INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Pulsar

Perception Understanding Learning Systems for Activity Recognition

Sophia Antipolis - Méditerranée

Theme : Vision, Perception and Multimedia Understanding

Activity Report

2010

Table of contents

Information related to Pulsar here (Pulsars ?)

1. Team

Research Scientists

François Brémond [Team Leader, DR2 Inria, HdR]
Guillaume Charpiat [CR1 Inria]
Sabine Moisan [CR1 Inria, HdR]
Annie Ressouche [CR1 Inria]
Monique Thonnat [DR1 Inria, HdR]

Faculty Member

Jean Yves Tigli [Assistant professor, Faculty Member Nice Sophia-Antipolis University, on secondment up to September 2010]

External Collaborators

Daniel Gaffé [Assistant professor, Faculty Member Nice University and CNRS-LEAT Member]
Jean-Paul Rigault [Professor, Faculty Member Nice Sophia-Antipolis University]
Philippe Robert [Professor, CHU Nice University]
Jean Yves Tigli [Assistant professor, Faculty Member Nice Sophia-Antipolis University, from September 2010]
Christophe Tornieri [Keeneo, up to August 2010]

Technical Staff

Bernard Boulay [Development Engineer, COFRIEND and QUASPER Projects]
Etienne Corvée [Development Engineer, COFRIEND, VANAHEIM and VideoID Projects]
Erwan Demairy [IR2 INRIA for the SUP Software Platform's ADT, up to October 2010]
Chedli Farhani [Development Engineer, VICOMO Project]
Julien Gueytat [Development Engineer, SWEET HOME Project]
Jose-Luis Patino Vilchis [Development Engineer, COFRIEND and VANAHEIM Projects]
Leonardo Rocha [Development Engineer, CIU Santé, SWEET HOME and VICOMO Projects]
Sofia Zaidenberg [Development Engineer, VANAHEIM Project]
Nadia Zouba [Development Engineer, VICOMO and COFRIEND Projects from February 2010 to November 2010]
Daniel Zullo [Development Engineer, CIU Santé and VICOMO Projects]

PhD Students

Slawomir Bak [Nice Sophia-Antipolis University, VideoID Grant]
Piotr Bilinski [Nice Sophia-Antipolis University, Paca Grant]
Duc Phu Chau [Nice Sophia-Antipolis University, Paca Grant]
Emmanuel Mulin [Nice Sophia-Antipolis University, CHU Nice]
Anh-Tuan Nghiem [Nice Sophia-Antipolis University, SIC Project, up to August 2010]
Guido-Tomas Pusiol [Nice Sophia-Antipolis University, CORDIs]
Rim Romdhame [Nice Sophia-Antipolis University, CIU Santé Project]
Malik Souded [Nice Sophia-Antipolis University, Keeneo CIFRE Grant]
Nadia Zouba [Nice Sophia-Antipolis University, up to February 2010]

Post-Doctoral Fellows

Ikhlef Bechar [ARC Bioserre, up to december 2010]
Mohammed-Bécha Kaâniche [up to June 2010]
Vincent Martin [Arc Monitore]
Sagar Sen [Project VICOMO, from May 2010]

Administrative Assistant

Christine Claux [AI Inria]

Others

Tanvi Barnajee [Internship EGIDE Inria from May 2010 up to August 2010]
Oscar Carrillo [Internship EGIDE Inria from March 2010 up to September 2010]
Ezequiel Cura [Internship EGIDE Inria from February 2010 up to August 2010]
Saurabh Goyal [Internship EGIDE Inria from May 2010 up to August 2010]
Santiago Hurtado [Internship EGIDE Inria from June 2010 up to December 2010]
Ratnesh Kumar [Internship ARC Bioserre from January 2010 up to July 2010]
Zouhair Ramadhane [Internship EGIDE Inria up to June 2010]
Silviu Serban [Internship from September 2010]
Anja Schnaars [Internship ANR Monitore up to March 2010]

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 semantic 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* (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, National Taiwan University (Taiwan), MICA (Vietnam), IPAL, I2R, NUS (Singapore), University of Southern California, University of South Florida, University of Maryland (USA).

2.2. Highlights

Pulsar is a Project team created in January 2008 in the continuation of the Orion project. Since 2008, we have started designing a new software platform called SUP (Scene Understanding Platform) for activity recognition based on sound software engineering paradigms. We have now (October 2010) a new operational SUP platform detecting and tracking mobile objects, which can be either humans or vehicles, and recognizing their behaviours. This SUP platform is the backbone of the team experiments to implement the new algorithms proposed by the team in perception, understanding and learning. We have studied a meta-modeling approach to support the development (e.g. specification) of video understanding applications based on SUP.

This year, we have also designed an algorithm for detecting mobile objects including a new background subtraction algorithm and a new shadow removal algorithm which both compute a fine distribution of pixel values for foreground and background images. We have started to improve the event recognition by handling event uncertainty. We have also continued original work on learning techniques such as data mining in large multimedia databases based on offline trajectory clustering. For instance, we have been able to learn six everyday activities (e.g. eating, preparing meal) of six elderly people living in the Ger'home laboratory by processing in an unsupervised way more than 6 x 4 hours of video recordings.

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

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 supervision 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 guarantee 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. Moreover 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 Machines) 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 recognize meaningful semantic 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

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 The Orion/Pulsar spin off 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

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 [13], 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.

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, when 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.

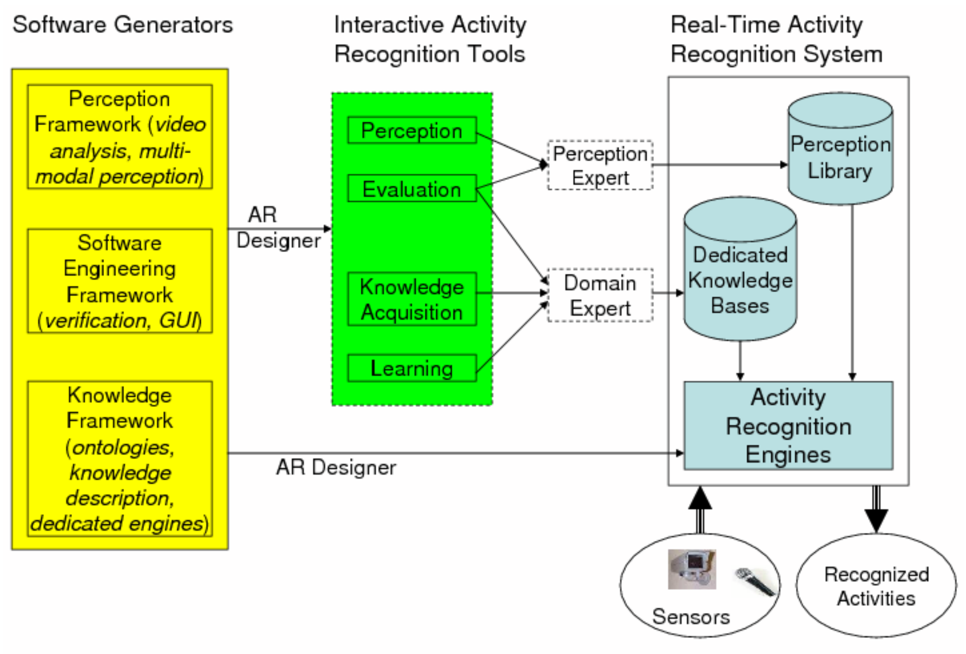


Figure 1. Global Architecture of an Activity Recognition Platform.

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) [75] 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.

More specifically, we first wish to use MDE to assist video-surveillance designers as well as software developers in configuring a video system before and during its deployment. A second challenge is to exploit models at runtime to cope with the dynamic aspects of video surveillance systems. Both approaches rely on software models, in our case *feature diagrams* [68], [62], [76], [58] and model transformations.

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 [12]. 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. 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. 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

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 in transport domain, applications in biologic domain or applications in medical domain.

4.2. Video Surveillance

The growing feeling of insecurity among the population led private companies as well as 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, AVITRACK CARETAKER, SERKET and CANTATA and more recently, European projects VICoMo 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, Barcelona, 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 [63]. 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 interesting *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, 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 at supporting 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 Général des Alpes Marimes (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. We have also started collaboration with Nice hospital to monitor Alzheimer patients with the help of geriatric doctors. The aim of the project is to design an experimental platform, providing services and allowing to test their efficiency.

5. Software

5.1. SUP

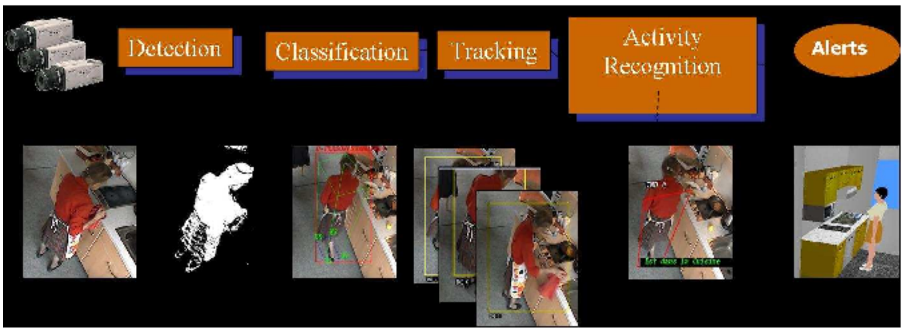


Figure 2. Components of the Scene Understanding Platform (SUP).

SUP is a Scene Understanding Software Platform written in C and C++ (see Figure 2). SUP is the continuation of the VSIP platform. SUP is splitting the workflow of a video processing into several modules, such as acquisition, segmentation, etc., until scenario recognition. Each module has a precise interface, and different plugins implementing these interfaces can be used for each step of the video processing. This generic architecture is designed to facilitate:

1. integration of new algorithms in SUP;
2. sharing of the algorithms among the team.

Currently, 15 plugins are available, covering the whole processing chain. Several plugins are using the Genius platform, an industrial platform based on VSIP and exploited by **Keeneo**, the Orion/Pulsar spin off created in July 2005.

Goals of SUP are twofolds:

1. From a video understanding point of view, to allow the researchers of the Pulsar team can share the implementations of their researches through this platform.
2. From a software engineering point of view, to integrate the results of the dynamic management of the applications when applied to video surveillance.

5.2. ViSEval

ViSEval is a software dedicated to the evaluation and visualisation of video processing algorithm outputs. The evaluation of video processing algorithm results is an important step in video analysis research. In video processing, we identify 4 different tasks to evaluate: detection of physical objects of interest, classification of physical objects of interest, tracking of physical objects of interest and event recognition.

The proposed evaluation tool (ViSEvAl, visualisation and evaluation) respects three important properties:

- To be able to visualise the algorithm results.
- To be able to visualise the metrics and evaluation results.
- For users to easily add new metrics.

The ViSEvAl tool is composed of two parts: a GUI to visualise results of the video processing algorithms and metrics results, and an evaluation program to evaluate automatically algorithm outputs on large amount of data. An XML format is defined for the different input files (detected objects from one or several cameras, ground-truth and events). XSD files and associated classes are used to check, read and write automatically the different XML files. The design of the software is based on a system of interfaces-plugins. This architecture allows the user to develop specific treatments according to her/his application (e.g. metrics). There are 6 interfaces:

1. The video interface defines the way to load the images in the interface. For instance the user can develop her/his plugin based on her/his own video format. The tool is delivered with a plugin to load JPEG image, and ASF video.
2. The object filter selects which objects (e.g. objects far from the camera) are processed to compute the evaluation. The tool is delivered with 3 filters.
3. The distance interface defines how the detected objects match the ground-truth objects based on their bounding box. The tool is delivered with 3 plugins comparing 2D bounding boxes and 3 plugins comparing 3D bounding boxes.
4. The frame metric interface implements metrics (e.g. detection metric, classification metric, ...) which can be computed on each frame of the video. The tool is delivered with 5 frame metrics.
5. The temporal metric interface implements metrics (tracking metric,...) which are computed on the whole video sequence. The tool is delivered with 3 temporal metrics.
6. The event metric interface implements metrics to evaluate the recognized events. The tool is delivered with 4 metrics.

