to combine image processing, neural learning, and *a priori* knowledge to design a system complete from video acquisition to behavior analysis. The ultimate goal of our system is to integrate a module for insect behavior analysis. Indeed, recognition of some characteristic behaviors is often closely related to epicenters of infestation. Coupled with an optimized spatial sampling of the video cameras, it can be of crucial help for rapid decision support.

Most of the studies on behavior analysis have concentrated on human beings. We intend to extend cognitive vision systems to monitor non-human activities. We will define scenario models based on the concepts of *states* and *events* related to intertesting *objects*, to describe the scenarios relative to white insect behaviors. We shall also rely on ontologies (such as a a video event one). Finally, in the long term, we want to investigate data mining for biological research. Indeed, biologists require new knowledge to analyze bioagressor behaviors. A key step will to be able to match numerical features (based on trajectories and density distributions for instance) and their biological interpretations (e.g., predation or center of infestation).

This work takes place in a two year collaboration (ARC BioSERRE) between Pulsar (INRIA Sophia Antipolis - Méditerranée), Vista (INRIA Rennes - Bretagne Atlantique), INRA Avignon UR407 Pathologie Végétale (Institut National de Recherche Agronomique), CREAT Research Center (Chambre d'Agriculture des Alpes Maritimes) started in 2008.

4.4. Medical Applications

In the medical domain, Pulsar is interested in the long-term monitoring of people at home, which aims to support the caregivers by providing information about the occurrence of worrying change in people behavior. We are especially involved in the GER'HOME project, funded by the PACA region and Conseil General (CG06), in collaboration with two local partners: CSTB and Nice City hospital. In this project, an experimental home that integrates new information and communication technologies has been built in Sophia Antipolis. The purpose concerns the issue of monitoring and learning about people activities at home, using autonomous and non-intrusive sensors. The goal is to detect the sudden occurrence of worrying situations, such as any slow change in a person frailty. The aim of the project is to design an experimental platform, providing services and allowing to test their efficiency.

The context of monitoring epileptic patients using videos is also investigated in collaboration with Marseille City hospital (Prof. P. Chauvel, La Timone). One purpose in terms of activity monitoring is to model the changes in behavior for a given patient when in ceasure. The ultimate goal is to cluster and/ or classify ceasures of patient groups, so that the localization of the brain lesion of a given patient is more easily determined.

5. Software

5.1. Pegase

Since September 1996, the Orion team (and now the Pulsar team) distributes the program supervision engine PEGASE, based on the LAMA platform. The Lisp version has been used at Maryland University and at Genset (Paris). The C++ version (PEGASE+) is now available and is operational at ENSI Tunis (Tunisia) and at CEMAGREF, Lyon (France).

5.2. SUP

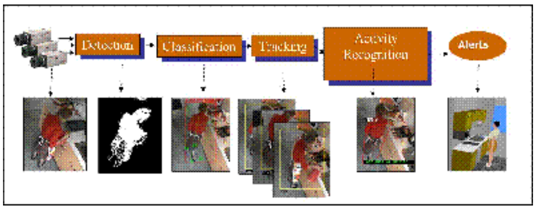

Figure 2. Components of the Scene Understanding Platform (SUP).

SUP is a Scene Understanding Software Platform written in C and C++ (see figure 2) based on Orion team platform (VSIP: VideoSurveillance Intelligent Platform). Four modules of the VSIP platform have been registered at APP (the French agency for patrimony protection) in 2005. These modules are:

1. VSIP-DMM contains the global architecture for data and module management;
2. VSIP-OD contains the image processing algorithms in charge of a video stream of one camera (mobile object detection, classification and frame to frame tracking);
3. VSIP-STA contains the multi-camera algorithms for the spatial and temporal analysis (4D) of the detected mobile objects;
4. VSIP-TSR contains the high level scenario recognition algorithms and scenario representation parsers.

Several versions of VSIP have been transferred to industrial partners: in 2003 to **Bull**, to **Thales**, and to **Ciel** (integrator in Toulon), to **Vigitec** a Belgium specialist in Videosurveillance, in July 2005 to **Reading** in UK. VSIP has been exploited by **Keeneo**, the Start-up created in July 2005 by the research team.

6. New Results

6.1. Scene Understanding for Activity Recognition

Participants: Slawomir Bak, Bernard Boulay, François Brémond, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Claudio Costaglia, Frédéric Fabre, Adrian Hoitan, Mohamed Bécha Kaâniche, Thi-Lan Le, Vincent Martin, Sabine Moisan, Mohamed Moumene, Anh-Tuan Nghiem, Guido Pusiol, Mohamed Siala, Suresh Sundaram, Monique Thonnat, Valéry Valentin, José Luis Patino Vilchis, Nadia Zouba, Marcos Zúñiga.

6.1.1. Introduction

This year, Pulsar has tackled several scene understanding issues and proposed new algorithms in the three following research axes:

- Perception: object detection and object tracking algorithms, human gesture recognition;
- Understanding: activity recognition using people postures;
- Learning: relevance feedback for video retrieval, incremental learning of video events and online trajectory clustering.

6.1.2. Video Pest Detection in Greenhouses

Participants: Frédéric Fabre, Adrian Hoitan, Vincent Martin, Sabine Moisan.

Inside a greenhouse, attacks from insects or fungi are fast and frequent. This implies almost immediate decisions to prevent irreversible proliferation. Integrated Pest Management promotes prophylactic, biological and physical methods to fight bio-aggressors while minimizing the use of pesticides. This approach is promising but requires frequent and precise observations of plants that are not compatible with production constraints.

Since the cost of video cameras is decreasing, it becomes realistic to equip greenhouses with such sensors. We thus propose an automatic system based on video analysis for *in situ* insect detection, classification, and counting. Our target application is the detection of bio-aggressors on plant organs such as leaves. The goal of this work is to define an innovative decision support system, which handles multi camera data and follows a generic approach to adapt to different categories of bio-aggressors. This approach is non-destructive and non-invasive. It will allow producers to take rapid remedial decisions. The major issue is to reach a sufficient level of robustness for continuous surveillance. To this end, vision algorithms (segmentation, classification, tracking) must be adapted to cope with illumination changes, plant movements, or insect characteristics. Our work takes place within the framework of cognitive vision [48]. We propose to combine image processing, neural learning, and *a priori* knowledge to design a complete system, from video acquisition to behavior analysis.

The first prototype of our decision support system is under tests in a rose greenhouse with a network of five wireless cameras. The position, number, and nature of video cameras are critical to obtain an optimized video sampling in terms of cost/accuracy. In this experiment, we choose to position the video cameras uniformly in the horizontal plane in order to optimize the horizontal sampling in terms of canopy area covering. The video cameras currently observe sticky traps in order to detect in real-time flying insects. In the future, we plan to locate other video cameras directly on plant organs as recommended by agronomic expertise, e.g. on growing stems for early detection of mature white flies.

The first objective is to detect and track bioagressors. In our case, the objects of interest are small, complex, and they evolve in a dynamic environment. We are thus developing *adaptive* vision methods at different levels (acquisition, detection, and tracking) to provide robust results: segmentation and classification should be able to cope with illumination changes during daytime and tracking algorithms should accommodate plant movements. Our system, named DIViNe[1], is composed of several modules. The complete system is described in Figure 3.

This year, in order to allow a tractable data flow on the network, we have mainly worked on the intelligent acquisition module [30], [31] that records images only when an insect motion is detected. After acquisition, visual data extracted from images (color, texture, shape, and size) an analysis module should detect regions that may correspond to insects in different contexts. We plan to use video analysis algorithms combined with *a priori* knowledge about the visual appearance of insects to detect. The classification module then has to select interesting regions, retaining only the ones corresponding to target insects and count them. Finally, we intend to use scenario recognition techniques to analyze insect behaviors such as egg laying or intra-guild predation.

[1]Detection of Insects by a Video camera Network.

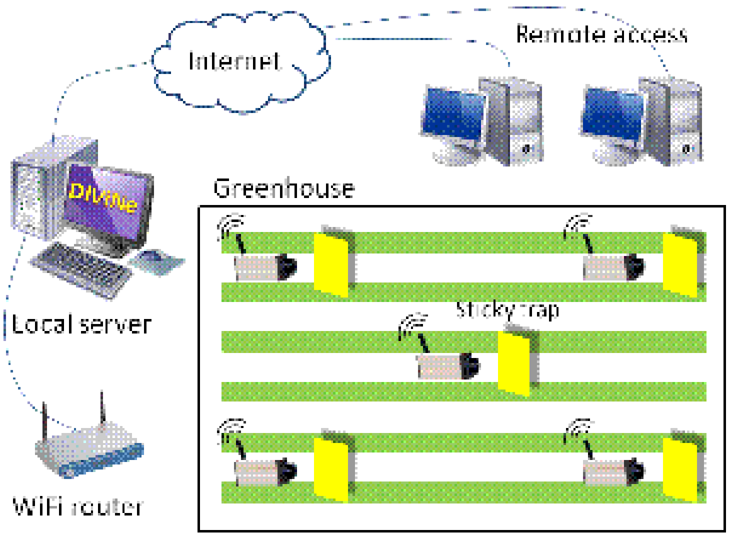

Figure 3. Overview of the DIViNe system and the wireless video camera network.

DIViNe currently detects a few types of pests (mature white flies, aphids), but within two years we intend to detect most of the common greenhouse crop bioagressors. Such a system can detect low infestation stages, which helps producers to rapidly decide on possible treatment.

6.1.3. Real-time Monitoring of TORE Plasma

Participants: Vincent Martin, Mohamed Moumene, François Brémond, Monique Thonnat.

This project aims at developing an intelligent system for the real time surveillance of the plasma evolving in Tore Supra or other devices. The first goal is to improve the reliability and upgrade the current real time control system operating at Tore Supra. The ultimate goal is to integrate such a system into the future ITER imaging diagnostic. In this context, a first collaboration has recently started between the Imagery and Diagnostic Group of CEA Cadarache and the Pulsar project-team. The goal is to detect events (expected or not) in real-time, in order to control the power injection for the protection of the Plasma Facing Components (PFC). In the case of a known event, the detection must lead to the identification of this event and finally to the appropriate reaction (decrease the RF power injected into the plasma, increase or decrease the fuelling). Otherwise, a learning process is proposed to assimilate this new type of event. In such way, the objective of the project is twofold: machine protection and thermal event understanding.