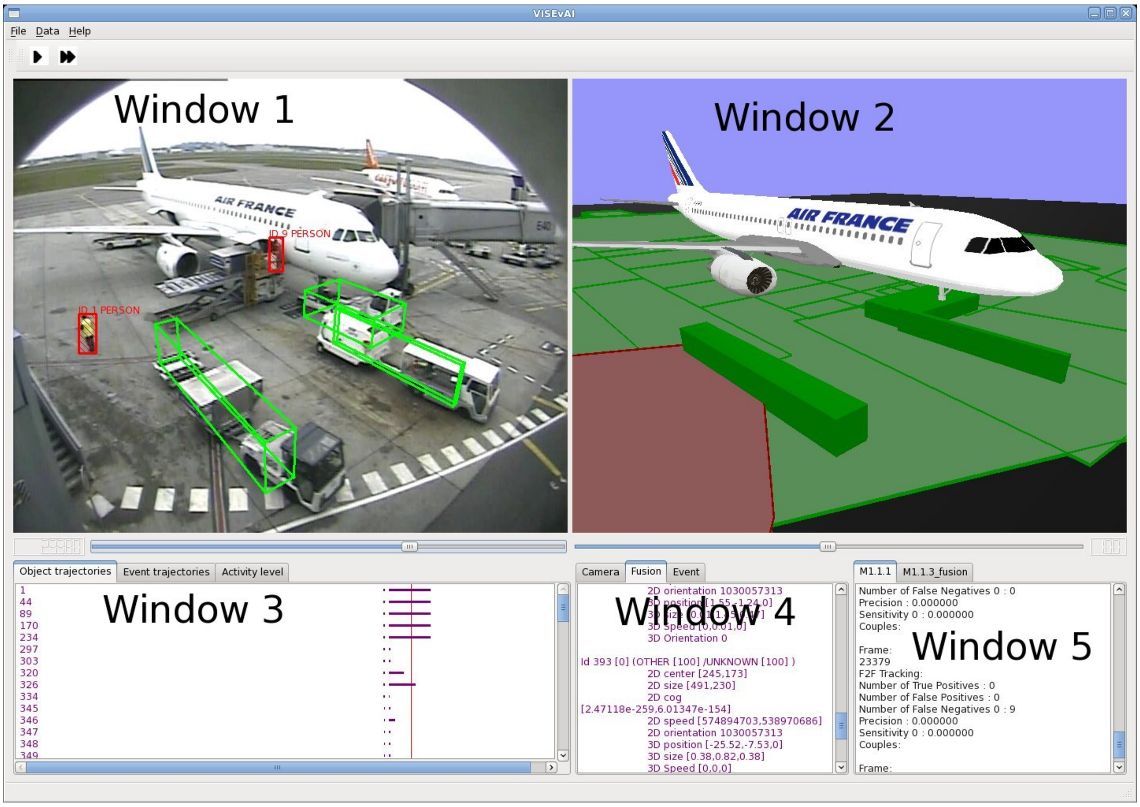


Figure 3. GUI of the ViSEvAl software

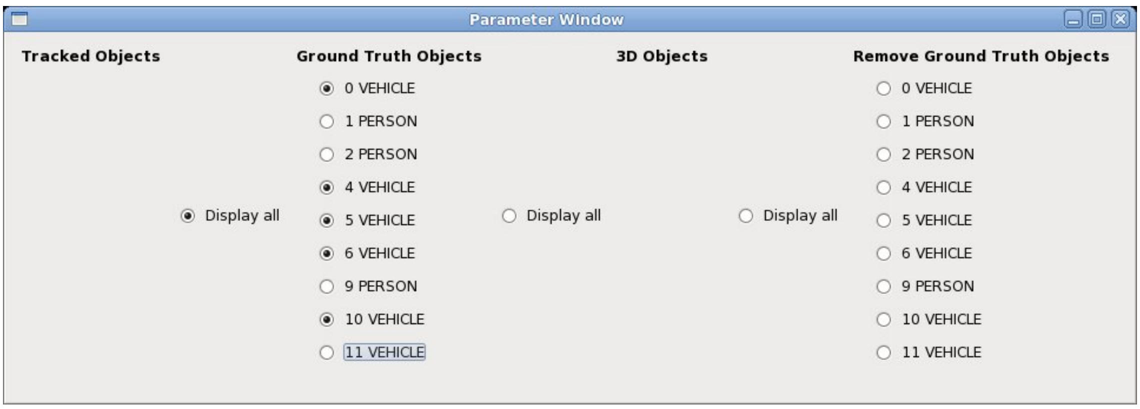


Figure 4. The object windows enables users to choose the object to display

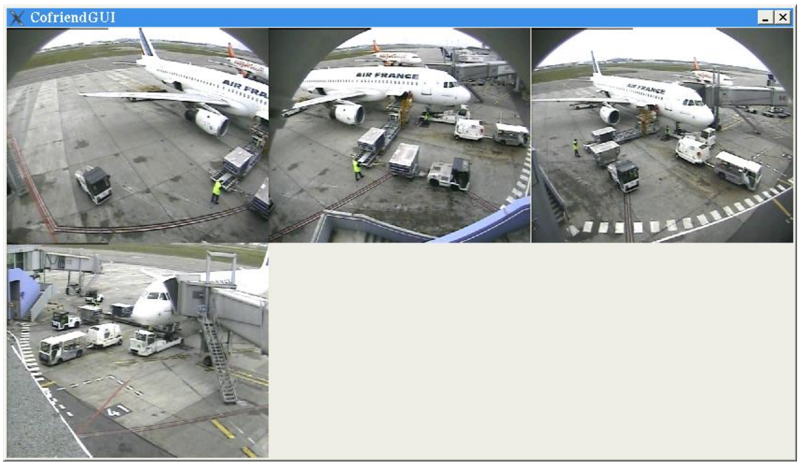


Figure 5. The multi-view window

The GUI is composed of different parts (see Figure 3):

- Window 1: the video part displays the current image and information about the detected and ground-truth objects (bounding-boxes, identifier, type,...).
- Window 2: the 3D virtual scene displays a 3D view of the scene (3D avatars for the detected and ground-truth objects, context, ...).
- Window 3: the temporal information about the detected and ground truth objects, and about the recognized and ground-truth events.
- Window 4: the description part gives detailed information about the objects and the events,
- Window 5: the metric part shows the evaluation results of the frame metrics.
- The object window enables the user to choose the object to be displayed (see Figure 4).
- The multi-view window displays the different points of view of the scene (see Figure 5).

The evaluation program saves, in a text file, the evaluation results of all the metrics for each frame (whenever it is appropriate), for all video sequences and for each object of the ground truth.

The ViSEvAl software was tested and validated into the framework of the Cofriend project through its partners (Akka,...). The tool is also used by IMRA, Nice hospital, Institute for Infocomm Research (Singapore),... The software version 1.0 was delivered to APP (French Program Protection Agency) on August 2010. The tool is available on the web page : http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl_Description.html

5.3. 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).

5.4. Clem

The *Clem Toolkit* [64](see Figure 6) is a set of tools devoted to design, simulate, verify and generate code for LE [17] [74] programs. This latter is a synchronous language supporting a modular compilation. The language also supports automata possibly designed with a dedicated graphical editor. The *Clem toolkit* comes with a simulation tool. Hardware description (Vhdl) and software code (C) are generated for LE programs. Moreover, we also generate files to feed the NuSMV model checker [61] in order to perform validation of program behaviors.

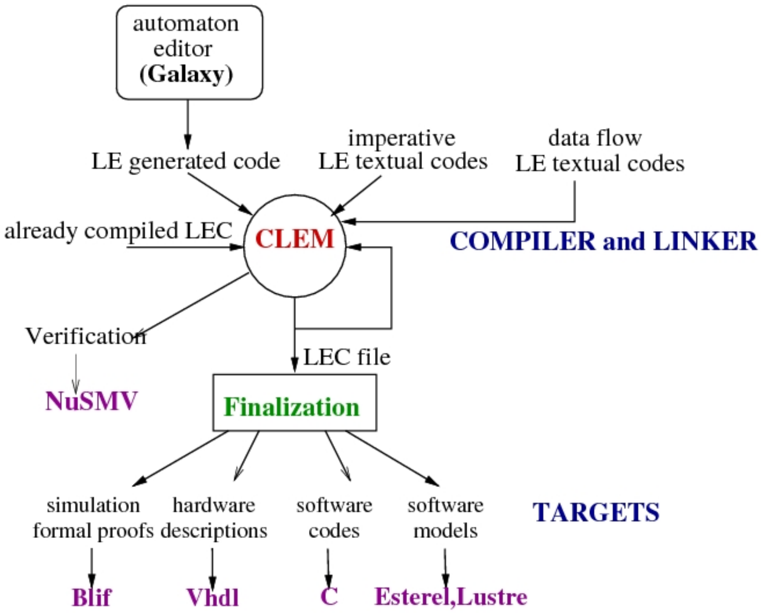


Figure 6. The Clem Toolkit

6. New Results

6.1. Introduction

This year Pulsar has tackled several issues related to its two main research axes : scene understanding for activity recognition and software engineering for activity recognition.

6.1.1. Scene Understanding for Activity Recognition

Participants: Slawomir Bak, Ikhlef Bechar, Tanvi Barnajee, Piotr Bilinski, Bernard Boulay, François Brémond, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Ezequiel Cura, Chedli Farhani, Saurabh Goyal, Julien Gueytat, Mohammed-Bécha Kaâniche, Ratnesh Kumar, Vincent Martin, Sabine Moisan, Emmanuel Mulin, Anh-Tuan Nghiem, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Leonardo Rocha, Rim Romdhame, Silviu Serban, Anja Schnaars, Malik Souded, Monique Thonnat, Sofia Zaidenberg, Nadia Zouba, Daniel Zullo.

This year Pulsar has proposed new algorithms in computer vision (people detection and people re-identification), in reasonning (activity recognition and uncertainty handling) and in learning (activity modeling based on offline trajectory clustering). More precisely, the new results for this research axis concern:

1. The early detection of pest attacks in a greenhouse by using computer vision (6.2);
2. The detection of mobile objects based on a fine distribution of pixel values for both foreground and background images (6.3);
3. People tracking combining SIFT and Particle Filter (6.4);
4. A control framework for tracking algorithms handling context variations (6.5);
5. The human detection and re-identification in a multi-camera network (6.6);
6. The tracking of groups of people (6.7);
7. Action recognition (6.8);
8. An extendable event recognition algorithm: SED (6.9);
9. Real-time Monitoring of TORE Plasma (6.10);
10. Monitoring activities for Alzheimer Patients (6.11);
11. Trajectory Clustering for Activity Learning (6.12);
12. Using Texture Descriptors and Edge Detectors in Image/Video Segmentation with New Optimization Tools (6.13);
13. Learning How to Learn and Understand : Automatic Model Construction (6.14).

6.1.2. Software Engineering for Activity Recognition

Participants: François Brémond, Bernard Boulay, Oscar Carrillo, Erwan Demairy, Chedli Farhani, Daniel Gaffé, Julien Gueytat, Santiago Hurtado, Ratnesh Kumar, Sabine Moisan, Zouhair Ramadhane, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Sagar Sen, Monique Thonnat, Jean-Yves Tigli, Christophe Tornieri, Sofia Zaidenberg, Daniel Zullo.

This year Pulsar has improved the SUP platform. This latter 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 studied the development of a scenario recognition module relying on formal methods to support activity recognition in SUP platform. We began to study the definition of multiple services for device adaptive platform for scenario recognition. We continue to develop the Clem toolkit around a synchronous language dedicated to activity recognition applications. The new results related to this research axis concern:

1. The SUP Software Platform (6.15);
2. Model-Driven Engineering and Video-surveillance (6.16);
3. Scenario Analysis Module (6.17)
4. Multiple Services for Device Adaptive Platform for Scenario Recognition (6.18);
5. SUP and WComp Platforms Interoperability (6.19);
6. The Clem Workflow (6.20).

6.2. Early detection of pest attacks in a greenhouse by using computer vision

Participants: Ikhlef Bechar, Ratnesh Kumar, Vincent Martin, Sabine Moisan, Leonardo Rocha.

This project concerns non-invasive integrated pest management methods for greenhouse crops by using computer vision. A network of sensors (namely video-cameras) is placed in the greenhouse in order to detect some harmful insect species of interest (e.g., whiteflies, greenflies). The acquired videos are processed in real time on a PC to classify harmful pests of interest in the scene and to estimate their population over time. The challenges pertain to the big size of the video frames, the very low spatial resolution and color contrast of the pests, the online constraint and the environmental issues (e.g. lighting changes, presence of outliers).

We had previously developed a first version of a vision system consisting of several modules: video acquisition, trap extraction, background subtraction, insect presence detection, image patch voting, insect detection, insect classification, and insect tracking which cooperate to achieve a satisfactory final vision system. This year we improved this system as follows :

- A new segmentation algorithm to take into account the shadowing problem in order to increase detection sensitiveness.
- A module to detect quick lighting changes (due for instance to the presence of human operator near the the camera); to enable the system to stop the processing during this period and to resume afterwards.
- A new tracking algorithm. The previous tracking algorithm was based on maximum correlation tracking. It has shown some non-robustness due the low SNR of the objects. Thus, we replaced the maximum correlation tracking by a new tracking criterion which updates the position of a previously detected insect in the current frame by using a weighted average of all possible positions of the insect in the current frame.

We acquired *in situ* during spring and summer 2010 many videos under various environmental conditions including shadowing due to leaves motion, slow/quick lighting changes, presence of outliers in the FOV of the cameras and so forth. We tested our new version against them and we noticed significant improvements. We could validate the new system by showing its acceptable performance in terms of accuracy and computational time.

We also developed an algorithm to classify whitefly and greenfly, as a three class separation problem (we also have samples of background from segmentation). The inputs to this algorithm are the blobs obtained from segmentation algorithm. The shape of both these flies being very similar combined with poor spatial resolution due to low-quality optical system and JPEG compression makes it challenging for a classifier to be robust on features based on area, elongation and color. Hence we adopted a two stage procedure. In the first stage we separate whiteflies from greenflies and background samples. Then the samples which are labeled negative (w.r.t. whitefly) are fed to another classifier and we make a decision whether we have a greenfly or background. For classification machine we used support vector machines. We evaluated three feature extraction methods (Color data, Pyramidal Histogram of Gradients and Gabor Filter Bank) [39]. Based on the results, we implemented Gabor filter bank based feature extraction.

This classification algorithm was incorporated into a complete processing chain and tested during summer 2010 in a tomato greenhouse in Avignon, equipped with a network of six Wifi cameras. There where 6 processing chains instances running simultaneously on the 6 cameras. The video cameras observed sticky traps and detected trapped flying insects. We conducted different experiences to measure the performance of the software w.r.t. manual counting. The system can run and count insects in real time in a multiple camera system, even from a remote location through an ssh conection. We could compare the results from a human operator to the results of the automatic counting. The system has more than 57% accuracy, and can be improved, e.g., by adding a background subtraction algorithm and on line learning and adaptation.

6.3. Detection of mobile objects based on a fine distribution of pixel values for both foreground and background images

Participants: An Tuan Nghiem, François Brémond.

Detecting mobile objects is an important task in many video analysis applications such as video surveillance, people monitoring, video indexing for multimedia. Among various object detection methods, the ones based on adaptive background subtraction are one of the most popular. Normally, this approach requires a background subtraction algorithm and a shadow removal algorithm. The background subtraction algorithm is responsible for modeling the background to detect foreground regions (the regions which are different from the background) in the current frame. Then the shadow removal algorithm will remove foreground regions corresponding to shadow and highlight so that the remaining foreground regions only correspond to objects of interest. To realize this objective, we have proposed a new background subtraction algorithm and a new shadow removal algorithm.

The proposed background subtraction algorithm is more robust to difficult situations compared to other popular background subtraction algorithms such as Gaussian Mixture model, Kernel Density Estimation, or Codebook model. Most of these popular algorithms are based on the hypothesis that pixel values associated to the background occur more frequently than pixel values belonging to objects of interest. Therefore if a pixel value occurs more than a given threshold, it is classified as background. However, there are cases when this hypothesis is incorrect. For example, in outdoor scenes, tree leaves move due to the wind and pixel values associated to the leaves may occur only few times. Therefore, these pixel values will be classified as foreground by background subtraction algorithms. On the other hand, a pixel value may occur frequently because people in the scene wear similar clothes. Consequently, this pixel value will be classified as background. To overcome these problems, the proposed background subtraction algorithm collects information not only about the number of occurrences of pixel values but also about the length of continuous occurrence (in how many continuous frames in average, a pixel value occurs) and the occurrence density (pixel values occur continuously without interruption or not). With the information about the length of continuous occurrence, the proposed algorithm can distinguish pixel values of tree leaves when they move with pixel values of objects of interest. With the information about the occurrence density, the proposed algorithm can distinguish pixel values of newly added background such as the displacement of a chair with the pixel values of people passing by.

The proposed shadow removal algorithm also uses the hypothesis that shadow / highlight does not change the chromaticity of the background and it has homogeneous effect on adjacent pixels. However, unlike many other shadow removal algorithms, to correctly verify the invariance of chromaticity and the homogeneity, the proposed shadow removal algorithm relies on a camera model and an illumination model. Based on these models, we can show the advantages / disadvantages related to the chromaticity representation and homogeneity verification. With these models, we also propose a new chromaticity verification and a new homogeneity verification. The proposed verifications are less affected by the characteristics of the camera and the illumination change than other popular chromaticity verification (e.g. UV in YUV color space) and homogeneity verification (e.g. LPB).