The sytem takes multimodal data as inputs: plasma parameters (plasma density, injected power by heating antenna, plasma position, ...), infrared and visible images, and signals coming from others sensors as spectormeters and bolometers. Recognized events are returned as outputs with their characteristics for further physical analysis.

In this application, we benefict from the large amount of available data accumulated during thousands of pulses and for several devices. We rely on an ontology-based representation of the *a priori* domain expert knowledge of thermal events. This thermal event ontology is based on visual concepts and video event concepts useful to decribe and recognize a large variety of events occuring in thermal videos.

This year, we have focused on the conception of the thermal event ontology and the developpement of a first system prototype for the automatic identification of some basic thermal events, mainly electrical arcing (see figure 4) and UFO (see 5). More precisely, we have proposed a first version of the ontology composed of 5 classes (i.e. thermal events) and used this ontology for the manual annotations of 44 pulses corresponding to 600 event annotations (Mohamed Amine Moumene's internship). Then we have worked on the detection problem. To this end, different state-of-the-art motion detection algorithms (based on background modeling) have been implemented and tested on the Tore Supra infrared images. The developped thermal event recognition prototype overperforms the previous detection system in terms of precise localization and discrimination between thermal events. A paper on this subject has been recently submitted to the IEEE I2MTC conference.

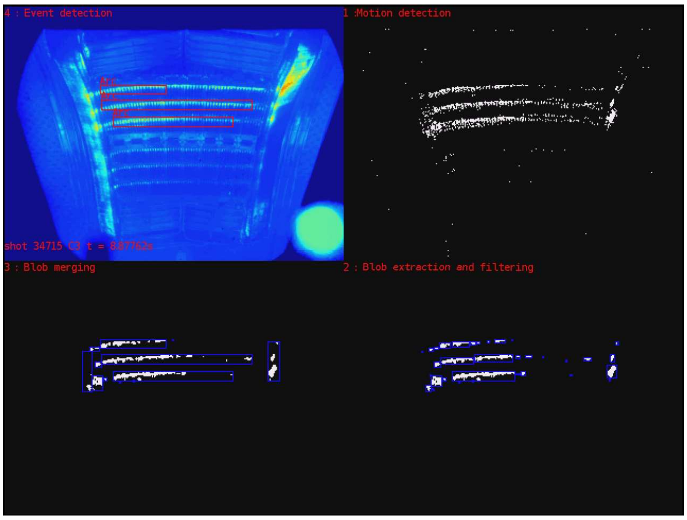

Figure 4. Electrical arcing event recognition at different steps.

From this first attempt, several directions are investigated:

- implement the low-level algorithms (motion detection) on a FPGA to reach the required framerate (50 fps for Tore Supra IR cameras) for real-time detection;
- enrich the ontology for the detection and recognition of other abnormal thermal events such as disruption precursors and carbon flakes;
- fuse multi-sensor data for both learning purposes and system reliability enhancement. To this end, we aim at applying a bayesian framework in the event recognition process so as to take into account uncertainty.

6.1.4. Shadow Removal in Indoor Scenes

Participants: Anh-Tuan Nghiem, François Brémond, Monique Thonnat.

Removing shadows is a crucial task for many video surveillance systems. To do this, there are two main approaches: chromaticity-based and texture-based. Each of them has its own strengths and weaknesses. Hence, combining these two approaches would improve the algorithm accuracy. However, it would be very slow because the texture-based approach is often time-consuming. To get around the problem, the proposed algorithm uses a chromaticity constraint together with a simple texture verification. The chromaticity consistency is verified in HSV as well as in RGB color spaces. The texture verification is based on the local coherency (over a pixel neighborhood) of intensity reduction ratio between shadows and background. Besides, a constraint on the ratio of intensity has also been imposed to better discriminate shadows from

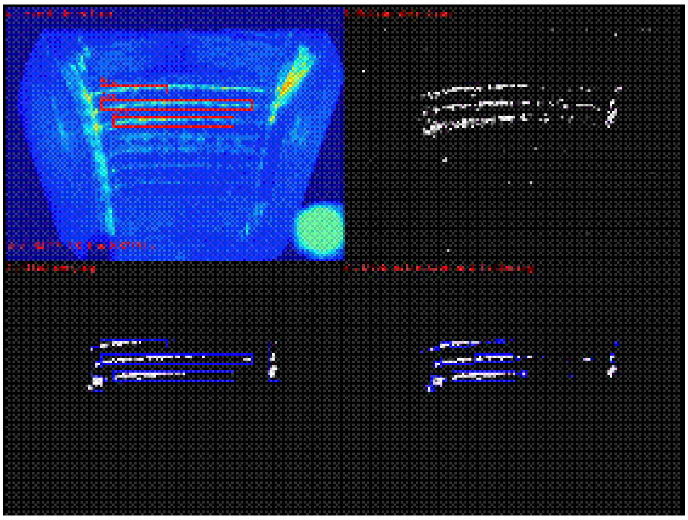

Figure 5. UFO detection and tracking using a particle filter and tracking algorithm.

background. To validate these constraints we have constructed a shadow model based on Phong illumination model. The shadow model shows that those constraints can be satisfied in most of indoor scene with flat surfaces. In our experiment, comparing to the popular chromaticity based algorithm, in average, the proposed algorithm obtains better results on both precision (78% over 74.9%) and sensitivity (69.7% over 67.3%) of object detection task. The detail of this algorithm is published in [33] .

6.1.5. Learning Shape Metrics

Keywords: *Learning*, *Perception*.

Participants: Guillaume Charpiat, Bernard Boulay.

Precise object detection or classification requires the notion of shape. On the one hand, shape priors can help image segmentation or tracking, and on the other hand, shape is a huge source of information for scene analysis (posture, gesture, etc.). We consequently aim at modeling shapes. Building shape models by hand is however complex and cannot be guaranteed to fit the real data observed. Therefore, our approach consists in learning shapes, i.e. in building statistical models of shapes from training sets of shapes. Further works would then be to use these shape models in image-based frameworks [26], [39], [23].

Shape evolutions, as well as shape matchings or image segmentation with shape prior, involve the preliminary choice of a suitable metric in the space of shapes. Instead of choosing a particular metric, we propose a framework to learn shape metrics from a set of examples of shapes. First, we design a criterion to compute point-to-point matching between pairs of shapes (possibly topologically different), which is especially reliable for close shapes. Dealing with topological changes is important since we consider video sequences such as walking human silhouettes.

Then, given a training set of shapes, we compute pairwise matchings with this criterion and use them to build a manifold-like structure on the set of shapes. An example is given in figure 6. The matchings are useful to transport fields, in particular they are able to transport a deformation observed on a particular shape to any another shape, with a reliability which depends on the distance between the two shapes. We estimate the metric in the tangent space of any shape, based on transported deformations, weighted by their probability. This way of processing allows the consideration of non-dense samples. Indeed, usual 3D models of human silhouettes have about 30 degrees of freedom, so that there is no hope in expecting a dense training set of shapes.

The estimated metric can then be turned into a shape prior or a matching prior. The framework is particularly well suited for video analysis because it relies on close shape matching, and opens new perspectives in segmentation learning. We also formulate the task of finding the best metric as a classical minimization problem, and prove that the metric we compute is the optimal one.

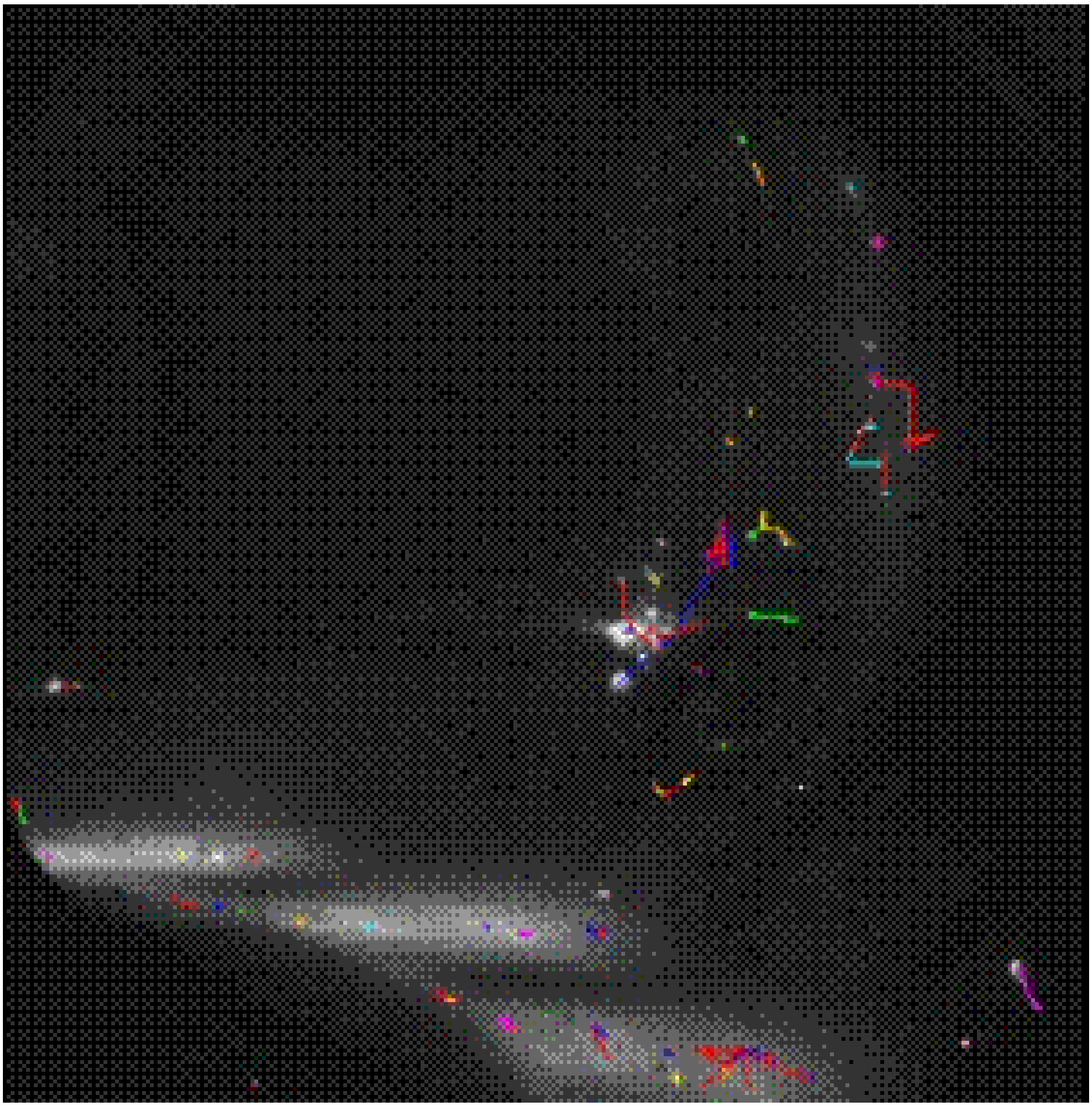

Figure 6. Colors on one shape, propagated via transports to other shapes from a same video sequence, in order to visualize computed correspondences.

6.1.6. People Detection in Complex Scene Using a Cascade of Boosted Classifiers Based on Haar-like-features

Participants: Mohamed Siala, Francois Brémond, Monique Thonnat.

Detecting people in real world videos is the challenging video surveillance application of the VIDEO-ID project. We have proposed an approach that detects different poses of a person in an image using a cascade of boosted classifiers based on Haar-like-features similarly to the work of Viola and Jones, which is implemented on the OpenCV library.

Adaboost optimizes a combination of weak classifiers to obtain strong classifiers. A weak classifier combines several human descriptors (Haar features) on a sub-window of human size and aims at filtering out other sub-windows without human (low false negatives rate). A cascade combines several layers of strong classifiers. We have used the Adaboost classifiers and changed the training samples to adapt the algorithm to fixed cameras. We have defined four human body poses (frontal, left side, right side, back) and one occluded body. For each pose we have trained 15 layers of strong classifiers separately, and we have obtained five detectors for all the poses. During the training phase, we have used about 1600 positive samples from one scene (Trecvid Video) with varied people appearances, and more than 1800 background images from varied scenes as negative samples. We have used specific video samples to take advantage of the fixed cameras.

Figure 7. Example of people detection using the Adaboost classifiers.

During the training phase, the cascade parameters (number of layers) and threshold values (false positive rate (0.6) and false negative rate (0.1) for the strong classifiers) are tuned to give a higher detection rate. The proposed algorithm is able to detect humans, see illustration in figure 7, with the following rates: TP = 900/1540, FN = 640/1540 and FP = 1900/2800. These rates have been obtained by taking 90% of the annotated samples for training and 10% for testing.

We have compared our work with the training samples provided by the OpenCV library: TP = 50/1540, FN = 1490/1540 and FP = 100/150. Our results outperform the OpenCV training samples. The algorithm is at a preliminary stage but can work in difficult conditions, it gives a good detection despite occlusion and different human poses. Concerning future work we are planning to test different descriptors such as Histogram of Oriented Gradients.

6.1.7. Online Adaptive Neural Classifier for Robust Tracking

Participants: Slawomir Bak, Suresh Sundaram, Claudio Costaglia, Francois Bremond, Monique Thonnat.

This project focuses on online tracking of mobile objects. The information coming from motion segmentation is fused with an online adaptive neural classifier. Then, the classifier is used to differentiate the object from its local background. The neural classifier adapts to the change in illumination and appearance throughout the video. The feedback from the motion segmentation helps in avoiding drifting problems due to similar appearance in the local background region.

The initialization of new targets is based on the motion information obtained from the object classification. Only classified objects are tracked. The basic building block of the neural classifier is a Radial Basic Function Network (RBFN). The input of the classifier are feature vectors extracted from objects and their local background region. To compute the error in the learning part we compare outputs of the neural network and results of the object(inside)/background (surroundings) separation scheme.

We have tested our approach on many data sets. We have used the video obtained from the GERHOME laboratory which is a realistic site reproducing the environment of a typical apartment of an elderly person. The laboratory promotes research in the domain of activity monitoring and assisted living. We have also performed our algorithm on ETISEO video sequences such as a metro scene (ETI-VS1-MO-7-C1) where a man drops his bag. This is an example of video with weakly contrasted objects. The approach has been also tested with TREC Video Retrieval Evaluation (TRECKVID 2008) data obtained from Gatwick Airport surveillance system where many targets are crossing each other. This work has been published in [25]. Preliminary work has also been done for including KLT feature points and SIFT descriptors to the model of the tracked person.

6.1.8. People Trajectory Extraction and Repairing for Video Event Recognition

Participants: Duc Phu Chau, Etienne Corvee, Francois Brémond, Monique Thonnat.

In the CARETAKER project, the PULSAR team developed an algorithm to follow detected foreground objects (i.e. people) evolving in a scene captured in Rome and Turin underground stations. Simple events have been extracted from the interaction of people with the contextual content (subway equipment and delimited zones) of scenes viewed by surveillance cameras. However, the low image quality of the compressed captured images with large field of views with the occlusion problem in crowded areas make the tracking of each individual a hard task to perform. This task becomes even more complicated to achieve when real time operating is required. Two main functionalities for the multiple object tracking algorithm have been developed.

The first functionality of the long term tracker is to provide real time tracking information to the event detector while being synchronized with the live video. This tracker is based on region matching (optimised for a 10 to 20 frames buffer) using 2D and 3D object information criteria. It globally tracks groups of people in a scene. The second functionality of the multiple object tracker is the capability to repair trajectories such as noisy trajectories (caused for instance by misdetection) or broken or unfinished trajectories due to static or dynamic occlusion. The approach of this trajectory repairing technique is based on learning erroneous trajectories and the ones with potential lost and reappearing objects. This learning process allows us to associate a confidence value to each trajectory and to then discard noisy trajectories with low confidence values independently of the viewed scene. The learning process also allows us to learn automatically the contextual zones where lost or broken trajectories occur in order to repair them. Therefore the tracker does not require any user interaction. The obtained results on a video of 5 hours with 8000 detected trajectories: 54% of noise or broken trajectories are filtered, 70% of incomplete trajectories were used to repair trajectories and increase the rate of complete trajectories from 9% to 20%. Figure 8 shows two examples of repaired trajectories. An individual is badly segmented on the top left frame (due to colour similarity between the background floor and his/her clothes). In the next captured frame (top middle frame) this individual is well segmented but the tracker looses the track (it creates a new one) due to the too large disparity in the 3D location. However, when the trajectory repairing technique is used, the individual keeps being tracked with his/her original track identifier as shown in the top right frame. The second example in the bottom three frames shows similar results where in this case the individual is not being detected at all for several frames. This causes the individual trajectory to be broken i.e. split into two distinct trajectories. The tracker manages to repair and keeps the original track.

The event detection algorithm uses the tracked objects obtained by the multiple object tracker described above and the contextual information about the viewed scene as inputs. Results were obtained online and offline along with the obtained tracking results (several days) and events ranging from *inside zone*, *close to* to *loitering* (i.e. *stays inside zone*), *use equipment* and *queuing at equipments*. The obtained events analysed by offline statistical analysis (see next) allowed us to enhance our evaluation on the tracking algorithms. These statistical results validate the fact that repaired trajectories make the number of tracked objects slightly increase and these objects stay longer in the zones than when non repaired. In general, the overall evaluation has shown that the operational and well tested long term tracker provides reliable enough tracking information for meaningful events for online (end user's terminal display) and offline analysis. This work has been published in [34].

6.1.9. Multi-object Tracking Using Reliability Measures

Keywords: *Perception.*

Participants: Marcos Zúñiga, François Brémond, Monique Thonnat.

We have proposed a new video understanding approach for multi-object tracking. This approach calculates reliability measures of the tracked object attributes, for improving the robustness of tracking. Reliability can be defined as the confidence or degree of trust we have on a measurement. This approach is composed of three tasks:

Figure 8. Two examples of repaired trajectories

First, at each video frame, a segmentation task detects the moving regions, represented by bounding boxes enclosing them.

Second, a new 3D classifier associates to each moving region an object class label (e.g. person, vehicle) and a 3D parallelepiped described by its width, height, length, position, orientation, and visual reliability measures of these attributes.

Finally, a new multi-object tracking algorithm uses these object descriptions to generate tracking hypotheses about the objects evolving in the scene. Reliability measures associated to the object features are used to perform a proper selection of valuable information.

This tracking approach has been validated using video-surveillance benchmarks publicly accessible.

6.1.10. Human Gesture Recognition

Keywords: *Learning*, *Understanding*.

Participants: Mohamed Bécha Kaâniche, François Brémond, Monique Thonnat.

We propose a learning method for Human Gesture Recognition which is based on classifying gestures against temporal 2D gesture descriptors. A set of body gestures has been predefined: sitting, standing, falling, bending, hand waving, applausing. We suppose that several image sequences illustrate each gesture. The beginning and the end of each sequence correspond to the beginning and the end of the related gesture. Our goal is to learn classifiers for gesture recognition in new videos. The proposed approach is composed of three stages.

- First, we compute for each detected human in the scene a set of 2D local descriptors, based on feature points (KLT) and Histograms of Oriented Gradients, which are tracked over time to build temporal 2D descriptors.
- Second, we learn the temporal 2D descriptors for a given gesture creating code-words from a learning video data set.
- Third, we classify the gesture of a person in a new video by extracting the person temporal 2D descriptors as new code-words and comparing them with learned code-words.

For the first stage, we have developed a feature point tracking algorithm based on Kanade Lucas Thomasi tracker. The tracker is extended by combining Histograms of Oriented Gradients and feature points to form 2D tracking descriptors. The temporal 2D gesture descriptors are extracted in four steps:

1. *Object Detection* is performed by background subtraction to determine moving regions and then by applying a people classifier. The people bounding boxes define a mask for feature point extraction.
2. *Feature Points* are extracted for each detected person using Shi and Thomasi Corner detector. The masks generated by the previous step are used to define image regions where to extract features.
3. *2D descriptors* are computed for each feature point by considering a descriptor block centered on it. The descriptor block is based on Histogram of Oriented Gradients and consists of 3x3 cells. Each cell is a 5x5 or 7x7 pixel patch.
4. *Temporal 2D Descriptors* are builded by tracking 2D descriptors. The tracker minimizes a quadratic error function between current 2D Descriptor and possible 2D Descriptors in the next frame. When several 2D descriptors in the next frame satisfies a minimum distance, the tracker considers the descriptors which have the most visual evidence similarity (gray scale intensity difference).

For the second stage, k-means algorithm is used to cluster the temporal 2D descriptor into video-words annotated with the related gesture for each learning video. The same video-word can be associated to different gestures. Then we compact the video-word set to obtain a code-book of optimal code-words by applying the algorithm of Maximization of Mutual Information (MMI).

For the third stage, we will apply a learn-and-predict strategy with one-versus-all method using SVM binary classifiers.