The proposed background subtraction algorithm and the shadow removal algorithm have been evaluated using the videos of the ETISEO evaluation project. Comparing with the results of the participants in the ETISEO project, the results of the proposed algorithms are always at top 3 among 20 participants which are the leading teams in the world for object detection. In another experiment, these algorithms were also used to detect mobile objects in 1h video from the project GERHOME. Although the video is difficult with many displacements of contextual objects and many illumination changes, the detection results are still quite good (precision = 0.95, sensitivity = 0.86 in terms of objects) compared to GMM (precision = 0.87, sensitivity = 0.81). The proposed algorithms have also been validated through live demonstrations in Paris subway station and Bobigny Official Building in the framework of the SIC project.

The two proposed algorithms have been described in the PhD thesis of Anh-Tuan Nghiem entitled: “Adaptive algorithms for background estimation to detect moving objects in videos”([25]).

6.4. People tracking combining SIFT and Particle Filter

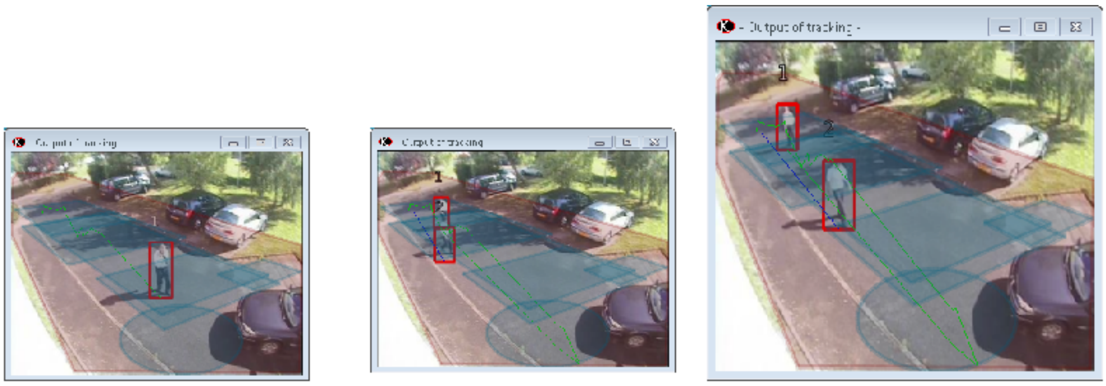
Participants: Malik Souded, François Brémond.

Our aim is to implement a mechanism for tracking people through a camera network, in order to allow us, at every moment, to locate a person moving in an area covered by the camera network, and to obtain its entire trajectory since its first appearance.

We chose to base our approach on the re-identification paradigm, whenever a person appears on the field of view of one camera. This re-identification can be done only with a strong visual signature learned and assigned to each individual moving in the video camera network. Learning this visual signature requires a good tracking of each individual as it is visible in the field of view of one camera. Thus, a robust single-camera tracking of individuals is very important, and it represents the first part of our work.

We propose an approach for real-time tracking of people and objects from videos. Initially, identification of objects of interest to track is performed by detecting moving objects. This detection is done using a background subtraction algorithm. Once the moving objects are detected, a first classification based on the 3D dimensions of objects is done. Then, for each newly detected object of interest, we assign a first signature composed of SIFT (Scale-Invariant Feature Transform) points of interest, around which we compute the descriptors of the same name. SIFT descriptor is widely reported in the literature and is known as one of the most robust descriptors, which led to the creation of a large number of variants (PCA-SIFT, GLOH, DAISY, etc.). We associate to each object of interest a set of SIFT descriptors that we track, which allows us to track the object itself. Tracking SIFT points is performed using a Bayesian model estimation method called "Particle Filter" (or sequential Monte-Carlo method). After initialization of particle filter for each SIFT point at its detection, the point is tracked with a succession of predictions and updates.

We have tested the mono-camera tracking approach on a set of video sequences with different difficulty levels, including illumination changes, variations of moving speed, and cases of crossing and partial occlusions. We have obtained encouraging results (see Figure 7).



(a) Example of tracking (b) a single person is tracked (c) a second person appears in the scene and is also tracked

Figure 7. The two persons are crossing each other (b and c). The tracking is well performed during and after crossing.

6.5. A control framework for tracking algorithms handling context variations

Participants: Duc Phu Chau, François Brémond, Monique Thonnat.

Although there are many studies aiming at solving tracking problems related to context variations, tracking algorithms are still not able to perform well in many different scene conditions. One tracker can obtain good results in one video but may fail in other ones. In order to overcome this limitation, we propose a control framework for tracking algorithms which is able to adapt itself to different contexts. The main principle of the framework is to choose a tracker among a set of available tracking algorithms and tune its parameters to get the best possible performance. To do that, we have proposed two new tracking algorithms with different approaches.

The first tracking algorithm is based on the combination of a Kalman filter with a trajectory filtering process. The mobile object trajectories are usually fragmented because of occlusions and mis-detections. This trajectory filtering process, named global tracker, aims at removing the noisy trajectories and fusing the fragmented trajectories belonging to the same mobile object. The method has been tested with five videos with different difficulty levels (Gerhome project, TRECVid dataset, ETISEO dataset). The results show that the proposed tracker gets good performance in all cases and outperforms several trackers from the state of the art (see Figure 8). This work will be published in [60].

The second tracking algorithm is based on a feature pool which is adaptable to different contexts. An offline learning process is proposed to search for useful features and to tune their weights depending on each considered context. During the online tracking process, the system uses a temporal window for establishing the matching links between detected objects. This enables the system to compute correctly the trajectory even if the mobile objects are mis-detected in some frames. We have tested the proposed tracking algorithm with four different contexts (Caretaker, TRECVid, ETISEO building, ETISEO subway). The experimentation has shown that the proposed tracking algorithm has also obtained good performance in different contexts.

More work still needs to be done to understand the limitations of the proposed trackers. We are currently defining a process to learn the relationship between tracking algorithms and video contexts.

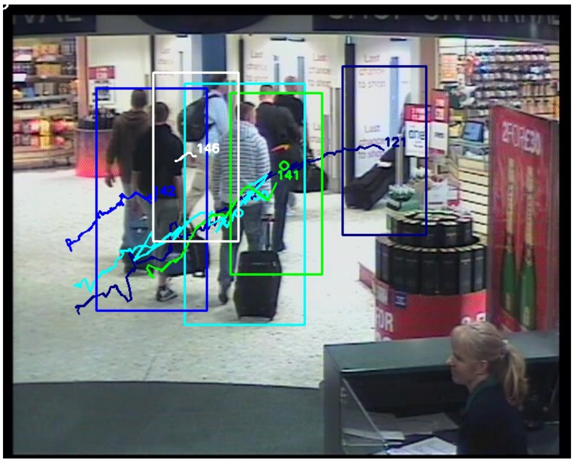


Figure 8. Illustration of the first tracker results for TRECVid video

6.6. Human detection and re-identification in a multi-camera network

Participants: Slawomir Bak, Etienne Corvée, Silviu-Tudor Serban, François Brémond, Saurabh Goyal.

In many systems there is a need to identify individuals or determine whether a given individual has already appeared somewhere in a network of cameras. Human identification is an important and still unsolved problem in computer vision. Before people can be re-identified, they need to be assigned with a reliable signature which can be provided by multiple instances of the people detected in different cameras. However, from the detection of people in independent cameras to their identity recognition in another one, many issues need to be handled such as the large complexity in human appearance. For instance, clothing can strongly alter human silhouette, and people activities make human pose varies.

The first stage of the re-identification approach consists in detecting and tracking people in one camera. We use Histograms of Oriented Gradient i.e. HOG features as human appearance feature. Human appearances are then learned using a hierarchical tree learning algorithm. The people detection is integrated in a vision platform that takes into account additional cues such as 3D and motion constraints to enhance the detection precision and limit the number of false detections.

We are also working on a face detector using Local Binary Pattern i.e. LBP to detect faces in challenging scenarios where faces are seen from far distances, rotated, with low image quality and video sequences where people are moving. We are also integrating several features (e.g. HOG, LBP and color) to be combined to enforce the people detection robustness. Detected faces can then be combined with people detection when detected in a frontal view for face re-identification purposes. The detected people are then tracked from correspondence from a frame to frame basis.

The second stage consists in retrieving the tracked people in different cameras. As a human signature we proposed a new appearance model based on spatial covariance regions extracted from human body parts. The human body parts are automatically detected using Histograms of Oriented Gradients (HOG) to establish spatial correspondences between individuals. An overview of the overall human identification framework is presented in Figure 9.

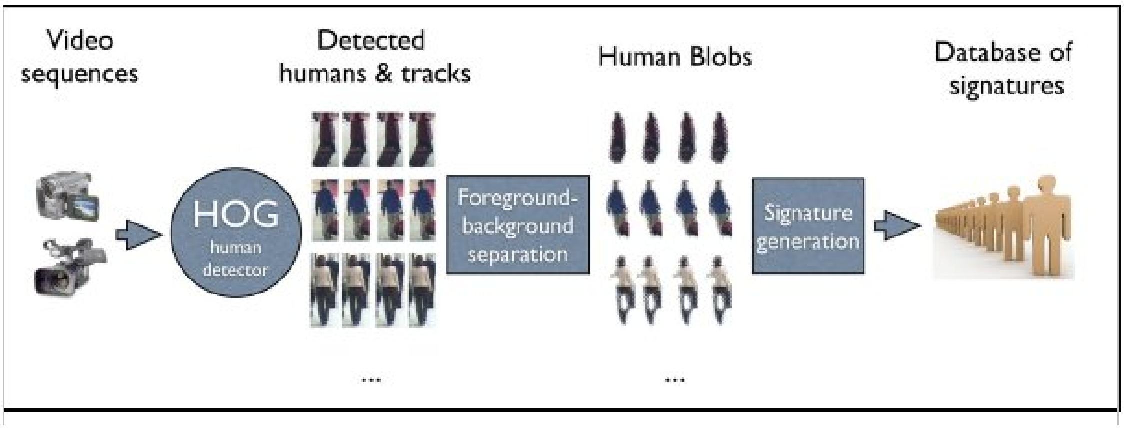


Figure 9. Overview of the proposed human re-identification framework.

The invariant signatures for comparing different regions are generated by combining color and structural information. The color information is captured by covariance descriptors based on color channels and their gradients. Invariance to differences in ambient illumination is achieved by histogram equalization. The structural information is represented by covariance matrices computed over body part areas as shown in Figure 10. Moreover, the spatial alignment is handled by the use of spatial pyramid matching in the two-dimensional image space of the region of interest. The results show that the proposed approach outperforms state of the art methods on the TrecVid dataset.

6.7. Tracking of groups of people

Participants: Sofia Zaidenberg, François Brémond.

In the framework of the European project VANAHEIM 7.3.3, we address the problem of tracking groups of people. Based on frame to frame mobile object detection, we try to detect which mobiles form a group and to follow the group through its lifetime. We define a group of people as two or more persons being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less chaotic: people may join or split from the group, one or more can disappear temporarily (occlusion or disappearance from the field of view) but reappear and still be part of the group. The motion detector which detects and labels mobile objects may also fail (misdetections or wrong labels). Analyzing trajectories over a temporal window allows handling this instability more robustly. The long term goal of group tracking is the recognition of abnormal behaviors such as violence or vandalism. The group tracking algorithm is divided

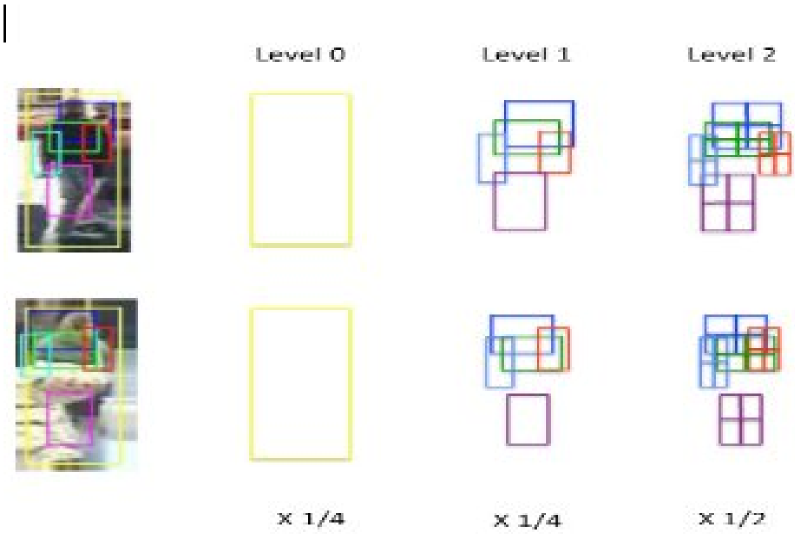


Figure 10. Example of a three-level pyramid. Level 0 corresponds to the full body part. Level 1 and level 2 correspond to the rest of detected body parts and grids inside body parts, respectively.

into four main steps: (1) update of groups, (2) creation of groups, (3) split/merge of groups and (4) termination of groups. Each group has a coherence coefficient. This coefficient is a weighted sum of three quantities characterizing a group: the group density (average of distances between mobiles), the similarity of mobile's speed and the similarity of their motion directions. Other criteria may be used, such as the distance of each mobile to the center of gravity of the group or the maximum and minimum speeds of group members. The update of a group consists in re-calculating the group coherence with a mobile from the current frame. If adding the mobile does not put the coherence over a defined threshold, the mobile is added to the group. A pre-selection can be made by only considering mobiles that are close enough to the center of gravity of the group. After the update step, all mobiles that have not been assigned to a group are analyzed to form new groups. Depending on its size, a mobile can be labeled as a GROUP_OF_PEOPLE. In this case, a group is created with this mobile. Otherwise, pairs of mobiles are tested and if their coherence is over the coherence threshold, they are assigned to form a group. The temporal window of analysis allows to eliminate falsely created groups. For example, two people can be close for a few frames while getting off a metro train because of the narrow doors, but then take independent paths. Groups are created with a delay T. If the current frame is designated by t_c, a group is created at frame $t_c - T$ if the corresponding moving regions form a group that would be observed from frame $t_c - T$ to t_c. During this window, a group can be lost and found again. If no mobiles are added to the group for one or several frames, the group still exists and mobiles can be added to the group later in time (for example if the group left temporarily the field of view or was occluded). If no mobiles are added to the group for T frames, the group is terminated.

6.8. Action Recognition

Participants: Piotr Bilinski, François Brémond.

We propose a new approach for human action recognition in videos. In the first step of our algorithm, in order to compute moving objects and by the same to reduce the computational space size, we apply background subtraction method, we group foreground, neighboring pixels and we classify them as moving objects or noise. In the next step, for each detected moving object, we compute corner points using the Features from

Accelerated Segment Test (FAST) corner detector algorithm and for each such feature point, we compute a descriptor. The proposed descriptor is based on the Histogram of Oriented Gradients approach, very successful and powerful descriptor in object detection domain of computer vision. We use these descriptors to track in consecutive frames detected corner points, choosing neighboring points with the most similar descriptors. We process each video sequence from our database in the same way. We obtain a set of point trajectories for each video sequence. Having such video description, we evaluate our approach using machine learning techniques. Using k-means algorithm, we cluster all the point trajectories from the action learning videos and we classify all the trajectories into action classes from the testing videos using k-nearest neighbor algorithm. In order to estimate how accurately our approach performs on an independent dataset, we use cross validation technique. We evaluate our approach on two public video datasets (the KTH and Weizmann human action datasets, see Figure 11).

This project contains also a second aim, to distinguish similar activities such as sitting down or standing up by healthy and Alzheimer patients. We collect a real, relatively high-resolution human action dataset. For this purpose, we collaborate with CHU at Nice hospital, where we record video sequences. Our action dataset contains several types of human actions performed by different people: people standing up and sitting down, people walking, playing cards, reading a magazine, watching TV, etc. Our goal is to collect a new dataset with real human actions to allow us to evaluate action recognition approaches in this real environment.

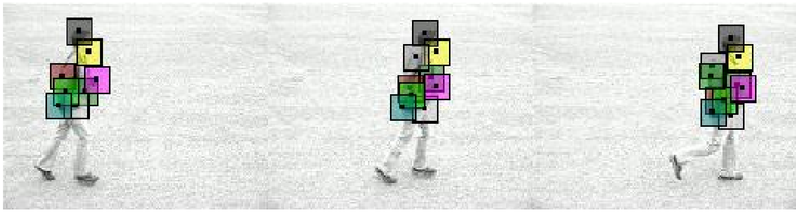


Figure 11. Examples from the KTH dataset of person jogging. Colored squares represent the calculated descriptors, whose centers are the tracked corners points.

6.9. Extendable Event Recognition Algorithm: SED

Participants: Bernard Boulay, François Brémond.

This year we have improved the event recognition task of the SUP platform. This task recognizes a set of events which have been predefined by end-users by taking as input tracked objects which have been detected by vision tasks. The new algorithm named SED (Simple Event Detector) verifies logical and temporal constraints based on the attributes (e.g. posture, position,...) of the objects of interest (e.g. person,...). This work is based on [24]. The algorithm is composed of a generic recognition engine and a declarative descriptions of the events (including an ontology of the domain) to be recognized for a specific application. The ontology of the domain (including the definition of primitive events and more complex events directly related to end-users) is taken as a parameter of the event recognition. That is to say the event recognition engine does not change even if it is applied to another domain. The ontology is first defined by the objects of interest of the application (e.g. person, zone, ...). We propose a simple language to describe these objects based on predefined basic types (CSDouble, CSString, CSTimeStamp,...):

```
class Mobile
 {
   const false;
   CS3DDPoint Position;
   CS3DDPoint Size3D;
```

```
class Person:Mobile
  {
    const false;
    CSInt Posture;
  }
```

```
  CSDouble Speed;
  CSDouble Orientation;
  CSString SubType;
 }
```

Then end-users can add new attributes or new types of object. The engine is able to handle heritage between objects (e.g. a Person is a Mobile). The user can model their scenarios using the object attributes (operator $\rightarrow$) and using the historical values (list of previous values) of the attributes (operator $\rightarrow *$). To help the user writing in an intuitive way the constraints of the scenario model, we propose a language to describe new operators by defining specific rules. For instance the user can write "Person in Zone" instead of "Person $\rightarrow$ Position in Zone $\rightarrow$Vertices" by defining the rule:

```
operator in
{
  arguments((o1 : Person) (o2 : Zone))
 attributes((o1,Position)(o2, Vertices))
}
```

13 operators are modeled and can be classified in 3 categories:

- logical operators (e.g. accession($\rightarrow$), =, $<$, not(!), duration,...)
- spatial operators (e.g. in, distance)
- temporal operators (Allen algebra (meet, before,...))

We update the description format of the scenario models proposed in [24]:

- to control some specific low level processing (called ControlledProcess). For instance, according to a recognized event, a PTZ camera can be controlled to zoom on a specific zone. It can also be used to detect objects which are split into pieces by the image segmentation process.

```
ControlledProcess (
  (CameraControl,
  (Object z1)(Camera ptz3)(Priority HIGH) (ControlType CENTER_ZOOM)
  (Duration 1200)(SVT AFT_LOADING))(MergeObject, (Object v1)(Threshold 10))
)
```

- to manage confidence (and uncertainty) of the recognized events by associating confidence weights for the event components (expressing how significant are the event components compared to each other). The user can also associate a threshold for a specific scenario model to decide whether the scenario is considered as recognized or not.

```
CompositeEvent(FWD_CN_LoadingUnloading_Operation_Starts[0.7],
    PhysicalObjects( (v1 : Vehicle), (v2 : Vehicle), (z1 : Zone), (z2 : Zone) )
    Components( (c1 : CompositeEvent FWD_LD_Positioning(v1, z1)[0.6] )
                (c2 : CompositeEvent FWD_TS_Positioning(v2, z2) [0.4]) )
    Constraints( (c1 before c2) )
    Alarm ((Level : VERYURGENT))
)
```

A scenario model is composed of :

- a name,
- a list of physical objects of interest,
- at most two event components,
- a list of constraints (at most one temporal constraint and several logical constraints),
- a list of specific treatments (e.g. low level processing).

To obtain a real time recognition (video cadence), optimisation is made in two directions:

- the scenario tree which describes all the previously recognized scenario models is constructed by taking into account the last event component which has triggered the recognition of the main event,
- the partially instantiated scenarios are saved to avoid recomputing the similar situation at the next frames.

The engine is operational and it has been tested and used in the framework of the European Cofriend project to monitor activities of people and vehicles around a parked aircraft in Toulouse airport (Loading/Unloading luggage,...).

We focus now on designing probabilistic operators to give a confidence value to the recognized scenarios. For instance the operator "in" should give a probabilistic value (e.g. the person is inside the zone with a probability of 0.9) depending how far the object is detected. The next task is to generalise the confidence for events defined with any operator.

6.10. Real-time Monitoring of TORE Plasma

Participants: Vincent Martin, Guillaume Charpiat, François Brémond.

This topic is studied in the framework of the moniTORE project. It is a collaborative research action between CEA and INRIA Pulsar team which aims at monitoring fusion plasma in tokamaks by means of imaging (infrared and visible) diagnostics. Imaging diagnostics will be strongly involved in the machine protection function of future devices like ITER. Up to 36 video cameras monitoring 70% of the inner vessel components are planned for ITER. In this context, new approaches and techniques for real-time monitoring have to be developed for the automatic detection and recognition of abnormal overheating or abnormal plasma events.

We have developed a prototype system (a demonstrator) set up on the French tokamak Tore Supra. This unique intelligent vision system is able to recognize different types of thermal events [27] observed during plasma operation. The goal of this system is twofold. First, it can be seen as a computer-aided system for plasma-wall interaction understanding (plasma physics issues) providing to plasma physicists a new source of information through thermal event databases. Second, it tends to prove that real-time recognition of thermal events can be useful for safety operation, for instance to ensure the protection of plasma facing components. An illustration of the detection and the recognition of several thermal events is given in Figure 12. The demonstrator operates in parallel of the existing real-time image analysis system for one camera view. Costly algorithms (pixel-wise processing) have been parallelized on a FPGA [41] to reach real-time constraints (max. processing time = 20 ms per frame). High level reasoning processes are achieved using multithreaded CPU resources. To this end, we have developed a new modular software C++ platform named PINUP (Plasma ImagiNg data Understanding Platform). We are currently assessing the performance of the system against the previous system.

6.11. Monitoring activities for Alzheimer Patients

Participants: Rim Romdhame, Emmanuel Mulin, François Brémond, Philippe Robert, Monique Thonnat.

The aim of this work is to propose a constraint-based approach for video event recognition with probabilistic reasoning for handling uncertainty. This work has two aspects: video understanding and health care monitoring.

Concerning the video understanding aspect we have extended the event recognition algorithm proposed in [24] to take into account the uncertainty of the recognition. We computed the precision of the 3D information characterising the mobile objects moving in the scene. 3D information is used to compute the spatial probability of the event. We take as input the mobile object (i.e. person) coordinates at each time instant t and the 3D coordinates of contextual objects (i.e. 3D projection coordinates of zones and equipment on the ground). The algorithm computes the distance of the person to the different borders of the contextual zone. This distance is zero when the person is inside the zone and the distance is high if the person is far from the zone. Based on this distance, we compute the Gaussian probability that a primitive state occurs within the

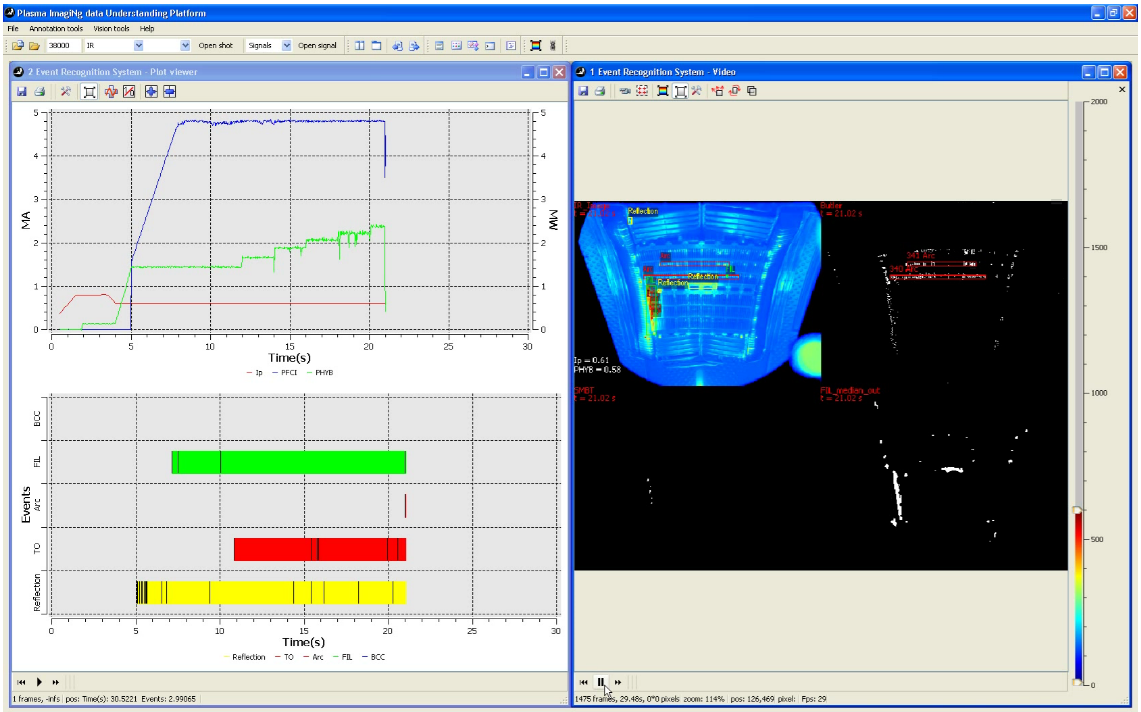


Figure 12. Example of detection and recognition of thermal events.

zone. The event is recognized when the computed probability is over a predefined threshold. We have also computed the mobile object reliability. The reliability computation method consists mainly in estimating how much the mobile object features vector deviate from an estimated distribution model. For that, a vector of five features (motion step, velocity, direction, 3D height and 3D width) that best characterize the mobile object is extracted. For each feature value we compute respectively the instantaneous reliability at time instant t as a Gaussian probability density function. This work has been published in [47].

Concerning the health care monitoring aspect, we have worked in close collaboration with clinicians from Nice hospital to evaluate the behaviours of Alzheimer patients. We have first done a strong effort to model the behaviours of Alzheimer patient using our event modeling formalism. The result is 69 models which compose our knowledge base of events for health care applications. With the help of clinicians we have established a scenario protocol. The scenario is composed of three parts: (1) directed activities (10 min), (2) semi directed activities (20 min), (3) free activities (30 min). A first objective consists in defining the strategy to adopt to detect early symptoms of Alzheimer disease with the use of the proposed video understanding approach. This work was presented in [46]. Experiments have been performed in a room of Nice hospital equipped with 2 camera videos where 18 elderly volunteers have spend between 15 min to 1 hour (see Figure 13). Volunteers include Alzheimer, MCI (mild cognitive impairment) and healthy elderly.

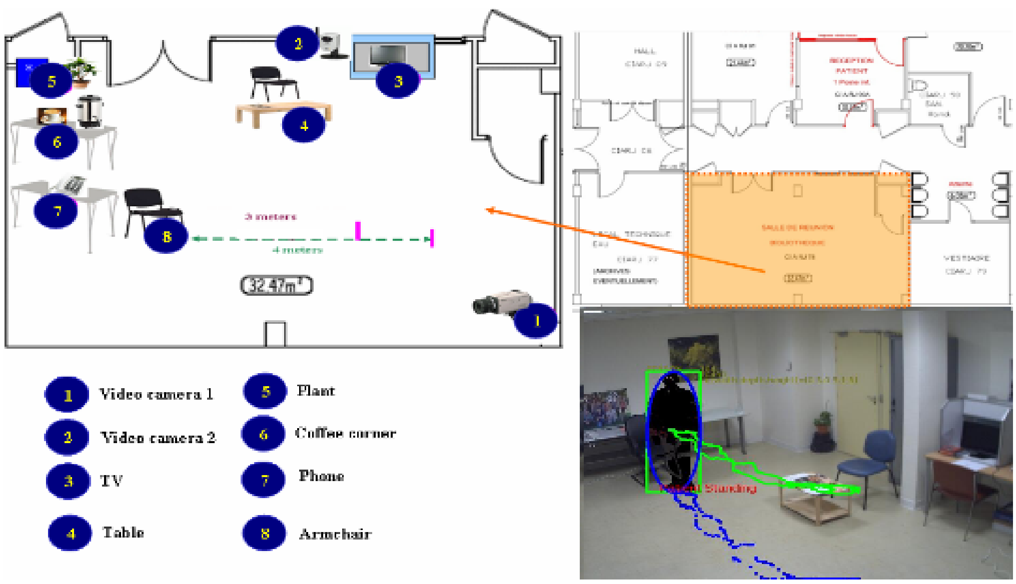


Figure 13. Experimental site at Nice hospital.

6.12. Trajectory Clustering for Activity Learning

Participants: Jose Luis Patino Vilchis, Guido Pusiol, François Brémond, Monique Thonnat.

The discovery, in an unsupervised manner, of significant activities observed from a video sequence, and its activity model learning, are of central importance in our team to build up on a reliable activity recognition system. We have deepened our studies on activity extraction employing trajectory information. In previous work we have shown that rich descriptors can be derived from trajectories; they help us to analyze the scene occupancy and its topology and also to identify activities [70], [71], [73], [59]. Our new results show how trajectory information can be more precisely employed, alone or in combination with other features for the extraction of activity patterns. Two application domains are currently being explored: 1) Monitoring of elderly people at home; 2) Monitoring the ground activities at an airport dock-station (COFRIEND project).