6.1.11. Visualisation Tool for Activity Recognition Evaluation

Keywords: *Evaluation*, *Graphical Tool*, *Learning*, *Understanding*.

Participants: Bernard Boulay, François Brémond, Monique Thonnat.

Co-FRIEND aims at designing a framework for understanding human activities in real world environment, through a cognitive vision approach, identifying objects and events, and extracting sense from scene observations. The framework will manage uncertainty and will adapt dynamically algorithms to change in environment, and will create a meaningful analysis of the scene. To validate this framework, a heterogeneous sensor network (wide angle and PTZ cameras and GPS installed on vehicles) has been deployed on Toulouse Airport for monitoring all servicing operations around aircraft parked on aprons.

Also a visualisation tool has been designed for quick visualisation of the interpretation results and quick evaluation of the vision and cognitive algorithms. This tool is composed of a graphical interface to visualise the processed video sequences and to friendly display the processing results. In particular, the bounding boxes of the detected objects and the classification type are displayed on a background image. A 3D virtual scene is built with a 3D representation of the objects (with 3D parallelepipeds) corresponding to the detected physical objects. To easily visualise the tracking information, lines representing the presence of the detected objects all over the sequence are displayed. The same display is made for the recognised events.
Thanks to this tool the algorithm capability of video processing and activity recognition can be easily highlighted. The tool has been used for the COFRIEND and GerHome project.

6.1.12. Monitoring Activities of Daily Living (ADLs) of Elderly, based on 3D Key Human Postures

Participants: Nadia Zouba, Valéry Valentin, Bernard Boulay, François Brémond, Monique Thonnat.

In previous work, we have modeled a set of interesting activities for elderly at home by combining video cameras with environmental sensors. However, some activities at home require detecting a fine description of human body such as postures. For this, we have used a human posture recognition algorithm [3] in order to detect in real time the posture of a person evolving in the scene. This algorithm needs 3D predefined models of human postures. We have adapted this algorithm for homecare applications, by modeling a set of interesting human postures which are useful to detect activities at home and critical situations for elderly (e.g. falls). For this purpose, we have modeled ten 3D key human postures which are displayed in figure 9: standing (a), standing with arm up (b), standing with hands up (c), bending (d), sitting on a chair (e), sitting on the floor with outstretched legs (f), sitting on the floor with flexed legs (g), slumping (h), lying on the side with flexed legs (i), and lying on the back with outstretched legs (j). Each of these postures plays a significant role in modeling and recognizing a set of interesting activities at home. For example, the posture standing with arm up is used to detect when a person reaches and opens kitchen cupboards and his/her ability to do it. The posture standing with hands up is used to detect when a person is carrying an object such as plates. More details about the proposed postures and the approach are described in [37].

To validate our models, we have tested a set of human activities in the Gerhome laboratory. Two experiments have been performed. One with actors, and the second one with real elderly people.
In the first experiment, we have acquired ten video sequences with only one human actor. The duration of each video is about ten minutes and each video contains about 4800 frames (about eight frames per second). In this experiment, we have tested some normal activities such as: open and close kitchen cupboards, use microwave and warm up a meal. We have also tested two abnormal activities: "feeling faint" and "falling down". Figure 10-a shows the recognition of the "feeling faint" situation and the 3D visualization of this recognition. Figure 10-b shows the recognition of the "falling down" situation and the 3D visualization of this recognition.

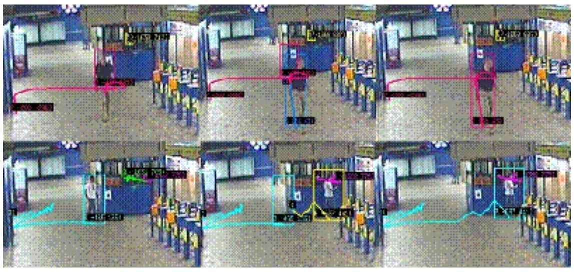

Figure 9. The proposed 3D human postures

In the second experiment, while evolving in the laboratory, fourteen volunteers (aged from 60 to 85 years) have been observed, each one during 4 hours and 14 video scenes have been acquired by 4 video cameras (about ten frames per second). In this experiment the fourteen volunteers were asked to perform a set of household activities, such as preparing meal, taking meal, washing dishes, cleaning the kitchen, and watching TV. Each volunteer was alone in the laboratory during the experiment. Figure 11-a shows the recognition of "bending in the kitchen" activity and the 3D visualization of this recognition. Figure 11-b shows the recognition of "taking meal" activity and the 3D visualization of this recognition.

(a) (b)

Figure 10. (a) "Feeling Faint" situation. (b) "Falling Down" situation.

(a) (b)

Figure 11. (a) "Bending" activity. (b) "Taking Meal" activity.

These preliminary results show that the proposed human postures play an important role in detecting interesting activities at home. A finer video analysis (especially by analyzing fine gestures) is needed to better measure people activities. More evaluation especially in real and natural environment on long periods is required to assess the robustness of the monitoring system. Similarly, more work is needed to establish a behavior profile of the observed person and to determine the significant changes in this profile.

6.1.13. Incremental Event Learning

Keywords: *Learning*.

Participants: Marcos Zúñiga, François Brémond, Monique Thonnat.

We propose a new approach for video event learning. The only hypothesis is the availability of tracked object attributes. The approach incrementally aggregates the attributes and reliability information of tracked objects to learn a hierarchy of state and event concepts. Simultaneously, the approach recognizes the states and events of the tracked objects. This approach proposes an automatic bridge between the low-level image data and higher level conceptual information. This new learning approach is unsupervised, incremental, and continuous.

The approach has been evaluated for more than two hours of an elderly care application. The results show the capability of the approach to learn and recognize meaningful events occurring in the scene. Also, the results show the potential of the approach for giving a description of the activities of a person (e.g. approaching to a table, crouching), and to detect abnormal events based on the frequency of occurrence.

6.1.14. Surveillance Video Indexing and Retrieval

Participants: Thi-Lan Le, Monique Thonnat, François Brémond.

The goal of this work is to propose a general approach for surveillance video indexing and retrieval. Based on the hypothesis that videos are preprocessed by an external video analysis module, this approach is composed of two phases : indexing phase and retrieval phase.

In order to profit output of various video analysis modules, a general data model consisting of two main concepts : objects and events is proposed. The indexing phase that aims at preparing data defined in the data model performs three tasks. Firstly, two new key blob detection methods in the object representation task choose for each detected object a set of key blobs associated with a weight. Secondly, the feature extraction task analyzes a number of visual and temporal features on detected objects. Finally, the indexing task computes attributes of the two concepts and stores them in the database.

The retrieval phase starts with a query of user and is composed of 4 tasks. In the formulation task, user expresses his query in a new rich query language. This query is then analyzed by the syntax parsing task. A new matching method in the matching task aims at retrieving effectively relevant results for the query. Two proposed methods in the relevance feedback task allow to interact with the user in order to improve retrieved results.

The key blob detection method has improved results of one method in the state-of-the-art. The analysis of query language usage shows that many queries at different abstract levels can be expressed. An example query is illustrated in Fig. 12. This query aims at retrieving indexed persons that are close to vending machine 1 and similar to a sub-image. The matching method has proved its performance in comparison with two other methods in the state of the art.

The complete approach has been validated on two video databases coming from two projects: CARETAKER and CAVIAR. Videos of the CARETAKER project are analyzed by the VSIP platform of Pulsar team while videos coming from CAVIAR project are manually annotated. Experiments have shown how the proposed approach is efficient and robust to retrieve the objects of interest and the complex events from surveillance videos. A part of this work has been published in [28], [29].

6.1.15. Knowledge Discovery for Activity Extraction in Video Data

Keywords: *Data-mining*, *Learning*, *Understanding*.

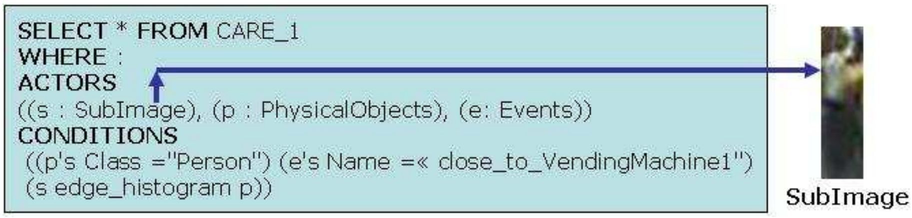

Figure 12. Expression of query "Finding persons that are close to the vending machine 1 and similar to a sub-image" in our query language

Participants: José Luis Patino Vilchis, Etienne Corvée, Guido Pusiol, François Brémond, Monique Thonnat.

We have developed an approach for agglomerative clustering of trajectories from mobile objects detected in videos in order to provide meaningful information on people behaviours and interactions with contextual (static) objects in the scene. This approach has been presented in [14], [15], [2] where agglomerative clustering techniques were employed to find patterns of trajectories and patterns of activity on real videos from two metro stations, one in Torino, the other in Roma, Italy.

The clustering approach is characterised by :

- Optimizing trajectory clustering with the use of optimal weights;
- On-line learning of trajectory clusters to process large amounts of data ;
- Computation of statistics for long-time recordings.

1) Optimizing trajectory clustering with the use of optimal weights. Each trajectory is composed by a set of six features. F1: the entry point; F2: the exit point; F3: the direction; F4: the sum of the distance between all subsequent pairs of points; F5: the distance from entry to exit; F6: the mean variation of the angle between all subsequent pairs of points.

By giving different weights to each feature, the final partition of the trajectory clusters will be changed. The optimal weights are those that allow the clustering algorithm to better match an user-defined behavior ground truth. The ground truth represents the expected paths between zones of interest in the scene and corresponding to behaviors predefined by users such as, 'From Vending machines to gates', 'From south doors to gates', 'etc'... A cost function is defined to evaluate the agreement between the clustering results and the ground-truth. The cost function is mainly made up of three performance measures: **TruePositive** is the number of correct associations between resulting Clusters and Ground-truth, **FalsePositive** is the number of resulting Clusters that could not be associated to the Ground-truth, **Dispersion** indicates the number of ground-truth trajectories that have been distributed into different Clusters. In an iterative way, a genetic-algorithm optimization minimizes the cost function. The experimental results have shown that the proposed technique helps to minimize dispersion. Minimization of Dispersion not only enables to merge similar behaviors, but also helps to split clusters that do not correspond to the same behavior [35].