6.12.1. Monitoring of elderly people at home

To understand daily activities at home-care applications, we propose a novel framework. The framework is capable of discovering and modeling activities (e.g. "Cooking", "Eating") in unstructured scenes (i.e. "an apartment"). The modeled activities can be recognized in unseen video datasets.

The framework links basic visual information (i.e., tracked objects) to the discovery and recognition of activities by constructing an intermediate layer of Primitive Events in a completely unsupervised way.

The primitive events capture the spatial transitions of the person between interesting regions of the scene. The interesting regions are learned describing a topology that is general enough to be shared by all individuals.

The composition of primitive events is very informative about the description of most of activities. Thus we search for particular patterns within the primitive event layer to discover interesting activities. The patterns of primitive events are then used as generic activity models in order to recognize automatically the main activities for a new observed person.

The results show 90% of correct recognized activities such as "Eating", "Cooking", "Sitting on the armchair". Nevertheless, the method fails at discriminating different activities at the same spatial position (i.e. “Sitting on the armchair/Standing near TV”). To solve this problem, recently we have extended the framework introducing the concept of actions and adding optical flow information that can describe the motion of an activity (see Figures 14 and 15).

Figure 14. Representative actions of activities recognized: (a)”Standing near TV”, (b) “Sitting on the armchair”.

The new results achieve an accuracy of 99% for the recognition of activities such as “Sitting on the armchair”, “Standing near TV”, “Preparing Meal”, “(Going) Kitchen to Bathroom”, “Eating”, etc. The approach can be used to recognize most of the interesting activities in a home-care application and has been published in [43].

6.12.2. Monitoring the ground activities at an airport dock-station

The COFRIEND project aims at creating a system for the recognition and interpretation of human activities and behaviours. Although activity models can be built by experts of the domain, this might be a hard and time consuming task depending on the application and the spectrum of activities that may be observed. The challenge consists thus in discovering, in a unsupervised manner, the significant activities observed from a video sequence. Our contribution is a novel approach for activity extraction and knowledge discovery from videos. Spatial and temporal properties from detected mobile objects are modeled employing soft computing

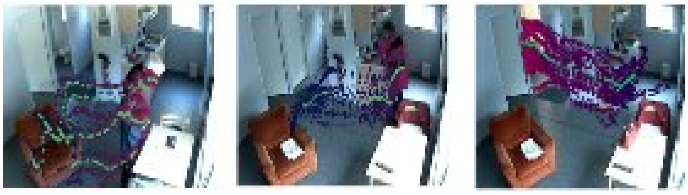


Figure 15. Representative actions of activities recognized: (a)"Armchair to Table", (b) "Armchair to Kitchen", (c) "Kitchen to Bathroom".

relations, that is, spatio-temporal relations graded with different strengths. Our system works off-line and achieves the extraction of activity patterns from the video. This system is composed of three modules: the trajectory speed analysis module, the trajectory clustering module, and the activity analysis module. The first is aimed at segmenting the trajectory into segments of fairly similar speed. The second is aimed at obtaining behavioural displacement patterns indicating the origin and destination of mobile objects observed in the scene. We achieve this by clustering the mobile trajectories and also by discovering the topology of the scene. The latter module is aimed at extracting more complex patterns of activity, which include spatial information (coming from the trajectory analysis) and temporal information related to the interactions of mobiles observed in the scene, either between themselves or with contextual elements of the scene. Spatial and temporal properties from detected mobile objects are modeled employing soft computing relations. These can then be aggregated employing typical soft-computing algebra. A clustering algorithm based on the transitive closure calculation of the final relation allows finding spatio-temporal patterns of activity. An example of discovery is given in the Figure 16. This approach has been applied to dock-station monitoring at the Toulouse airport [43], [44], [53]

6.13. Using Texture Descriptors and Edge Detectors in Image/Video Segmentation with New Optimization Tools

Participants: Anja Schnaars, Guillaume Charpiat.

Scene understanding requires first the extraction of information from the images or the video. As in classical computer vision, the two main approaches to image/video analysis are patch-based recognition of actions (as in object detection, with SIFT/HOG/3D features) and object segmentation (with priors on objects or with background subtraction, for instance). We will focus here on the second approach, i.e. segmentation. We aimed first at analysing the expressive power of texture for segmentation. We considered several classes of texture (sky, grass, clothes, etc.) and rewrote the task of segmentation as a texture classification task, with spatial coherency. We studied varied texture descriptors (i.e. many local descriptors such as SURF, co-occurrence matrices, Gabor filters, gradient histograms...) as well as varied edge detectors. Based on training sets of textures and of manually segmented images, the texture descriptors were used to learn the probability that each pixel of an image belongs to such or such texture class, while the edge detectors were used to learn the probability that neighboring pixels have different texture classes, as a spatial coherency criterion (see Figure 17). The problem was then formulated as a Markov random field and solved classically with graph cuts, similarly to [49]. Graph cut is a very powerful optimization tool that is guaranteed to give optimal solutions or good local minima (depending on the class of problems) in low computational time.

However, graph cuts in the literature are known to solve only a sub-class of Markov random field problems, for instance a condition on interaction terms between neighboring pixels is usually required : the one of convexity or submodularity. This condition is problematic since in practice it allows only to enforce coherency between neighboring pixels and not to enforce variations, while the information given by edges (variations) is usually more reliable than the one given by texture descriptors (because of texture ambiguities, for instance sky and

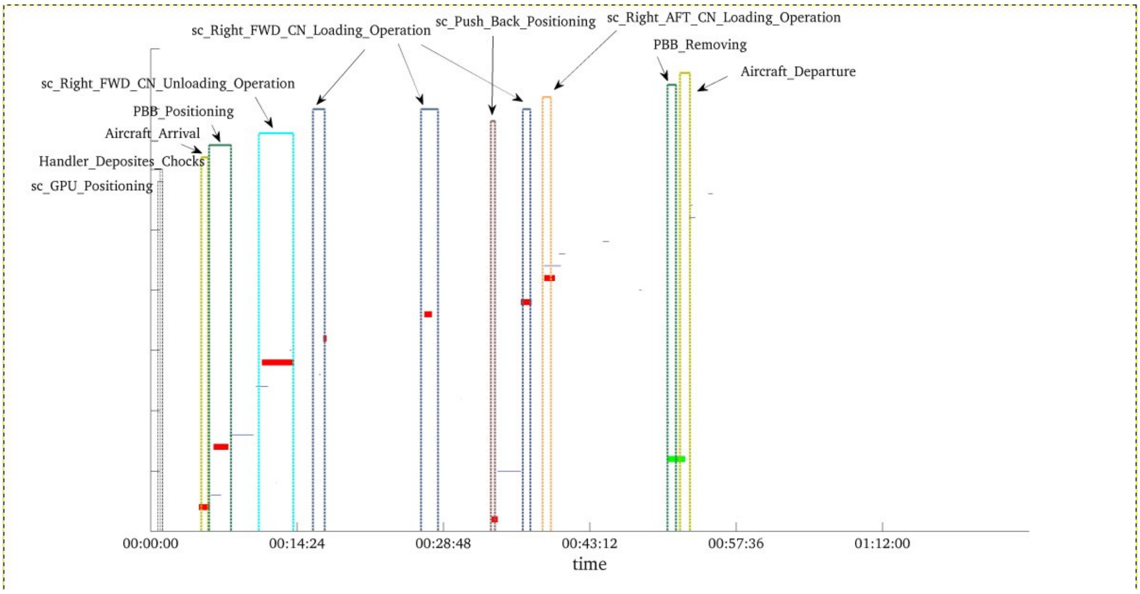


Figure 16. Activity patterns discovered in a sequence of activities taking place while servicing a plane in its dock-station. Red lines indicate a match with activities described in a Ground-Truth. Green line indicates one discovered activity merging two Ground-Truth activities. Blue lines indicate discovered activities not included in the Ground-Truth.

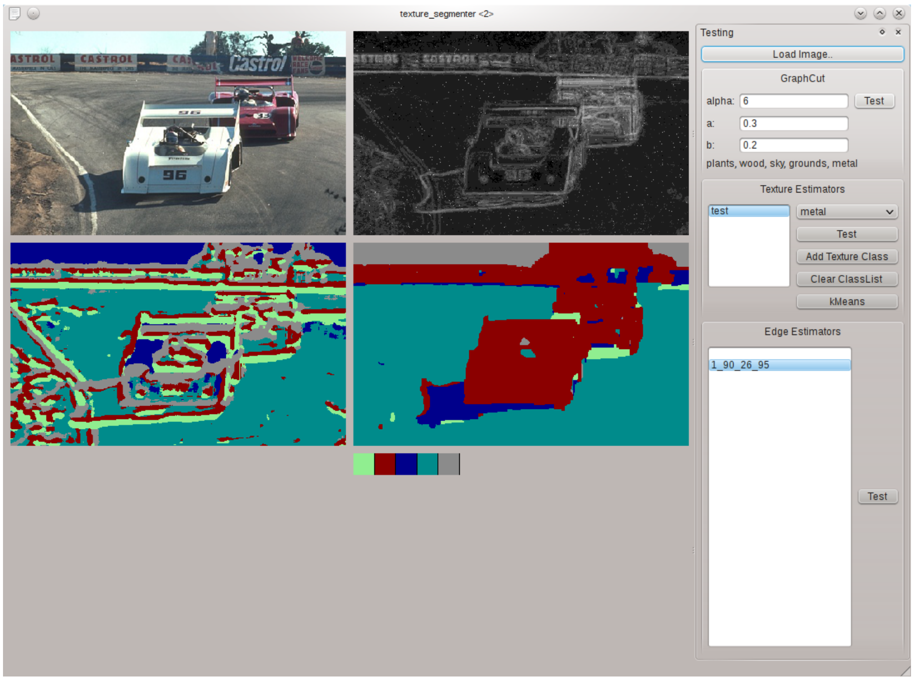


Figure 17. The interface for texture-based segmentation. The user chooses possible texture classes and the segmentation is performed accordingly.

skin textures are very similar). We checked the origin of the submodularity condition and realized that it is specific to a particular way of constructing graphs to represent Markov random fields (MRF). This particular construction is used in all the literature involving graph cuts in computer vision, but no one so far seemed to have noticed that there exist other possible ways to represent MRF as graphs. We found a way to express the (truly) exhaustive family of energies minimizable by a graph cut, whatever graph construction is used. We studied this family and also proposed ways to project MRF that do not belong to the family onto it, in order to solve exactly approximate problems instead of solving approximately the exact problems. We finally exhibited new kinds of energies that can now be minimized with graph cuts, including non-submodular energies which favor certain types of label variations between neighbors, i.e. edges in an image to be segmented.

We developed also another optimization tool for image segmentation [36], with Siqi Chen and Richard Radke from the Rensselaer Polytechnic Institute (USA), dealing with shape optimization in a level-set representation, correcting a mistake often made in the literature of shape evolutions.

6.14. Learning How to Learn and Understand : Automatic Model Construction

Participants: Ezequiel Cura, Guillaume Charpiat.

In scene understanding as in any task to be automatized, one may prefer to replace hand-designed criteria/methods by criteria/methods that are learned, in order to better fit the data or the subject of study. However, paradoxically, current statistical learning approaches show a common drawback : they offer particular schemes for learning, and thus only build pre-defined models, of which only a few parameters can vary. The question is then to know whether it would be possible to learn how to learn, in order not to be constrained to belong to parameterized families of pre-existing algorithms, but to be able to build new ones if needed, to be creative somehow. We formulated the problem as the one to find the best representation of data, based on compactness and efficiency criteria (Kolmogorov and computational complexity). This leads us to hierarchical structures based on redundancy detection in data. The important issues are then strategic : strategies to detect redundancy, meta-strategies to create and test these strategies, etc (see Figure 18). Main concerns are then : optimization methods, creation of efficient heuristics, reasonable use of stochastic approaches.

6.15. SUP Software Platform

Participants: Erwan Demairy, Chedli Farhani, Julien Gueytat, Leonardo Rocha, Christophe Tornieri, Daniel Zullo, François Brémond.

SUP is a software platform developed by PULSAR team, written in C and C++ for generating activity recognition systems. These systems should be able to perceive, analyze, interpret and understand a 3D dynamic scene observed through a network of sensors.

These activity recognition systems are a combination of algorithms developed by members of Pulsar or state of the art computer vision libraries. The SUP dissemination is targeted for use in real-world applications requiring high-throughput.

SUP is made as a framework allowing several computer vision workflows to be implemented. Currently, the workflow is static for a given application but our goal is to make it dynamic. A given workflow is the composition of several plugins, each of them implementing an algorithmic step in the video processing chain (i.e. the segmentation of images, the classification of objects, etc.). The design of SUP allows to execute at run-time the selected plugins.

Several plugins are available:

- 2 plugins are wrappers on industrial implementations of video processing algorithms (made available by Keeneo). They allow a quick deployment of a video processing chain encompassing image acquisition, segmentation, blob construction, classification and short-term tracking. These algorithms are robust and efficient algorithms, but with the drawback that some algorithms can lack accuracy.

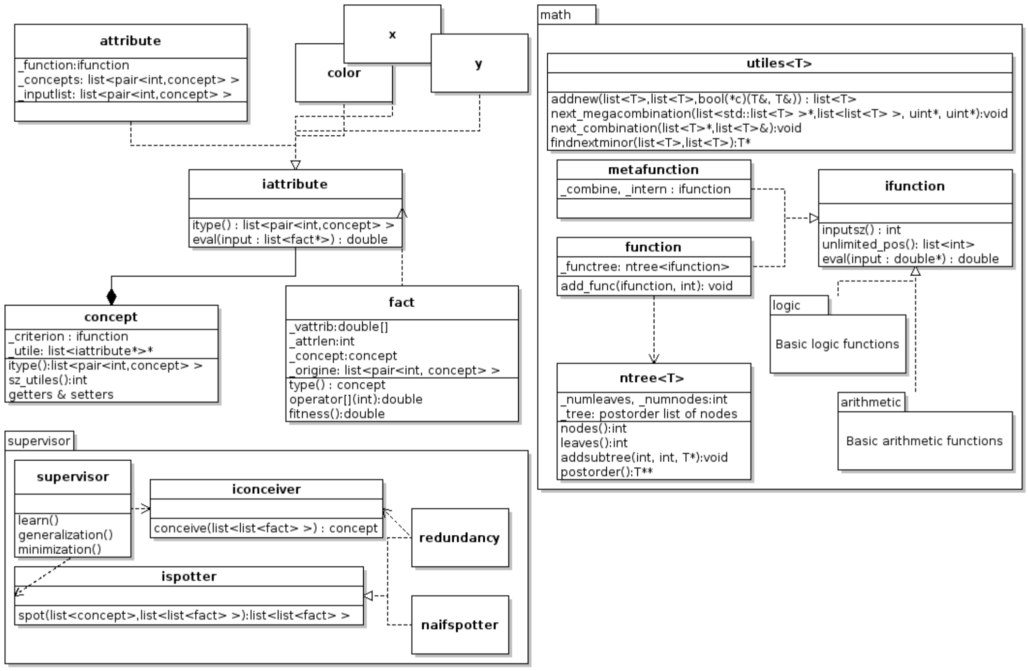


Figure 18. The automatic model constructor.