2) On-line learning of trajectory clusters to process large amounts of data. All the behaviours cannot be exhaustively and a priori defined. Thus online learning capabilities turn out to be compulsory when dealing with human interaction monitoring and behaviour characterization. We have addressed such adaptive capability by implementing a competitive neural network approach. The algorithm constructs a partition of the input data, and a leading prototype for each cluster, such that every element in a cluster is within a distance T of the leading prototype and the gradient descent technique is used to optimize this parameter. The algorithm makes one pass through the data basically with two alternatives:

1. assigning each element to the cluster whose leading prototype is the closest;

2. making a new cluster, and a new leading prototype, for cases that are not close to any existing leader.

This approach has allowed to process 14 mornings of video in the Roma metro, that is: 77 hours of video, 12000 detected mobiles, 18000 detected video events. Experimental results over extracted trajectories allowed to identify 133 different trajectory types. The most common trajectories are *From main entry to gates* (19% from dataset), *From gates to platform* (9% from dataset). The most rare trajectories are *From main entry to exit gates* (less than 1% from dataset), *From hall to exit gates* (less than 1% from dataset).

3) Computation of statistics for long-time recordings. Statistical information is obtained from the attributes of the mobile objects and the contextual objects as well as their interactions. For the current application (monitoring of two metro stations) we have calculated four main statistics: **a. number of users:** the total number of people interacting with a contextual object. **b. percentage of use:** the ratio between the total period of time a contextual object is in use to the total observation time. **c. interaction duration:** the mean time a user spends when interacting with a contextual object. **d. mobile object histogram:** gives the number of appearances for all involved mobile objects per object type. The detailed description of the statistics computation can be found in [38], [24]. As an illustrative example, the statistics for 14 mornings of video show, according to the number of users, similar activities between weekdays but significant difference between weekdays / Sunday. A week day contains in average 120 people every 5 min, a Sunday contains less than 50 people observed in the same period. Also people are quicker to travel through the hall when the station is busy: people take 10 to 25s (6h-7h) and less than 10s (7h-11h30).

This work was evaluated by Rome and Turin subway managers as well by European experts at the CARETAKER live demonstration in Rome on the 24th of September. The overall future work includes developing more trajectory features and further testing of their optimal weighting. We will also develop new evaluation criteria that will complement the dispersion measure to decide when a cluster is well created and whether a structural change must be done on the cluster hierarchy by merging or splitting of clusters.

6.2. Software Engineering for Activity Recognition

Participants: François Brémond, Edouardo Cruz Gonzalès, Germàn Delbianco, Erwan Demairy, Cecilia Diaz, Vivien Fighiera, Adrian Hoitan, Sabine Moisan, Anh-Tuan Nghiem, Domnica Pavel, Annie Ressouche, Jean-Paul Rigault, Alejandro Sanchez, Monique Thonnat, Christophe Tornieri, Marcos Zuniga.

6.2.1. Introduction

This year Pulsar has developed a new software platform: the SUP platform. It is the backbone of the team experiments to implement the new algorithms proposed by the team in perception, understanding and learning. We study a meta-modeling approach to support the development of video surveillance applications based on SUP. We also introduce the notion of relation in our knowledge description language. We continue to develop the CLEM toolkit around a synchronous language dedicated to activity recognition applications. We study the management of concurrent access to components in a component platform dedicated to the fast design of applications in the field of ubiquitous computing.

6.2.2. SUP Software Platform

Participants: Monique Thonnat, François Brémond, Erwan Demairy, Christophe Tornieri.

The Scene Understanding Platform (SUP) is the platform for the PULSAR project-team to recognize and learn activity based on video and sensors analysis. This platform has two main objectives:

- The software architecture of SUP allows new algorithms to be plugged in the platform, be architected in such a way that new algorithms can be seamlessly plugged.
- To help the valorisation of the scientific results by allowing them to be transfered to the Keeneo company.

The software architecture of SUP relies on a set of design patterns (factories, commands and strategies being the most prominents) in order to allow an easy substitution of one algorithm by another. From a video analysis point of view, the main algorithms are: (i) video stream acquisiton; (ii) segmentation; (iii) classification; (iv) frame-to-frame tracking; (v) long-term tracking; (vi) event recognition; (vii) learning.

Concerning offline algorithms, knowledge discovery and video indexing are in the process to be integrated.

The algorithms that can be used in SUP can come from the old platform of the project, from the Genius Keeneo library, or be new algorithms developed directly in SUP. In a second step, the SUP platform will be extended to include results from meta-modeling done in PULSAR in order to ease the configuration of the whole tool chain.

The SUP platform is currently validated by designing an application for early detection of plant diseases.

SUP is the backbone of the team experiments to design PULSAR applications based on video and multi-sensor analysis.

6.2.3. Meta-modeling for Video Surveillance

Participants: Edouardo Cruz Gonzalès, Cecilia Diaz, Sabine Moisan, Jean-Paul Rigault.

This year we started the COSIMO project, a COLOR collaboration between INRIA, the University of Nice (Rainbow team) and a startup company (Keeneo). The goal is to explore and enrich Model Driven Engineering (MDE) techniques and model transformations to support the development of product lines for domains presenting multiple variability factors. Video surveillance is a good candidate for which Pulsar has an extensive experience, both for specifying video surveillance tasks and implementing the corresponding computer systems. Of course, for the latter aspect, we target the SUP platform, a joint development between Pulsar and Keeneo 6.2.2.

Conforming with the model engineering approach, we need both a generic model of video surveillance applications (business model) and a model of video processing components and chains (platform model). Moreover, we also need to model automatic (or semi-automatic) transformations from the first model to the second. This general approach uses meta-modeling techniques and is represented in figure 13. A difficult issue is to handle variability factors both at application level (observation context, objects of interest, surveillance task, lighting conditions...) and at component assembly level (selection of algorithms, parameter tuning...). Yet, variability should not be expressed at the expense of genericity so that the results of the study may go beyond the video surveillance domain.

This year, we have defined a first partial version of both models using "feature diagrams" [46]; this representation is well suited to describe variability factors. We have also identified some transformation rules with the help of a video surveillance expert. Furthermore, we are developing a prototype of a generic feature diagram editor to enter the previous models, as a first step toward a graphic tool to assist the specification and design of video surveillance processing chains. The feature diagram editor is implemented using Eclipse meta-modeling facilities (EMF, Ecore, GMF...); model transformations will be implemented owing to an existing tool available within Eclipse (such as Kermeta, ATL...). Ultimately, as shown on figure 13, the user (a video surveillance expert) should be able to input a task specification and, after automatically applying model transformations, should get as output a set of valid solutions. For the time being, it will be the user's responsibility to extract an assembly of components to implement the final processing chain.

6.2.4. First-Class Relations in Knowledge Representation Languages

Participants: Sabine Moisan, Jean-Paul Rigault, Alejandro Sanchez.

We are working on a generic object-oriented component framework (BLOCKS) for knowledge-based systems (KBS) [32] supporting several kinds of reasoning tasks (classification, planning, resource allocation, video understanding, activity recognition...). The framework provides C++ components implementing basic concepts of KBS such as *frames* (sorts of classes), *slots* (structured attributes) and *daemons* (methods automatically triggered in case of slot or frame access), *inference rules*, and *inference engines*... It also allows the definition of *Domain Specific Languages* (DSL) that are used by experts to enter their knowledge. Once translated into

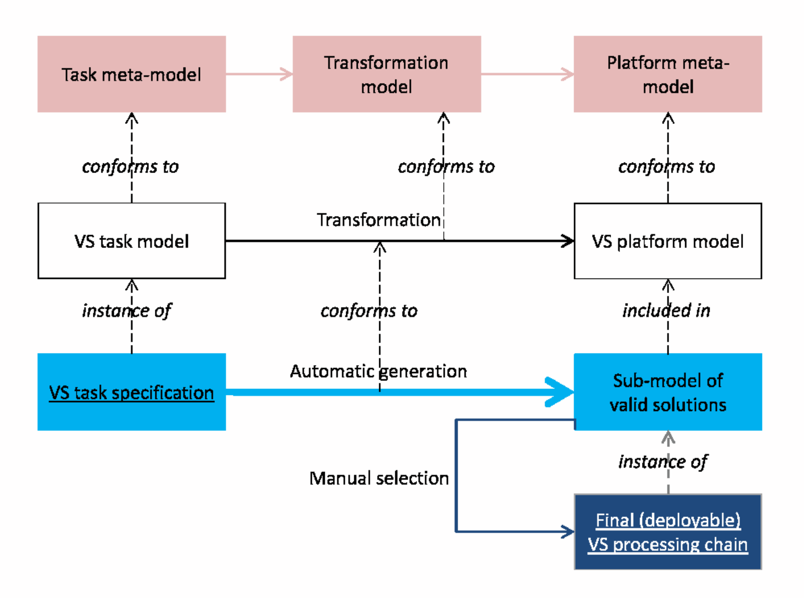

Figure 13. Model driven engineering approach for video surveillance

C++ code, this expertise is linked with the engine to yield an executable KBS for solving (non expert) end users' problems.

Relations among objects constitute an indispensible part of the knowledge. These relations are intrinsic to the domain and may present various natures, e.g., spatial relations between objects in a scene, animals and their habitats, customer and service reservation... Therefore, it is important for experts to declare and manipulate them directly rather than embedding them into object attributes. Hence, we decided to introduce relations as first class concepts in the expert languages as well as in the C++ components.

Two important properties of our KBSs are worth mentioning. First, experts and end users are not programmers, all the C++ code is automatically generated from expert knowledge, no customization is made at construction nor at execution time. Second, our systems are usually embedded and autonomous; they manipulate *dynamic* (even real time) objects. Thus the state of objects evolves with time during the KBS execution, which may challenge existing relation consistency.

We chose to rely on the mathematical notion of a "relation", for the abstract definition as well as for the implementation. For the time being, we limit ourselves to binary relations. Hence, the relation $\mathcal{R}$ appears as a set of tuples (here, pairs) (a, b) where $a \in A$ and $b \in B$. Therefore, we reify the relations themselves, not the individual tuples.

In some cases, a collection of tuples is not sufficient. Thus, we also offer experts the possibility to attach constraints to a relation, that is predicates that each tuple must satisfy. These constraints add extra *necessary* conditions for a tuple to be part of a relation. For instance, in video understanding, two objects can be considered "close to" each other provided that they are both visible (not occulted) in the scene.

We extended the knowledge representation languages so that experts can define and manipulate binary relations. A relation as a whole may have its own attributes, multiplicities on both ends, and be associated with constraint predicates. Once translated into C++, a relation type becomes a class which represents a collection of tuples. These classes automatically handle constraint verifications when a new tuple is inserted as well as when objects change their state.

In a KBS, the object states may change due to the engine reasoning activity. In particular, related objects can change state independently and in any order. The consequence is that the constraints of some relation tuples may become false, which invalidates the tuple itself. We are interested in those constraints that may be *temporarily* violated, provided that they hold at some system observation point. For instance, relation invalidity may be due to serialization of updates which are in fact logically concurrent. Since evolutions cannot be predicted by the expert nor by the engine, we must provide an automatic mechanism to cope with these relation inconsistencies.

To cope with these temporary inconsistencies, we have explored several solutions and we propose a flexible “check on demand” solution, associated with a temporary blacklisting and remedial strategies. When an object changes its state, we simply mark all relation tuples in which it is involved; the real check will only be performed when a marked tuple is accessed. Should the relation not hold, it is up to the expert to specify one among several “remedial strategies” (globally for all relations or specifically for some). The default strategy is to blacklist the invalid tuples until some expert rule decides what to do.

Introducing relations is an active line of research in the object oriented community. Our approach originality lies in handling dynamically checked constraints through blacklisting and expert parametrized strategies [36].

This year, we have defined an expert’s language to describe relations together with its parser and C++ code generator. The generated code uses a set of C++ classes which have been smoothly integrated into our BLOCKS framework. We are currently “reengineering” existing knowledge bases to take advantage of relations and to evaluate the expression power and performance issues.

6.2.5. The CLEM Workflow

Participants: Annie Ressouche, Germàn Delbianco.

The aim of this work is to develop a special purpose language (LE) devoted to activity recognition. This high level language allows to specify applications reacting to events coming from different sensors. This language is designed in collaboration with D. Gaffé (CNRS and UNSA). It agrees with Model Driven Software Developement philosophy which is now well known as a way to manage complexity, to achieve high re-use level, and to significantly reduce the development effort. Therefore, we benefit from a formal framework well suited to compilation and formal validation. In practice, we defined two semantics for LE: a *behavioral semantics* to define a program by the set of its behaviors, avoiding ambiguities in program interpretations; an *execution equational semantics* to allow **modular** compilation of programs into software and hardware targets (C code, VHDL code, FPGA synthesis, observers...). Our approach fulfills two main requirements of critical realistic applications: modular compilation to deal with large systems and model-based approach to perform formal validation.

The main originality of this work is to be able to manage both modularity and causality. Indeed, only few approaches consider a modular compilation because there is a deep incompatibility between *causality* and modularity. Causality means that for each event generated in a reaction, there is a causal chain of events leading to this generation. No causal loop may occur. Program causality is a well-known problem with synchronous languages, and therefore, it needs to be checked carefully. Thus, relying on semantics to compile a language ensures a modular approach but requires to complete the compilation process with a global causality checking. To tackle this problem, we introduced a new way to check causality from already checked sub programs and the modular approach we infer.

This year, we have improved the expressiveness of the language and now we offers three kinds of designs: (1) event-driven systems are specified with a textual synchronous language closed to Esterel language [44]; (2) state-chart-like description is a native construct and is well suited to the design of schedulers; (3) data flow applications can be described with the equation syntax the language offers. The language, its semantics and the algorithm we have to check causality with respect to our modular approach has been presented in [40]. The complete description of Clem workflow and the demonstration of equivalence of LE behavioral and equational

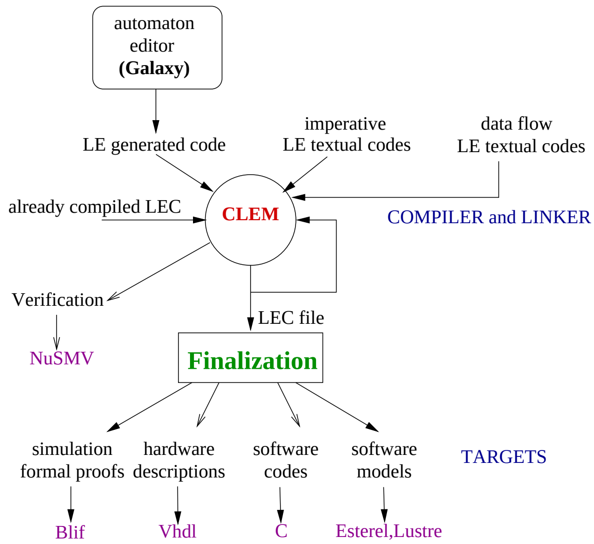

Figure 14. The CLEM Toolkit

semantics is detailed in [41]. This year, we have also improved the Clem toolkit with verification means that apply model checking techniques. In practice, we provide users with facilities to use the NuSMV [45] model checker. The Clem toolkit (see figure 14) has been presented in [27] and is now available for free download at: http://www.inria.sophia/pulsar/projects/Clem.

Finaly, during his internship, G. Delbianco has begun to improve LE with data handling. According to LE principle, only signals carry data. Data belong to predefined types (usual programming language types) or to some external types. Moreover, the type system definition supports two constructors: a union operator and a pair operator. German designed a complete type system "à la Milner" [49] and implements it in Haskell language [47]. To complete this work, a syntactic extension of LE language to support types is required. Moreover, the integration of the type checker in Clem must also be done.

6.2.6. Concurrent Acces Management in Component and Service Platforms

Participants: Annie Ressouche, Domnica Pavel, Vivien Fighiera.

This research axis concerns the management of concurrent access according to the synchronous hypothesis in component and service platforms with an application to web service composition. Indeed, this year we have initiated a collaboration with the Rainbow team (CNRS/UNSA) in the framework of a Color action on this topic. The design of applications by composition and assembly of components relies on the reuse of existing components (it is why libraries of reusable component by assembly and composition, appear). The assembly and composition of components require to obey the "usage contract" of components. But, the existing approaches don't really succeed to ensure that this usage contract is not violated during application designing. Only a formal analysis of the component behavior models associated with a well sound modelling of composition operation will allow us to secure the respect of the usage contract. Within the Rainbow project, a component platform (WComp) is dedicated to the fast design of applications in the domain of ubiquitous computing. The component oriented approach of WComp relies on three main paradigms: (1) a basic Web Services architecture; (2) a programming model of Web Services dynamic composition; (3) an adaptation mechanism associated with composite services. A challenge is to adapt the behavior of an application with respect to its execution context which permanently evolves. Some bounds appear during composition and adaptation of services, when unanticipated concurrent accesses to services can induce an unexpected behavior of the designed application. It is what we call “concurrent access conflicts”.

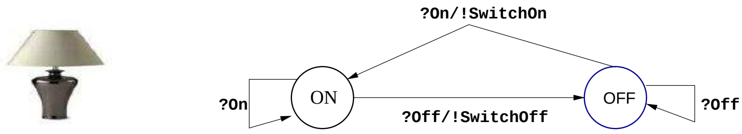

Figure 15. Component lamp behavior

This year, we studied the definition of a local synchronous model of component behavior as solutions to resolve these concurrent access conflicts to components and services. In practice, D. Pavel during her internship developed a synchronous language LS to express how access to components must be managed. Thus, we can build in WComp specific components that check the behavior described by LS programs. For instance, the figure 15 represents a component lamp and its expected behavior. We can describe this behavior as a LS program and a specific component will be generated to represent it. To tackle the problem of concurrent access, LS offers two operators on behaviors: a sequence and a synchronous parallel operators. Moreover, a mathematical semantics for LS has been introduced to build a model of LS programs that allow us to benefit from model checking techniques, to prove that the behavior described by a LS program verify expected properties (mainly safety properties).

Beyond the theoretical study, V. Fighiera did the practical improvement of the WComp platform to generate the specific components designed with LS programs. Actually, he is improving WComp with facilities to call the NuSMV [45] model checker in order to prove properties about the generated component behavior.

7. Contracts and Grants with Industry

7.1. Industrial Contracts

The Pulsar team has strong collaborations with industrial partners through European projects and national grants. In particular with RATP, STMicroelectronics, Bull, Thales, Sagem, Alcatel, Bertin, AKKA, Solid, Metro of Turin (GTT), Metro of Roma (ATAC) and Keeneo.

8. Other Grants and Activities

8.1. European Projects

Pulsar team has been involved this year in three European projects: a project on crowd behavior analysis (SERKET), a project on multimedia information retrieval (CARETAKER) and a project on machine learning and activity monitoring (COFRIEND).

8.1.1. SERKET Project

SERKET was a European ITEA project in collaboration with THALES R&D FR, THALES Security Syst, CEA, EADS and Bull (France); Atos Origin, INDRA and Universidad de Murcia (Spain); XT-I, Capvidia, Multitel ABSL, FPMs, ACIC, BARCO, VUB-STRO and VUB-ETRO (Belgium). It began at the end of November 2005 and ended in July 2008. The main objective of this project was to develop software engineering techniques for videosurveillance and in particular to analyze crowd behaviors and to help in terrorist prevention.

8.1.2. CARETAKER Project

CARETAKER was a STREP FP6 European project that began in March 2006 and ended in November 2008. The main objective of this project was to discover information in multimedia data. The prime partner was Thales Communications (France) and other partners were: Multitel (Belgium), Kingston University (UK), IDIAP (Switzerland), Roma ATAC Transport Agency (Italy), Metro of Turin (GTT), SOLID software editor for multimedia data basis (Finland), and Brno University of Technology (Czechia). Our team was in charge of modeling, recognizing and learning scenarios for frequent or unexpected human activities from both video and audio events.

8.1.3. COFRIEND Project

COFRIEND is a European project in collaboration with Akka, University of Hamburg (Cognitive Systems Laboratory), University of Leeds, University of Reading (Computational Vision Group), Toulouse-Blagnac Airport. It has begun at the beginning of February 2008 and will last 3 years. The main objectives of this project is to develop techniques to recognise and learn automatically all servicing operations around aircraft parked on aprons.

8.2. International Grants and Activities

Pulsar is involved in an academic collaboration with MICA and University of Hanoi in Vietnam (a joint PhD will be defended at the beginning of 2009).

8.2.1. Joint Partnership with Vietnam

Pulsar has been cooperating with the Multimedia Research Center in Hanoi MICA on semantics extraction from multimedia data. Currently we continue through a joint supervision by A. Boucher and M. Thonnat of Lan Le Thi 's PhD on video retrieval (funded by an AUF grant).

8.3. National Grants and Collaborations

Pulsar Team has six national grants: the two first ones concern the implication of the team in a "pôle de compétitivité" and an ANR project on videosurveillance. We continue both our collaboration with INRA and our collaboration with STmicroelectronics. A collaboration in homecare domain involves a PhD student.