- Several implementations by the Pulsar team members which cover the following fields:
  1. Image acquisition from different types of image and camera video streaming.
  2. Segmentation removing the shadows.
  3. Two classifiers, one being based on postures and one on people detection.
  4. Four frame-to-frame trackers, using as algorithm:
     1. a simple tracking by overlapping,
     2. neural networks,
     3. tracking of feature points,
     4. tracking specialized for the tracking of people in a crowd.
  5. Three scenario recognizers, one generic algorithm allowing expression of probabilities on the recognized events, the second one focusing on the recognition of events based on postures and the third one (see section Extendable Event Recognition algorithm: SED in this document) uses the complete ontology of the domain as a parameter (e.g. the definition of objects of interest, scenario models, etc.).
  6. 3D animation generation, it generates a virtual 3D animation from information provided by different plugins of the processing chain together with 3D contextual environment.

From a software engineering point-of-view, the goal is to obtain a flexible platform being dynamically reconfigurable for the generated scene understanding systems to be autonomous and adaptable for handling changing environment.

SUP relies on DTK, a generic platform developed by the DREAM service at INRIA Research Center Sophia-Antipolis Méditerranée.

The purpose of DTK is to provide a software infrastructure allowing the generation of a new system by the composition of plugins, each plugin being an algorithmic step of the whole processing chain. SUP is oriented to help developers building activity recognition systems and describing their own scenarios dedicated to specific applications. By relying on the DTK software infrastructure, the possibilities are:

- To simplify the exchanges of algorithms between different INRIA teams using the DTK.
- To use the facilities already provided by the DTK allowing to compose quickly existing plugins. Currently a python interface is operational, and we plan to take advantage of the graphical composer to prototype quickly new work-flows, or reconfigure existing ones, for the experimentation conducted by the team.

In order to be confident on the results obtained with the SUP platform, an important effort is done to check:

- The correct behavior of the platform from a software engineering point of view, i.e. that the functionality of the SUP software is correctly provided, or is not broken by modifications.
- A qualitative evaluation tool (see ViSEvAl in this document) for the algorithms, which compares and assesses the results obtained with the algorithms to ground truth for several reference videos.

Both kinds of test are performed on a daily basis and on several hardware/software architectures.

6.16. Model-Driven Engineering and Video-surveillance

Participants: Santiago Hurtado, Sabine Moisan, Annie Ressouche, Jean Paul Rigault, Sagar Sen, Christophe Tornieri, François Brémond.

We have explored how model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP platform. The challenge is to cope with the many –functional as well as nonfunctional– causes of variability both in the video application specification and in the concrete SUP implementation. Hence, we first decided to separate these two concerns and we applied domain engineering to identify the reusable elements on both sides. This led to two models: a generic model of video surveillance applications (*application model*) and a model of configuration for SUP components and chains (*platform model*) [57]. Both of them are *feature models* expressing variability factors. Feature models compactly define all the features of a system and their valid combinations in an AND-OR graph. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify how the system should adapt to changes in its environment. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [69]. In particular, the inter-feature constraints can be analyzed as a SAT problem.

The application model describes the relevant concepts from the video surveillance stakeholders' point of view: characteristics and position of sensors, context of use (day/night, in/outdoors, target task). It also includes notions of quality of service (performance, response time, detection robustness, configuration cost...). The platform model describes the different software components and their assembly constraints (ordering, alternative algorithms...).

This year we focused on the dynamic variability of the context. Our aim is to cope with possible run-time change of implementation triggered by context variations (e.g., lighting conditions, changes in the reference scene, quality of service degradation, etc.) [56]. A context change at run-time corresponds to a new configuration in the application model and must lead to a new configuration of the platform model. Adaptation rules use the constraint language of feature models (i.e.propositional logic-based language).

We have developed a prototype of a dynamically adaptive vision system, Tekio (signifies adaptation in Japanese). Tekio is based on the Models@Runtime paradigm and contains a master service that orchestrates the different video services at runtime. A change in the context (application model) triggers change in orchestration configuration (platform model). For instance, this may imply tuning service parameters, replacing services by similar ones, or even changing the architecture of the processing chain. Tekio must determine a corresponding configuration that respects all the constraints and must set up this new configuration preferably without shutting down the whole system. The mapping between context and system itself is decided by a reasoning engine service. Moreover, the Tekio system can call services to send back and forth information in form of events/images to other possibly external services. For instance, it can send information to family members about detected accidents of elderly people at home.

Tekio is based on the OSGi standard for service-oriented computing. It is primarily implemented in Java on top of the Eclipse Equinox development platform. It contains a reasoning engine based on the formal specification language Alloy. We are currently performing empirical studies to validate Tekio for Quality of Service testing. We are also studying the scheduling problems raised by "hot swapping" components: this is essential to ensure consistent replacement of services in the orchestration while maintaining continuous activity.

Our first tests use OpenCV [79] as software platform. The following figures illustrate component replacement. In Figure 19, the system is configured to detect movements. When such a movement is detected, the processing chain is reconfigured to perform face detection. This determines whether we have a case of human intrusion in which case the detected face is sent to an external operator. The result of the reconfiguration is shown on Figure 20. Three services are replaced: segmentation, blob construction, and event construction.

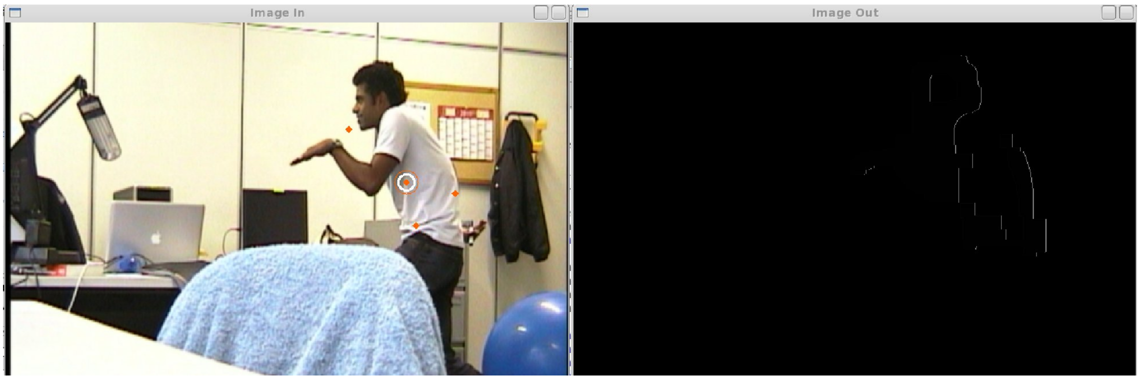


Figure 19. Movement detection context event with red dot indicating center of movement.

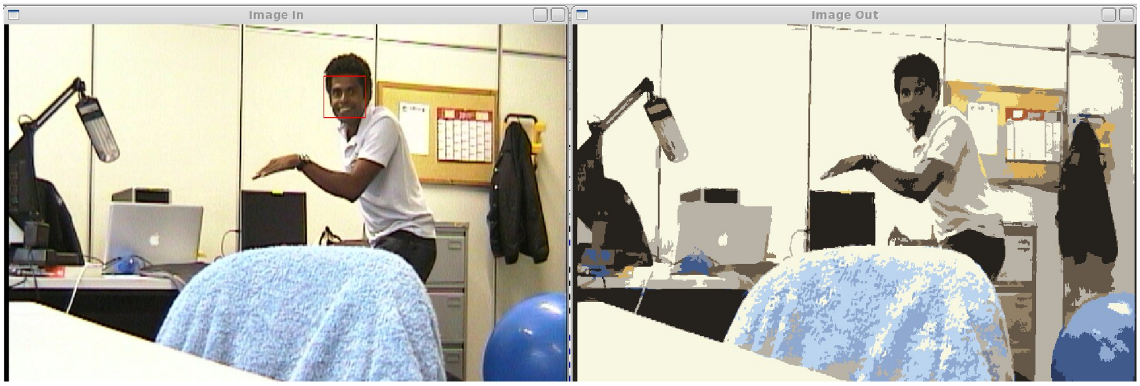


Figure 20. Face detection context event triggered by intrusion detection. Red box indicates face location.

6.17. Scenario Analysis Module

Participants: Bernard Boulay, Sabine Moisan, Annie Ressouche, Jean Paul Rigault.

We are studying and developing a module to analyze scenarios, together with models for scenarios recognition. Scenarios are mostly used to specify the way a system may react to input events from different kinds of sensors. We have started to design a recognition module (SAM) to define, analyze, simulate, and prove scenarios. Our models of scenarios deal with both real time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension and verification). Our purpose is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems [72] and we apply general modeling methods [65] to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model. The SAM module thus provides users with (1) a simulation tool to test scenario behaviors; (2) a generator of a recognizer for each scenario model; (3) an exhaustive verification of safety properties relying on model checking techniques our approach allows. The latter offers also the possibility to define safety properties to prove as “observers” [66] expressed in the scenario language.

This year we have completed SAM in order to address the life cycle of scenario instances. For a given scenario model there may exist several (possibly many) instances at different evolution states. These instances are created and deleted dynamically, according to the input event flow. The challenge is to manage the creation/destruction of this large set of scenario instances efficiently (in time and space), to dispatch events to expecting instances, and to make them evolve independently.

Presently, we rely on an existing synchronous language (Lustre [65]) to express the equational semantics of scenario models and to generate recognizers because this language offers simulation and verification means. But, to improve efficiency, we plan to build our own compiler and to generate recognizers directly from the Boolean equation systems modeling scenario models. This implies that we must supply our own simulation tool and that we interface with a model checking tool as NuSMV [61].

6.18. Multiple Services for Device Adaptive Platform for Scenario Recognition

Participants: Oscar Carrillo, Annie Ressouche, Jean-Yves Tigli.

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.15) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [78], [77], [67]) to dynamically adapt and recompose assemblies of components. These operations must obey the "usage contract" of components. The existing approaches don’t really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behavior models associated with a well sound modeling of composition operation may guarantee the respect of the usage contract.

We began to consider this topic in 2008 through a local collaborative action (SynComp) between Rainbow team (UNS) and Pulsar team. During this collaboration, the management of concurrent access in WComp has been studied as a main source of disturbance for the invariant properties. A modeling of the behavior of components and of their accesses in a synchronous model helps us to answer this problem. This approach allows us to benefit from model checking techniques to ensure that there are no unpredictable states of WComp components reached during a concurrent access. This year, the collaboration between Rainbow and Pulsar has been strengthened since Jean-Yves Tigli was on secondment at Pulsar team up to September 2010. The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioral model of the critical component and model checking techniques apply to verify safety properties about it. Thus, we consider that the critical component is validated. Then we proposed a sound (with respect to our mathematical formalism) composition operation between synchronous components. We proved that this operation preserves already separately verified properties of synchronous components. This operation is an answer to the multiple access to critical components. Actually, we supply a graphical interface to design both critical component behaviors and properties as observers in the synchronous language Lustre [65]. Then the validation of properties and the creation of the validated synchronous component is automatic [54].

6.19. SUP and WComp Platforms Interoperability

Participants: Zouhair Ramadhane, Jean-Yves Tigli, Annie Ressouche.

In this research axis, we propose an approach to activity recognition that combines the use of video cameras with environmental sensors to determine as many activities as possible. This approach consists in analyzing human behaviors and looking for changes in their activities. In particular, the goal is to collect and combine multi-sensor information to detect activities.

The SUP platform gathers a set of modules devoted to design applications in the domain of activity recognition. WComp (see 6.18) is aimed at assembling services which evolve in a dynamic and heterogeneous environment. Indeed, the services provided by SUP can be seen as complex high-level services whose functionalities depend on the SUP treatments; this latter dealing with the dynamic change of the environment. Although SUP can support a huge number of plug-ins and manage different sensors and scenarios, all sensors, primitive events and composite events have to be defined before SUP starts working. Nevertheless, considering SUP services as web services for devices, the devices associated with SUP services will be discovered dynamically by WComp and used with other heterogeneous devices.

Figure 21. SUP

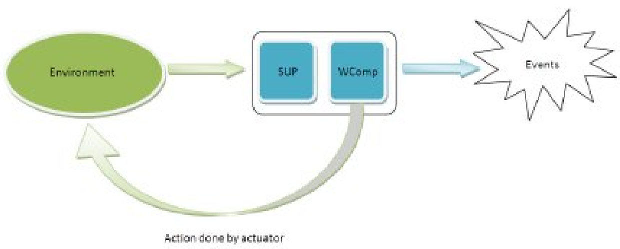


Figure 22. SUP+Wcomp

To cope with this goal, this year we have defined a new model in which data from sensors will be demultiplexed into an heterogeneous entity : the *Actor* which extends SUP Actor entity. In SUP, Actors are a high level abstraction obtained by filtering some tracked physical objects. Thus, Actors are managed as web services in WComp and we have defined in WComp a SUP Actor Generator to generate WComp component for each SUP Actor entity. Then, when a system recognizes a scene it can interact transparently with the environment. We also introduce a dynamic multi-sensor based on an activity recognition approach in which sensors and scenario can be managed while the applications are running. Applying our approach helped us define the scene in an easier way while SUP is running. Our system does not interact with the environment in one way (see Figure 21). It understands it and reacts to it (see Figure 22)

6.20. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit 6 around the language to design, simulate, verify and generate code for programs.

LE language 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 *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.

The equational semantics compute a Boolean equation system and we ensure both modularity and causality in computing all the partial orders valid for a system and we define a way to merge two partial orders. The algorithm which computes partial orders rely on the computation of two dependency graphs: the upstream (downstream) dependency graph computes the dependencies of each variable of the system starting from the input (output) variables. This way of compiling is the corner stone of our approach. Now we have three different algorithms to compute the partial orders valid for an equation system:

1. apply **PERT** method : inputs (resp. outputs) have date 0 and recursively increase of dates for each vertice in the upstream (resp downstream) dependencies graph;
2. apply **graph theory**:
   - compute the adjency matrix $\mathcal{U}$ of upstream (resp. downstream) dependencies graph;
   - the length of the maximal path from a variable v to system inputs is characterized by the maximal k such that $\mathcal{U}^k[v, i] \neq 0$ for all inputs i.
3. apply **fix point theory**: the vector of earliest (resp. lastest) dates can be computed as the least fix point of a momotonic increasing function.

To be modular, we define a technique to compose two already sorted equation systems : first, we memorize the two dependency graphs of equation systems. Second, we define two merge algorithms relying on two different techniques:

1. propagation of common variables dates adjustement;
2. fix point characterisation starting with the vectors of already computed dates and considering only the variables in the dependencies (upstream and downstream) of common variables

The fix point characterization helps us to prove that the merge algorithm is correct (i.e we get the same partial orders using the merge algorithm on two previously sorted equation systems or when sorting the union of the two equation systems considered).

This year we have also improved code generation with *SystemC* and we have written a general journal paper to describe our approach and the *Clem Toolkit* [31]

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 STMicroelectronics, Bull, Thales, Sagem, Alcatel, Bertin, AKKA, Metro of Turin (GTT), Metro of Paris (RATP) and Keeneo.

7.2. National and Regional Initiatives

Pulsar Team has eleven national and regional initiatives: the two first ones concern INRIA cooperative research initiatives (ARC). SIC and VIDEO-ID projects concern the implication of the team in a "pôle de compétitivité" and an ANR project on videosurveillance. Two projects concern long term people monitoring at home. We continue both our collaboration with INRA and our collaboration with STmicroelectronics. A new collaboration is devoted to qualification and certification of perception systems. We also continue our collaboration with IFIP and with Ecole des Mines de Paris.

7.2.1. BioSERRE

This project consists in building a new approach for managing a video camera network for early pest detection in greenhouses. Pulsar cooperates with Vista team (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.

7.2.2. MONITORE

Since April 2008, Pulsar has been involved in an Exploratory Action called MONITORE for the real-time monitoring of imaging diagnostics to detect thermal events in a tore plasma. This work is a preparation for the design of the future ITER nuclear reactor and is done in partnership with the Imaging and Diagnostics Group of the CEA Cadarache.

7.2.3. 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 which is a strategic initiative in perimeter security to address terrorism threats. More precisely the **SIC** project is funded for 42 months starting 2006 with industrial partners including Thales, EADS, BULL, SAGEM, Bertin, Trusted Logic. Two live demonstrations have been done in an official building in Bobigny (Paris) and in one Paris metro station.

7.2.4. 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.

7.2.5. 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 Conseil Général des Alpes Maritimes. 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 supervised and unsupervised event learning and recognition.

7.2.6. CIU-Santé

CIU-Santé (Centre d'Innovation et d'Usages en Santé) is a DGE French project from December 2008 to July 2011 on long-term monitoring of elderly people at Hospital with CSTB, the Nice City Hospital, Pôle SCS, Thales Aliena Space, Actis Ingenierie, Movea, CEA and UNSA. In this project two experimental rooms have been built in Nice-Cimiez Hospital for monitoring Alzheimer patients.

7.2.7. SWEET-HOME

SWEET-HOME is a ANR TECSAN French project from Nov 1st 2009 to 2012 (3 years) on long-term monitoring of elderly people at Hospital with Nice City Hospital, Actis Ingenierie, MICA Center (CNRS unit - UMI 2954) in Hanoi, Vietnam, SMILE Lab at National Cheng Kung University, Taiwan and National Cheng Kung University Hospital. INRIA Grant is 240 Keuros out of 689 Keuros for the whole project. SWEET-HOME project aims at building an innovative framework for modeling activities of daily living (ADLs) at home. These activities can help assessing elderly disease (e.g. Alzheimer, depression, apathy) evolution or detecting pre-cursors such as unbalanced walking, speed, walked distance, psychomotor slowness, frequent sighing and frowning, social withdrawal with a result of increasing indoor hours.

7.2.8. QUASPER: QUAlification et certification des Systèmes de PERception

Quasper R&D is a FUI project with national funding, from June 2010 to may 2012. It gathers 3 objectives to serve companies and laboratories:

1. to encourage R&D and the design of new perception systems;
2. to develop and support the definition of European standards to evaluate the functional results of perception systems;
3. to support the qualification and certification of sensors, software and integrated perception systems.

Target domains are Security, Transportation and Automotive. INRIA participates to the development of the evaluation platform. The coordinator is Thales Services SAS and other partners are Afnor, Akka, Duran, Inrets, Sagem Securite, ST Microelect., Thales RT, Valeo Vision SAS, Cea, Citilog, Institut d'Optique, Civitec, Sopemea, Erte, Hgh.

7.2.9. Semantic Interpretation of 3D Seismic 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's PhD at IFP. The topic is the Semantic Interpretation of 3D seismic images by cognitive vision techniques.

7.2.10. Scenario Recognition

Another collaboration with Ecole des Mines de Paris started in October 2006 through a joint supervision by A. Ressouche and V. Roy of Lionel Daniel's PhD at CMA. The topic is : "Principled paraconsistent probabilistic reasoning - applied to scenarios recognition and voting theory". The PhD has been defended in February 2010.

7.3. European Initiatives

Pulsar team has been involved this year in four European projects: (1) a project on multimedia information processing (VICoMo); (2) a project on machine learning and activity monitoring (COFRIEND); (3) a project concerning the study of autonomous monitoring of complex audio/video surveillance infrastructure(VANAHEIM); (4) a project to study the benefit of a Video-Surveillance system integrated in a global security infrastructure to address potential threats in European ports (SUPPORT).

7.3.1. ViCoMo

ViCoMo is a ITEA 2 European Project which has started on the 1st of October 2009 and will last 36 months.

The project is executed by a strong international consortium, including large high-tech companies (e.g. Philips, Acciona, Thales), smaller innovative SMEs (CycloMedia, VDG Security) and is complemented with relevant research groups and departments from well-known universities (TU Eindhoven, University of Catalonia, Free University of Brussels) and research institutes (INRIA, CEA List, Multitel). Participating countries are France, Spain, Finland, Turkey, Belgium and Netherlands.

The ViCoMo project is focusing on the construction of realistic context models to improve the decision making of complex vision systems and to produce a faithful and meaningful behavior. ViCoMo goal is to find the context of events that are captured by the cameras or image sensors, and to model this context such that reliable reasoning about an event can be perofrmed.

7.3.2. COFRIEND

This project 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 recognize and learn automatically all servicing operations around aircraft parked on aprons.

7.3.3. VANAHEIM (Autonomous Monitoring of Underground Transportation Environment)

VANAHEIM (Video/Audio Networked surveillance system enhAncement through Human-cEntered adaptIve Monitoring) is a FP7-ICT-2009-4 Cognitive Systems and Robotics IP European project which started on February 1st 2010 and will end in 2013 (42 months). The aim of this project is to study innovative surveillance components for the autonomous monitoring of multi-Sensory and networked Infrastructure such as underground transportation environment. The prime partner is Multitel (Belgium) and others partners are: INRIA Sophia-Antipolis (France), Thales Communications (France), IDIAP (Switzerland), Torino GTT (Italy) and Régie Autonome des Transports Parisiens RATP (France), Ludwig Boltzmann Institute for Urban Ethology (Austria) and Thales Communications (Italy). First, we propose to investigate unsupervised modeling techniques for audio/video streams content characterization. Second, moving one step beyond localization-based event recognition, we will investigate human-centered vision techniques for behavior monitoring applications. Conducted studies will be focused on monitoring how human beings act in a given infrastructure, and how they interact with it and with each other.

7.3.4. SUPPORT (Security UPgrade for PORTs)

SUPPORT is a FP7-SEC-2009-1 IP European project which started in June 2010 and will end in 2014 (48 months). The prime partner is BMT Group (UK) and other partners are: INRIA Sophia-Antipolis (France), Swedish Defence Research Agency (SE), Securitas (SE), Technical Research Centre of Finland (FI), MARLO (NO), INLECOM Systems (UK).

SUPPORT is addressing potential threats on passenger life and the potential for crippling economic damage arising from intentional unlawful attacks on port facilities, by engaging representative stakeholders to guide the development of next generation solutions for upgraded preventive and remedial security capabilities in European ports. The overall benefit will be the secure and efficient operation of European ports enabling uninterrupted flows of cargos and passengers while suppressing attacks on high value port facilities, illegal immigration and trafficking of drugs, weapons and illicit substances all in line with the efforts of FRONTEX and EU member states.

7.4. International Initiatives

Pulsar has collaborations with several international teams.

7.4.1. Collaborations with Asia

Pulsar has been cooperating with the Multimedia Research Center in Hanoi MICA on semantics extraction from multimedia data. Pulsar also collaborate with the National Cheng Kung University in Taiwan and I2R in Singapore.

7.4.2. Collaboration with U.S.

Pulsar is collaborating with Guillermo Sapiro's team, University of Minnesota, about the design of new shape distances and shape statistics. Pulsar is collaborating with the Department of Electrical, Computer, & Systems Engineering (ECSE), Rensselaer Polytechnic Institute, namely with Richard J. Radke, Siqi Chen, about the design of shape evolution methods for shape optimization. Pulsar also collaborates with the University of Southern California.

7.4.3. Collaboration with Europ

Pulsar collaborates with Multitel in Belgium and the University of Kingston upon Thames UK.

8. Other Grants and Activities

8.1. Spin off Partner

Keeneo 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. Animation of the scientific community

9.1.1. Journals

- 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.
- 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) and IEEE Transactions on Image Processing (TIP). He also reviewed for the British Machine Vision Conference (BMVC).

- Vincent Martin is a reviewer for the journals: Transactions on Instrumentation & Measurement (TIM), Computers & Electronics in Agriculture (COMPAG)
- M. Thonnat is a reviewer for the journals PAMI (IEEE Transactions Pattern Analysis and Machine Intelligence), CVIU Computer Vision and Image Understanding, Eurasip Journal on Image and Video Processing and gerontechnology Journal.

9.1.2. Conferences and Workshops

- M. Thonnat was a Program Committee member for the following conferences: CVPR10 (23rd IEEE International Conference on Computer Vision and Pattern Recognition), ECCV10 (11th European conference on computer vision), ICPR10 (20th IEEE International Conference on Pattern Recognition).
- M. Thonnat is a Program Committee member for CVPR11 (24th IEEE International Conference on Computer Vision and Pattern Recognition)
- M. Thonnat is general chair of the next International Conference on Computer Vision Systems ICVS 2011.
- J-P Rigault is a member of AITO, the steering committee for several international conferences including in particular ECOOP.

9.1.3. Invited Talks

- G. Charpiat was invited for a talk in the minisymposium *Metric and Riemannian methods in Shape Analysis*, during the SIAM Conference on Imaging Science, in Chicago, April 2010.
- F. Brémond was invited to give a talk at I2R, Singapore on the 10th of August 2010.
- M. Thonnat was invited to give the talk in the honor of A. Bijaoui (astronomer and member of the Academy of Sciences), at the 6th conference on Astronomical Data Analysis in Monastir, Tunisia on the 3rd of May 2010.
- M. Thonnat had an invited talk at the French senate on Services for the elderly in February 2010.
- M. Thonnat had an invited talk at the rencontres INRIA-Industrie on Digital industries for the health in Bordeaux on April the 15th 2010.

9.1.4. Publishing Activities

- F. Brémond is handling editor of the International Journal of Computing and Informatics since September 2006.
- 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.

9.1.5. Expertise

- F. Brémond is EC INFSO Expert in the framework of Ambient Assisted Living European FP7 (HERMES and VitalMind projects) 2007-2010.
- F. Brémond is member of the Scientific Committee of CSOSG 2010 of the French ANR.
- F. Brémond is network member of the 2nd European Network for the Advancement of Artificial Cognitive Systems, Interaction and Robotics (EUCogII network)
- S. Moisan is a member of the Scientific Council of INRA for Applied Computing and Mathematics (MIA Department).
- M. Thonnat was an expert to review a Cognitive System project IST eTRIMMS for the European Commission and to review research projects for the French government DGA.
- M. Thonnat was an expert for reviewing the LIRIS laboratory in Lyon for the National evaluation committee AERES.

- M. Thonnat is member of the scientific board of ENPC, Ecole Nationale des Ponts et Chaussees since June 2008.
- J.-Y. Tigli is member of the expert committee of the Sectoral consulting group (GCS3) of the Ministry for Higher Education and Research, DGRI A3, on "Ambient Intelligence" since 2009.

9.1.6. Miscellanous

- G. Charpiat is organizing a working group about shapes between several teams in INRIA Sophia-Antipolis.
- G. Charpiat's work on image colorization is the subject of a popular science article published by Interstices.
- j-P Rigault is 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.
- A. Ressouche was co-director of Lionel Daniel PhD at Ecole des Mines and reviewer for its defense.
- M Thonnat is deputy scientific director of INRIA in charge of the domain Perception, Cognition and Interaction since September 2009.
- M. Thonnat and F. Brémond are co-founders and scientific advisers of Keeneo, the videosurveillance start-up created to exploit their research results on the VSIP/SUP software.

9.2. Teaching

- Pulsar is a hosting team for the Master of Computer Science of UNSA.
- François Brémond : Teaching at EURECOM Master at Sophia Antipolis, 3hours per year, class on video understanding techniques.
- Annie Ressouche : Teaching at Master of Computer Science at Polytechnic School of Nice Sophia Antipolis University, lecture on Synchronous Languages and Verification (4 h).
- Annie Ressouche: Teaching at Ecole des Mines (Sophia Antipolis), lecture on Scade Toolkit and Verification (4h).
- Jean-Paul Rigault : Full professor at Polytech'Nice Sophia, the engineering school of the University of Nice-Sophia Antipolis ; teaching C, C++ (beginner and advance levels), system programming, software engineering (about 200 hrs per year) ; also teaching advanced C++ programming and object-oriented design for several industrial companies [51].

9.3. Thesis

9.3.1. Thesis in Progress

- Slawomir Bak : People Detection in Temporal Video Sequences by Defining a Generic Visual Signature of Individuals, Nice Sophia-Antipolis University.
- Piotr Bilinski : Gesture Recognition in Videos, Nice Sophia-Antipolis University.
- Duc Phu Chau: Object Tracking for Activity Recognition, Nice Sophia-Antipolis University.
- Emmanuel Mulin : Utilisation de l'Outil Vidéo dans l'Analyse des Troubles du Comportement chez les Patients Atteints de la Maladie d'Alzheimer, Nice Sophia-Antipolis University.
- Guido-Tomas Pusiol : Learning Techniques for Video Understanding, Nice Sophia-Antipolis University.
- Rim Romdhame : Event Recognition in Video Scenes with Uncertain Knowledge, Nice Sophia-Antipolis University.

- Malik Souded : Suivi d'Individu à travers un Réseau de Caméras Vidéo, Nice Sophia-Antipolis University.

9.3.2. Thesis Defended

- 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.

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”", August 2005, vol. 2005:14, p. 2359-2374.

[2] H. BENHADDA, J. PATINO, E. CORVEE, F. BREMOND, 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. BREMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System.*, in "Pattern Recognition Letter.", 2006, vol. 27, n^{o} 15, 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", 1998, vol. 48, 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", March 1993, vol. 57, n^{o} 2, p. 166–184.

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

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

[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", 2007, vol. 16, 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", April 1996, vol. 26, n^{o} 2, p. 113–123.

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

[12] S. MOISAN. *Knowledge Representation for Program Reuse*, in "European Conference on Artificial Intelligence (ECAI)", Lyon, France, July 2002, p. 240-244.

[13] S. MOISAN. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*, Université de Nice-Sophia Antipolis, April 1998, Habilitation à diriger les recherches.

[14] 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", November 2001, vol. 25, n^{o} 4, p. 501-507.

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

[16] J. PATINO, E. CORVEE, F. BREMOND, 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.