8.3.1. SYSTEM@TIC SIC Project

Pulsar is strongly involved in SYSTEM@TIC "pôle de competivité" and in particular in the **SIC** project Sécurité des Infrastructures Critiques ([43], [42]) which is a strategic initiative in perimeter security. More precisely the **SIC** project is funded for 42 months with the industrial partners including Thales, EADS, BULL, SAGEM, Bertin, Trusted Logic.

8.3.2. VIDEO-ID Project

Pulsar is participating to an ANR research project on intelligent video surveillance and people biometrics. The project lasts 3 years and will be over on February 2011. The involved partners are: Thales-TSS, EURECOM, TELECOM and Management SudParis, UIC, RATP, DGA, STSI

8.3.3. BioSERRE: Video Camera Network for Early Pest Detection in Greenhouses

Pulsar cooperates with Vista (INRIA Rennes - Bretagne Atlantique), INRA Avignon UR407 Pathologie Végétale and CREAT Research Center (Chambre d'Agriculture des Alpes Maritimes) in an ARC project (BioSERRE) for early detection of crop pests, based on video analysis and interpretation.

8.3.4. MONITORE: Real-time Monitoring of TORE Plasma

Pulsar is involved in an Exploratory Action called MONITORE for the real-time monitoring of imaging diagnostics to detect thermal events in tore plasma. This work is a preparation for the design of the futur ITER nuclear reactor and is done in partnership with the Imaging and Diagnosics Group of the CEA Cadarache.

8.3.5. Intelligent Cameras

This year Pulsar has completed a cooperation with STmicroelectronics. A PhD thesis on the design of intelligent cameras including gesture recognition algorithms is on going (Mohamed Becha).

8.3.6. Long-term People Monitoring at Home

Pulsar has a collaboration with CSTB (Centre Scientifique et Technique du Bâtiment) and the Nice City Hospital (Groupe de Recherche sur la Tophicité et le Viellissement) in the GER'HOME project, funded by the General Council 06. GER'HOME project is devoted to experiment and develop techniques that allow long-term monitoring of people at home. In this project an experimental home has been built in Sophia Antipolis and is relying on the research of the Pulsar team concerning unsupervised event learning and recognition.

8.3.7. Semantic Interpretation of 3D Sismic Images by Cognitive Vision Techniques

A cooperation took place with IFP (French Petrol Institute) and Ecole des Mines de Paris in the framework of a joint supervision by M Thonnat and M. Perrin of Philippe Verney PhD at IFP. The topic is Semantic Interpretation of 3D sismic images by cognitive vision techniques.

8.4. Local Collaborations

This year Pulsar team has two local collaborations in the framework of the COLOR (COoperation LOcale de Recherche) program: Cosimo and SynComp.

8.4.1. Synchrony in Component and Services Platforms

This year PULSAR has a collaboration with the Rainbow team (CNRS/UNSA) and the CMA (Ecole des Mines) in the framework of a COLOR action *SynComp* (http://www.inria.fr/sophia/pulsar/projects/SynComp). In this project we study the concurrent access management according to the synchronous hypothesis in component and service platforms (see section 6.2.6).

8.4.2. Meta-modeling for Video Surveillance

This year we started COLOR action *Cosimo* (COnfiguration Sûre d'applications de vision par l'Ingénierie des MOdèles), a collaboration between Pulsar, the University of Nice (Rainbow team) and a startup company (Keeneo). The goal is to study the possible contributions of model-driven engineering to the implementation of vision product lines (see section 6.2.3).

8.5. Patent

This year Pulsar has contributed to a patent filled in June 2008. This latter concerns a method to build a geological model relying on a cognitive vision technique. The inventors are: P. Verney (IFP) , J.F. Rainaud (IFP) , M. Perrin (Mines Paris) and M. Thonnat (INRIA).

8.6. Spin off Partner

Keeneo (http://www.keeneo.com) is a spin off of the Orion/Pulsar team which aims at commercializing video surveillance solutions. This company has been created in July 2005 with six co-founders from the Orion/Pulsar team and one external partner. We have a joint collaboration for building the new Pulsar SUP platform in the framework of an ADT (Action de Développement Technologique).

9. Dissemination

9.1. Scientific Community

- M. Thonnat is member of the editorial board of the journal Image and Vision Computing (IVC), and is co-editor of a special issue in the journal CVIU Computer Vision and Image Understanding.
- M. Thonnat is a reviewer for the journals PAMI (IEEE Transactions - Pattern Analysis and Machine Intelligence), CVIU Computer Vision and Image Understanding, and Eurasip Journal on Image and Video Processing.
- M. Thonnat is a Program Committee member for the following conferences: ECAI 08, ICVS08, ICPR08, CVPR09.
- M. Thonnat is an expert to review a Cognitive System project IST eTRIMMS for the European Commission, and for the French governement DGA.
- M. Thonnat is reviewer for the following theses: Frederic Lerasle (HDR, Laas Toulouse), Simon Conseil (Univ, Aix-MarseilleIII)), Suphot Chunwiphat (INP Grenoble), Pierre Maurel (ENS Paris).
- M. Thonnat is member of the scientific board of ENPC, Ecole Nationale des Ponts et Chaussées since june 2008.
- M. Thonnat is member of the INRIA Evaluation board since 2003.
- M. Thonnat is member of the scientific board of INRIA Sophia Antipolis (bureau du comité des projets) since September 2005.
- M. Thonnat had an invited talk at the Conference ECAI 2008, European Conferrence on Artificial Intelligence, Patras, Greece, July 2008.
- M. Thonnat and F. Brémond are co-founders and scientific advisors of Keeneo, the videosurveillance start-up created to exploit their research results on the VSIP/SUP software.
- F. Brémond is an ANR reviewer for the 2008 edition of Project Call: "Concepts, Systèmes et Outils pour la Sécurité Globale".
- F. Brémond is an Expert for EC INFSO in the framework of Ambient Assisted Living FP7.
- F. Brémond is reviewer for the journals: PAMI Transactions on Pattern Analysis and Machine Intelligence, CVIU Computer Vision and Image Understanding, PR Pattern Recognition, Transactions on Systems Man and Cybernetics, IJPRAI International Journal of Pattern Recognition and Artificial Intelligence, MVA Machine Vision and Applications Journal, IMAVIS Image and Vision Computing, The Open Cybernetics and Systemics Journal, JPRR Journal of Pattern Recognition Research, IVP EURASIP Journal on Image and Video Processing, IEEE Transactions on Multimedia, TKDE Transactions on Knowledge and Data Engineering, MEP Medical Engineering and Physics, IET Computer Vision, and Artificial Intelligence Journal.
- F. Brémond is Program Committee member of M2SFA2 2008 ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion, ICVS 2008-2009 International Conference on Computer Vision Systems, VNBA 2008 ACM International Workshop on Vision Networks for Behavior Analysis, AVSS 2008 IEEE Conference on Advanced Video and Signal based Surveillance, VS 2008 The Eighth IEEE International Workshop on Visual Surveillance, IWINAC 2009 3th International Work Conference on the Interplay between Natural and Artificial Computation, and BMVC 2008 British Machine Vision Conference.
- F. Brémond is a reviewer for the conferences and workshops: CVPR 2008-2009 Computer Vision and Pattern Recognition, INDIN 2008 6th IEEE International Conference on Industrial Informatics.
- F. Brémond had an invited Talk at VS 2008 (Visual Surveillance workshop, part of ECCV 2008, Marseille, Oct. 2008), at THEMIS 2008 (Tracking Humans for the Evaluation of their Motion in Image Sequences, part of BMVC 2008, Leeds, Sept. 2008) and at AViRS 2008 (Analyse Video pour le Renseignement et la Sécurité, Paris, April 2008)

- F. Brémond was a reviewer for the PhD defense of Mr Claudio Piciarelli at Udine University (Italy).
- F. Brémond is organizer of the GDR-ISIS day on Intelligent Video Surveillance, 17th of Dec. 2008, Paris.
- Sabine Moisan is a member of the Scientific Council of INRA for Applied Computing and Mathematics (MIA Department).
- Jean-Paul Rigault is a member of AITO, the steering committee for several international conferences including in particular ECOOP. He is also a member of the Administration Board of the Polytechnic Institute of Nice University.
- A. Ressouche is a member of the Inria Cooperation Locales de Recherches (Colors) committee.
- G. Charpiat is a reviewer for the journals: the International Journal of Computer Vision (IJCV), Transactions on Pattern Analysis and Machine Intelligence (TPAMI), the Journal of Mathematical Imaging and Vision (JMIV), Computer Vision and Image Understanding (CVIU), Image and Vision Computing Journal (IVC), Proceedings of the Royal Society, and IEEE Transactions on Image Processing (TIP).
- G. Charpiat is also a reviewer for the conference MICCAI 2008 : 11th International Conference on Medical Image Computing and Computer Assisted Intervention.

9.2. Teaching

- Pulsar is a hosting team for the master of Computer Science of UNSA.
- Teaching at Master EURECOM on Video Understanding (3h F. Bremond).
- Teaching at Master of Computer Science at Polytechnic School of Nice Sophia Antipolis University, course on Reactive Systems (24 h A. Ressouche).
- Jean-Paul Rigault resumed his teaching as a full professor at the Polytechnic School of Nice Sophia Antipolis University (Computer Science Department).

9.3. Thesis

9.3.1. Thesis in Progress

- Duc Phu Chau: Object Tracking for Activity Recognition, Nice Sophia-Antipolis University.
- Mohamed Bécha Kaâniche : Reconnaissance de gestes à partir de séquences videos, Nice Sophia-Antipolis University.
- Lan Le Thi : Semantic-based Approach for Image Indexing and Retrieval, Nice Sophia-Antipolis University and Hanoi University (Vietnam).
- Guido Pusiol : Learning Techniques for Video Understanding, Nice Sophia-Antipolis University.
- Anh Tuan Nghiem : Learning Techniques for the Configuration of the Scene Understanding Process, Nice Sophia-Antipolis University.
- Nadia Zouba : Multi Sensor Analysis for Homecare Monitoring, Nice Sophia-Antipolis University.

9.3.2. Thesis Defended

- Marcos Zúñiga : Unsupervised Primitive Event Learning and Recognition in Video, Nice-Sophia Antipolis University (defended on the 28th of November 2008).

10. Bibliography

Major publications by the team in recent years

[1] A. AVANZI, F. BRÉMOND, C. TORNIERI, M. THONNAT. *Design and Assessment of an Intelligent Activity Monitoring Platform*, in "EURASIP Journal on Applied Signal Processing, Special Issue on “Advances in Intelligent Vision Systems: Methods and Applications”", vol. 2005:14, 08 2005, p. 2359-2374.

[2] H. BENHADDA, J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Data Mining on Large Video Recordings*, in "5eme Colloque Veille Stratégique Scientifique et Technologique VSST 2007, Marrakech, Marrocco", 21st - 25th October 2007.

[3] B. BOULAY, F. BRÉMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System.*, in "Pattern Recognition Letter.", vol. 27, n^o 15, 2006, p. 1785-1796.

[4] F. BRÉMOND, M. THONNAT. *Issues of Representing Context Illustrated by Video-surveillance Applications*, in "International Journal of Human-Computer Studies, Special Issue on Context", vol. 48, 1998, p. 375-391.

[5] N. CHLEQ, F. BRÉMOND, M. THONNAT. *Image Understanding for Prevention of Vandalism in Metro Stations*, in "Advanced Video-based Surveillance Systems", Kluwer A.P. , Hangham, MA, USA, November 1998, p. 108-118.

[6] V. CLÉMENT, M. THONNAT. *A Knowledge-Based Approach to Integration of Image Procedures Processing*, in "CVGIP: Image Understanding", vol. 57, n^o 2, March 1993, p. 166–184.

[7] F. CUPILLARD, F. BRÉMOND, M. THONNAT. *Tracking Group of People for Video Surveillance*, Video-Based Surveillance Systems, vol. The Kluwer International Series in Computer Vision and Distributed Processing, Kluwer Academic Publishers, 2002, p. 89-100.

[8] B. GEORIS, F. BRÉMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", vol. 18, 2007.

[9] C. LIU, P. CHUNG, Y. CHUNG, M. THONNAT. *Understanding of Human Behaviors from Videos in Nursing Care Monitoring Systems*, in "Journal of High Speed Networks", vol. 16, 2007, p. 91-103.

[10] S. LIU, P. SAINT-MARC, M. THONNAT, M. BERTHOD. *Feasibility Study of Automatic Identification of Planktonic Foraminifera by Computer Vision*, in "Journal of Foramineferal Research", vol. 26, n^o 2, April 1996, p. 113–123.

[11] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", vol. 16, n^o 1, December 2004, p. 33-40.

[12] S. 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, April 1998.

[13] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Blocks, a Component Framework with Checking Facilities for Knowledge-Based Systems*, in "Informatica, Special Issue on Component Based Software Development", vol. 25, n^o 4, November 2001, p. 501-507.

[14] J. PATINO, H. BENHADDA, E. CORVEE, F. BRÉMOND, M. THONNAT. *Video-Data Modelling and Discovery*, in "4th IET International Conference on Visual Information Engineering VIE 2007, London, UK", 25th - 27th July 2007.

[15] J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Management of Large Video Recordings*, in "2nd International Conference on Ambient Intelligence Developments AmI.d 2007, Sophia Antipolis, France", 17th - 19th September 2007.

[16] M. THONNAT, M. GANDELIN. *Un système expert pour la description et le classement automatique de zooplanctons à partir d'images monoculaires*, in "Traitement du signal, spécial I.A.", vol. 9, n^{o} 5, November 1992, p. 373–387.

[17] M. THONNAT, S. MOISAN. *What Can Program Supervision Do for Software Re-use?*, in "IEE Proceedings - Software Special Issue on Knowledge Modelling for Software Components Reuse", vol. 147, n^{o} 5, 2000.

[18] M. 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, October 2003.

[19] M. THONNAT. *Toward an Automatic Classification of Galaxies*, in "The World of Galaxies", Springer Verlag, 1989, p. 53-74.

[20] A. TOSHEV, F. BRÉMOND, 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.

[21] V. T. VU, F. BRÉMOND, M. THONNAT. *Temporal Constraints for Video Interpretation*, in "Proc of the 15th European Conference on Artificial Intelligence, Lyon, France", 2002.

[22] V. T. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)", 9-15 September 2003.

Year Publications

Articles in International Peer-Reviewed Journal

[23] M. HOFMANN, F. STEINKE, V. SCHEEL, G. CHARPIAT, J. FARQUHAR, P. ASCHOFF, M. BRADY, B. SCHÖLKOPF, B. J. PICHLER. *MR-based Attenuation Correction for PET/MR: A Novel Approach Combining Pattern Recognition and Atlas Registration*, in "Journal of Nuclear Medicine", 11 2008.

[24] J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Extraction of Activity Patterns on Large Video Recordings*, in "Computer Vision", vol. 2, 2008, p. 108 - 128.

International Peer-Reviewed Conference/Proceedings

[25] S. BAK, S. SURESH, F. BRÉMOND, E. CORVEE, M. THONNAT. *Fusion of motion segmentation and learning based tracker for visual surveillance*, in "The International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), Lisbon, Portugal", 5 - 8 February 2008.

[26] G. CHARPIAT, M. HOFMANN, B. SCHÖLKOPF. *Automatic Image Colorization via Multimodal Predictions*, in "10th European Conference on Computer Vision, Berlin, Germany", D. FORSYTH, P. TORR, A. ZISSERMAN (editors), Springer, 10 2008, p. 126-139, http://eccv2008.inrialpes.fr/.

[27] D. GAFFÉ, A. RESSOUCHE. *The Clem Toolkit*, in "Proceedings of 23rd IEEE/ACM International Conference on Automated Software Engineering (ASE 2008), L'aquila, Italy", September 2008.

[28] T.-L. LE, A. BOUCHER, M. THONNAT, F. BRÉMOND. *A Framework for Surveillance Video Indexing and Retrieval*, in "International Workshop on Content-Based Multimedia Indexing (CBMI'08), UK", 18-20 June 2008, p. 338-345.

[29] T.-L. LE, M. THONNAT, A. BOUCHER, F. BRÉMOND. *A Query Language Combining Object Features and Semantic Events for Surveillance Video Retrieval*, in "The International MultiMedia Modeling Conference (MMM'08), Kyoto, Japan", 9-11 January 2008, p. 307-317.

[30] V. MARTIN, S. MOISAN. *Early Pest Detection in Greenhouses*, in "Visual Observation and Analysis of Animal and Insect Behavior (VAIB) Workshop at ICPR, Tampa, Fl. USA", December 2008.

[31] V. MARTIN, S. MOISAN, B. PARIS, O. NICOLAS. *Towards a Video Camera Network for Early Pest Detection in Greenhouses*, in "International Conference Endure 2008, La Grande Motte, France", October 2008.

[32] S. MOISAN. *Component-Based Support for Knowledge-Based Systems*, in "ICEIS, 10th International Conference on Enterprise Information Systems, Barcelona, Spain", June 2008.

[33] A.-T. NGHIEM, F. BRÉMOND, M. THONNAT. *Shadow Removal in Indoor Scenes*, in "IEEE International Conference on Advanced Video and Signal based Surveillance, Santa Fe, USA", 1-3 September 2008.

[34] C. D. PHU, F. BRÉMOND, E. CORVEE, M. THONNAT. *Repairing People Trajectories Based on Point Clustering*, in "The International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), Lisbon, Portugal", 5 - 8 February 2008.

[35] G. PUSIOL, J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT, S. SURESH. *Optimizing Trajectories Clustering for Activity Recognition*, in "1st International Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA 08), Marseille, France", 12th October - 18th October 2008.

[36] A. SANCHEZ, S. MOISAN, J.-P. RIGAULT. *Dynamic First-Class Relations for Knowledge-Based Systems*, in "Workshop on Relationships and Associations in Object-Oriented Languages (RAOOL), Nashville, TN, USA", http://www.mcs.vuw.ac.nz/raool/, OOPSLA 2008, October 2008.

[37] N. ZOUBA, F. BRÉMOND, M. THONNAT. *Monitoring Activities of Daily Living (ADLs) of Elderly Based on 3D Key Human Postures*, in "In the 4th International Cognitive Vision Workshop (ICVW 2008)", 12-15 May 2008.

National Peer-Reviewed Conference/Proceedings

[38] J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Data Mining for Activity Extraction in Video Data*, in "Extraction et Gestion des Connaissances EGC 2008, Sophia Antipolis, France", 29th January - 1st February 2008.

Scientific Books (or Scientific Book chapters)

[39] G. CHARPIAT, M. HOFMANN, B. SCHÖLKOPF. *Kernel Methods in Medical Imaging*, in "Handbook of Biomedical Imaging, Berlin, Germany", N. PARAGIOS, J. DUNCAN, N. AYACHE (editors), Springer, 12 2008, http://www.springer.com/computer/computer+imaging/book/978-0-387-09748-0.

[40] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, in "Software Engineering Research, Management and Applications", R. LEE (editor), Studies in Computational Intelligence, selected as one of the 17 best papers of SERA'08 conference, vol. 150, Springer, 2008, p. 157-171.

Research Reports

[41] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, Research Report, n^{o} 6424, INRIA, 01 2008, http://hal.inria.fr/inria-00213472.

[42] V. VALENTIN, F. BRÉMOND. *Algorithmes de détection de foule et de ses caractéristiques*, Research Report, SIC Internal Report, 10 2008.

[43] V. VALENTIN, F. BRÉMOND. *Algorithmes de reconnaissance de posture*, Internal Report, SIC Internal Report, 10 2008.

References in notes

[44] G. BERRY. *The Foundations of Esterel*, in "Proof, Language, and Interaction, Essays in Honor of Robin Milner", G. PLOTKIN, C. STEARLING, M. TOFTE (editors), MIT Press, 2000.

[45] A. CIMATTI, M. PISTORE, M. ROVERI. *NuSMV home page*, http://nusmv.irst.itc.it.

[46] K. CZARNECKI, U. W. EISENECKER. *Generative Programming: Methods, Tools, and Applications*, Addison Wesley, 2000.

[47] G. DELBIANCO. *tyLE: a Typed Extension to the LE Language*, Masters thesis, INRIA Sophia Antipolis Méditerranée, 2008.

[48] ECVISION. *A Research Roadmap of Cognitive Vision*, Technical report, n^{o} IST-2001-35454, IST Project, 2005.

[49] B. PIERCE. *Types and Programming Languages*, MIT Press, Cambridge, MA USA, 2002.

[50] T. STAHL, M. VOELTER. *Model-driven Software Development: Technology, Engineering, Management*, Wiley, 2006.