[17] 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, Springer, 2008, vol. 150, p. 157-171, selected as one of the 17 best papers of SERA'08 conference.

[18] 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.", November 1992, vol. 9, n^{o} 5, p. 373–387.

[19] 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", 2000, vol. 147, n^{o} 5.

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

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

[22] 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.

[23] 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.

[24] 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.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[25] A.-T. NGHIEM. *Algorithmes Adaptatifs d'Estimation du Fond pour la Détection des Objets Mobiles dans les Séquences Vidéos*, Nice Sophia-Antipolis University, Jun 2010, http://hal.inria.fr/tel-00505881.

[26] N. ZOUBA EP VALENTIN. *Fusion Multicapteurs pour la Reconnaissance d'Activités des Personnes Agées à Domicile*, Nice Sophia-Antipolis University, Jan 2010, http://hal.inria.fr/tel-00453021.

Articles in International Peer-Reviewed Journal

[27] V. MARTIN, J.-M. TRAVERE, F. BREMOND, V. MONCADA, G. DUNAND. *Thermal Event Recognition Applied to Protection of Tokamak Plasma-Facing Components*, in "IEEE Transactions on Instrumentation and Measurement", Apr 2010, vol. 59, n^{o} 5, p. 1182-1191, http://hal.inria.fr/inria-00499599.

[28] S. MOISAN. *Generating Knowledge-Based System Generators: a Software Engineering Approach*, in "International Journal of Intelligent Information Technologies", January 2010, vol. 6, n^{o} 1, p. 1–17.

[29] M. THONNAT. *Special issue on Intelligent Vision Systems*, in "Computer Vision and Image Understanding", May 2010, vol. 114, n^{o} 5, p. pp 501-502, http://hal.inria.fr/inria-00502843.

[30] N. ZOUBA, F. BREMOND, A. ANFOSSO, M. THONNAT, E. PASCUAL, O. GUERIN. *Monitoring elderly activities at home*, in "Gerontechnology", May 2010, vol. 9, n^{o} 2, http://hal.inria.fr/inria-00504703.

Articles in National Peer-Reviewed Journal

[31] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Théorie et Science Informatique", 2010, to appear, http://hal.inria.fr/inria-00524499.

International Peer-Reviewed Conference/Proceedings

[32] S. BAK, E. CORVÉE, F. BRÉMOND, M. THONNAT. *Person Re-identification Using Spatial Covariance Regions of Human Body Parts*, in "7th IEEE International Conference on Advanced Video and Signal-based Surveillance, AVSS - 2010", USA Boston, Aug 2010, http://hal.inria.fr/inria-00496116.

[33] S. BAK, E. CORVÉE, F. BREMOND, M. THONNAT. *Person Re-identification Using Haar-based and DCD-based Signature*, in "2nd Workshop on Activity Monitoring by Multi-Camera Surveillance Systems, AMMCSS 2010, in conjunction with 7th IEEE International Conference on Advanced Video and Signal-based Surveillance, AVSS - 2010", USA Boston, Aug 2010, http://hal.inria.fr/inria-00496051.

[34] I. BECHAR, S. MOISAN. *On-line Counting of Pests in a Greenhouse using Computer Vision*, in "VAIB 2010 - Visual Observation and Analysis of Animal and Insect Behavior", Istanbul, Turkey, Bob Fisher, August 2010, http://hal.inria.fr/inria-00515624.

[35] I. BECHAR, S. MOISAN, M. THONNAT, F. BRÉMOND. *On-line Video Recognition and Counting of Harmful Insects*, in "19th International Conference on Pattern Recognition (ICPR'2010)", Istanbul, Turkey, 2010, http://hal.inria.fr/inria-00515617.

[36] S. CHEN, G. CHARPIAT, R. RADKE. *Converting Level Set Gradients to Shape Gradients*, in "11th European Conference on Computer Vision", Greece Hersonissos, Springer, Sep 2010, http://hal.inria.fr/inria-00497222.

[37] E. CORVÉE, F. BRÉMOND. *Body parts detection for people tracking using trees of Histogram of Oriented Gradient descriptors*, in "AVSS - 7th IEEE International Conference on Advanced Video and Signal-based Surveillance", USA Boston, Sep 2010, http://hal.inria.fr/inria-00496980.

[38] M.-B. KAÂNICHE, F. BRÉMOND. *Gesture Recognition by Learning Local Motion Signatures*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)", USA San Franscico, CA, IEEE Computer Society Press, 2010, http://hal.inria.fr/inria-00486110.

[39] R. KUMAR, V. MARTIN, S. MOISAN. *Robust insect classification applied to real time greenhouse infestation monitoring*, in "Visual observation and analysis of animal and insect behavior (VAIB) Workshop at ICPR", Istanbul, Turkey, August 2010.

[40] T. L. LE, A. BOUCHER, M. THONNAT, F. BRÉMOND. *Surveillance video retrieval: what we have already done?*, in "Third International Conference on Communications and Electronics (ICCE 2010)", Viet Nam Nha Trang, 2010, http://hal.inria.fr/inria-00515574.

[41] V. MARTIN, G. DUNAND, V. MONCADA, J. MICHEL, J.-M. TRAVERE. *New FPGA image-oriented acquisition and real-time processing applied to plasma facing component thermal monitoringa*, in "18th Topical Conference on High-Temperature Plasma Diagnostics", USA Wildwood, NJ, AIP, October 2010, vol. 81, p. 1-4, http://hal.inria.fr/inria-00525445.

[42] E. MULIN, R. ROMDHANE, R. DAVID, J. LEE, N. ZOUBA, J. PIANO, A. DERREMAUX, F. BRÉMOND, P. ROBERT. *Intérêt des gérontechnologies dans la mesure de l'apathie chez les patients atteints de maladie d'Alzheimer*, in "9ème Congrés International Francophone de Gériatrie et Gérontologie (CIFGG 2010)", Nice (France), October 2010.

[43] J. L. PATINO VILCHIS, F. BRÉMOND, M. THONNAT. *Activity discovery from video employing soft computing relations*, in "2010 IEEE International Joint Conference on Neural Networks", Espagne Barcelone, IEEE, July 2010, http://hal.inria.fr/inria-00503047.

[44] J. L. PATINO VILCHIS, F. BREMOND, M. EVANS, A. SHAHROKNI, J. FERRYMAN. *Video Activity Extraction and Reporting with Incremental Unsupervised Learning*, in "7th IEEE International Conference on Advanced Video and Signal-based Surveillance", USA Boston, Aug 2010, http://hal.inria.fr/inria-00502836.

[45] G. PUSIOL, F. BRÉMOND, M. THONNAT. *Trajectory Based Activity Discovery*, in "7th IEEE International Conference on Advanced Video and Signal-based Surveillance", USA Boston, Aug 2010, http://hal.inria.fr/inria-00504634.

[46] P. ROBERT, R. ROMDHANE, F. BRÉMOND, E. MULIN. *Clinical Assessment using Information and Communication Technologies in Alzheimer Disease: Interest for Clinical Trials*, in "Proccedings of the third Conference Clinical Trials on Alzheimer's Disease", Toulouse (France), November 2010.

[47] R. ROMDHANE, F. BRÉMOND, M. THONNAT. *A framework dealing with Uncertainty for Complex Event Recognition*, in "AVSS 2010", USA Boston, Aug 2010, http://hal.inria.fr/inria-00502665.

[48] N. ZOUBA, F. BRÉMOND, M. THONNAT. *An Activity Monitoring System for Real Elderly at Home: Validation Study*, in "7th IEEE International Conference on Advanced Video and Signal-based Surveillance AVSS10", USA Boston, Aug 2010, http://hal.inria.fr/inria-00503864.

Scientific Books (or Scientific Book chapters)

[49] G. CHARPIAT, I. BEZRUKOV, Y. ALTUN, M. HOFMANN, B. SCHÖLKOPF. *Machine Learning Methods for Automatic Image Colorization*, in "Computational Photography: Methods and Applications", R. LUKAC (editor), CRC Press, October 2010.

[50] G. CHARPIAT, M. HOFMANN, B. SCHÖLKOPF. *Kernel methods in medical imaging*, in "Handbook of Biomedical Imaging", N. PARAGIOS, J. DUNCAN, N. AYACHE (editors), Springer, August 2010, http://hal.inria.fr/inria-00481569.

[51] S. MOISAN, J.-P. RIGAULT. *Teaching Object-Oriented Modeling and UML to Various Audiences*, in "Models in Software Engineering", S. GHOSH (editor), LNCS, Springer, 2010, vol. 6002, p. 40–54.

[52] J. L. PATINO VILCHIS, F. BRÉMOND, M. THONNAT. *Incremental learning on trajectory clustering*, in "Intelligent Paradigms in Safety and Security", P. REMAGNINO (editor), Springer-Verlag, December 2010, http://hal.inria.fr/inria-00503059.

[53] J. L. PATINO VILCHIS, F. BRÉMOND, M. THONNAT. *Intelligent Paradigms in Safety and Security*, in "Incremental learning on trajectory clustering", P. REMAGNINO (editor), Springer-Verlag, 2010, To be published in 2011.

Other Publications

[54] O. CARRILLO. *Composition et validation formelle dans une plateforme à base de composants dédiée à la conception d'applications en informatique ambiante*, Ecole Polytech'Nice, 2010.

[55] Z. RAMADHANE. *Plate-forme de composition de services pour dispositifs et de validation pour l'interprêtation de scènes pour la reconnaissance d activités en informatique ambiante*, ENIT Tunis, 2010.

References in notes

[56] M. ACHER, P. COLLET, F. FLEUREY, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Modeling Context and Dynamic Adaptations with Feature Models*, in "Models@run.time Workshop", Denver, CO, USA, October 2009, http://CEUR-WS.org/Vol-509/.

[57] M. ACHER, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Tackling High Variability in Video Surveillance Systems through a Model Transformation Approach*, in "ICSE'2009 - MISE Workshop", Vancouver, Canada, May 2009.

[58] D. S. BATORY. *Feature Models, Grammars, and Propositional Formulas*, in "SPLC", LNCS, 2005, vol. 3714, p. 7-20.

[59] C. CARINCOTTE, F. BRÉMOND, J.-M. ODOBEZ, L. PATINO, B. RAVERA, X. DESURMONT. *Multimedia Knowledge-based Content Analysis over Distributed Architecture*, in "The Networked and Electronic Media, 2009 NEM Summit - Towards Future Media Internet", Saint Malo, France, September 2009.

[60] D. P. CHAU, E. CORVÉE, F. BRÉMOND, M. THONNAT. *Robust Mobile Object Tracking based on Multiple Feature Similarity and Trajectory Filtering*, in "International Conference on Computer Vision Theory and Applications (VISAPP)", March 2011.

[61] A. CIMATTI, E. CLARKE, E. GIUNCHIGLIA, F. GIUNCHIGLIA, M. PISTORE, M. ROVERI, R. SEBASTIANI, A. TACCHELLA. *NuSMV 2: an OpenSource Tool for Symbolic Model Checking*, in "Proceeeding CAV", Copenhagen, Danmark, E. BRINKSMA, K. G. LARSEN (editors), LNCS, Springer-Verlag, July 2002, n^o 2404, p. 359-364, http://nusmv.fbk.eu/NuSMV/papers/cav02/ps/cav02.ps.

[62] K. CZARNECKI, S. HELSEN, U. EISENECKER. *Staged Configuration through Specialization and Multilevel Configuration of Feature Models*, in "Software Process: Improvement and Practice", 2005, vol. 10, n^o 2, p. 143-169.

[63] ECVISION. *A Research Roadmap of Cognitive Vision*, IST Project, 2005, n^o IST-2001-35454.

[64] 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.

[65] N. HALBWACHS. *Synchronous Programming of Reactive Systems*, Kluwer Academic, 1993.

[66] N. HALBWACHS, F. LAGNIER, P. RAYMOND. *Synchronous observers and the verification of reactive systems*, in "Third Int. Conf. on Algebraic Methodology and Software Technology, AMAST'93", Twente, M. NIVAT, C. RATTRAY, T. RUS, G. SCOLLO (editors), Workshops in Computing, Springer Verlag, June 1993.

[67] V. HOURDIN, J.-Y. TIGLI, S. LAVIROTTE, M. RIVEILL. *Context-Sensitive Authorization for Asynchronous Communications*, in "4th International Conference for Internet Technology and Secured Transactions (ICITST)", London UK, November 2009.

[68] K. KANG, S. COHEN, J. HESS, W. NOVAK, S. PETERSON. *Feature-oriented Domain Analysis (FODA) Feasibility Study*, November 1990, n^o CMU/SEI-90-TR-21.

[69] C. KÄSTNER, S. APEL, S. TRUJILLO, M. KUHLEMANN, D. BATORY. *Guaranteeing Syntactic Correctness for All Product Line Variants: A Language-Independent Approach*, in "TOOLS (47)", 2009, p. 175-194.

[70] L. PATINO, H. BENHADDA, E. CORVEE, F. BREMOND, M. THONNAT. *Video-data modelling and discovery*, in "4th IET International Conference on Visual Information Engineering, VIE 2007, London, UK", 25th -27th July 2007.

[71] L. PATINO, H. BENHADDA, E. CORVEE, F. BREMOND, M. THONNAT. *Extraction of Activity Patterns on large Video Recordings*, in "IET Computer Vision", june 2008, vol. 2, n^o 2, p. 108–128.

[72] A. PNUELI, D. HAREL. *On the Development of Reactive Systems*, in "Nato Asi Series F: Computer and Systems Sciences", K. APT (editor), Springer-Verlag berlin Heidelberg, 1985, vol. 13, chap. Logics and Models of Concurrent Systems, p. 477-498.

[73] G. PUSIOL, L. PATINO, E. CORVEE, F. BREMOND, M. THONNAT, S. SURESH. *Optimizing trajectory clustering for activity recognition*, in "1st International workshop on machine learning for vision-based motion analysis (MLVMA 08)", Marseille, France, 18th October 2008.

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

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

[76] M. SVAHNBERG, J. VAN GURP, J. BOSCH. *A taxonomy of variability realization techniques: Research Articles*, in "Software Practice and Experience", 2005, vol. 35, n^{o} 8, p. 705–754.

[77] J.-Y. TIGLI, S. LAVIROTTE, G. REY, V. HOURDIN, D. CHEUNG, E. CALLEGARI, M. RIVEILL. *WComp middleware for ubiquitous computing: Aspects and composite event-based Web services*, in "Annals of Telecommunications", 2009, vol. 64, n^{o} 3-4, ISSN 0003-4347 (Print) ISSN 1958-9395 (Online).

[78] J.-Y. TIGLI, S. LAVIROTTE, G. REY, V. HOURDIN, M. RIVEILL. *Lightweight Service Oriented Architecture for Pervasive Computing*, in "IJCSI International Journal of Computer Science Issues", 2009, vol. 4, n^{o} 1, ISSN (Online): 1694-0784 ISSN (Print): 1694-0814.

[79] WILLOW GARAGE. *OpenCV*, http://www.willowgarage.com/pages/software/opencv.