Activity Report 2011

Project-Team PULSAR

Perception Understanding Learning Systems for Activity Recognition

RESEARCH CENTER
Sophia Antipolis - Méditerranée

THEME
Vision, Perception and Multimedia Understanding

Table of contents

Project-Team PULSAR

Keywords: Computer Vision, Real-Time, Safety, Security, Semantics, Software Engineering, Video, Perception, Machine Learning

1. Members

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]

External Collaborators

Daniel Gaffé [Assistant professor, Faculty Member Nice University and CNRS-LEAT Member]
Veronique Joumier [Researcher Engineer, CHU Nice University]
Vincent Martin [Researcher engineer CEA]
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]

Technical Staff

Bernard Boulay [Development Engineer, COFRIEND and QUASPER Projects]
Etienne Corvée [Development Engineer, COFRIEND, VANAHEIM and VideoID Projects]
Hervé Falciani [Development Engineer, [EIT ITC Labs, since September 2011]
Julien Gueytat [Development Engineer, SWEET HOME Project]
Srinidhi Mukanahallipatna [Development Engineer, PAL AEN,since June 2011]
Jose-Luis Patino Vilchis [Development Engineer, COFRIEND and VANAHEIM Projects]
Leonardo Rocha [Development Engineer, CIU Santé, SWEET HOME and VICOMO Projects]
Ricardo-Cezar Rodrigues [Development Engineer, THALES, since July 2011]
Silviu-Tudor Serban [Development Engineer, QUASPER Project, since September 2011]
Sofia Zaidenberg [Development Engineer, VANAHEIM Project]
Daniel Zullo [Development Engineer, CIU Santé and VICOMO Projects, up to October 2011]

PhD Students

Julien Badie [Nice Sophia-Antipolis University, SWEET HOME Grant, since October 2011]
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]
Carolina Garate [Nice Sophia-Antipolis University, VANAHEIM Grant, since August 2011]
Ratnesh Kumar [Nice Sophia-Antipolis University, VANAHEIM Grant, since January 2011]
Emmanuel Mulin [Nice Sophia-Antipolis University, CHU Nice]
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]

Post-Doctoral Fellows

Carlos -Fernando Crispim Junior [PAL AEN, from November 2011]
Sagar Sen [Project VICOMO, up to August 2011]

Administrative Assistant

Christine Claux [AI Inria]

Others

Vasanth Bathrinarayanan [VICOMO Project, since April 2011 up to December 2011]
Julien Badie [QUASPER Project, since April 2011 to September 2011]
Carolina Garate [VANAHEIM Project, up to July 2011]
Nedra Nefzi [THALES, since May 2011 up toOctober 2011]
Jihed Joobeur [VICOMO Project, since April 2011]
Kandan Ramakrishnan [EGIDE Grant, since January 2011 up to June 2011]
Swaminathan Sankaranarayanan [QUASPER project, since September 2011]
Allen Wu [EGIDE Grant, since April 2011 up to September 2011]

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 COFRIEND, VANAHEIM, SUPPORT, DEM@CARE, VICOMO. 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 which designs cognitive vision systems for activity recognition based on sound software engineering paradigms. We have an operational platform, named SUP, for detecting and tracking mobile objects, which can be either humans or vehicles, and for 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 designed an efficient algorithm for detecting people in a static image based on a cascade of classifiers. We have also proposed a new algorithm for re-identification of people through a camera network. We have realized a new algorithm for the recognition of short actions and tested its performance on several benchmarking databases. We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. 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 frequent activities at the apartment of an elderly, in a subway station and on an airport tarmac.

We have also started two clinical trials to characterize the behaviour profile of Alzheimer patients compared to healthy older people.

Monique Thonnat was general chair of the International Conference on Computer Vision Systems (ICVS 2011).

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 formalizing 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 optimization 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: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, François Brémond.

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.

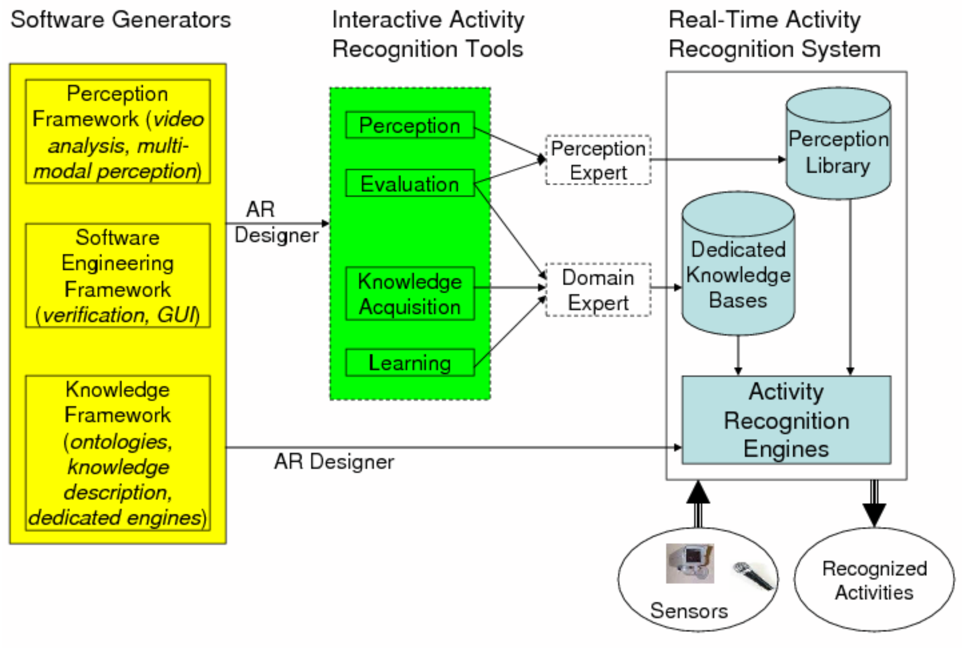


Figure 1. Global Architecture of an Activity Recognition Platform.

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) [73] 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* [65], [58], [74], [54] 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 [60]. 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

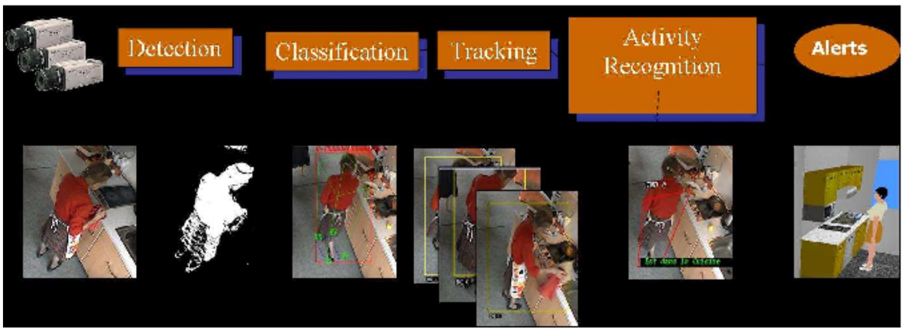


Figure 2. Components of the Scene Understanding Platform (SUP).

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 visualization 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, visualization and evaluation) respects three important properties:

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

The ViSEvAl tool is composed of two parts: a GUI to visualize 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.

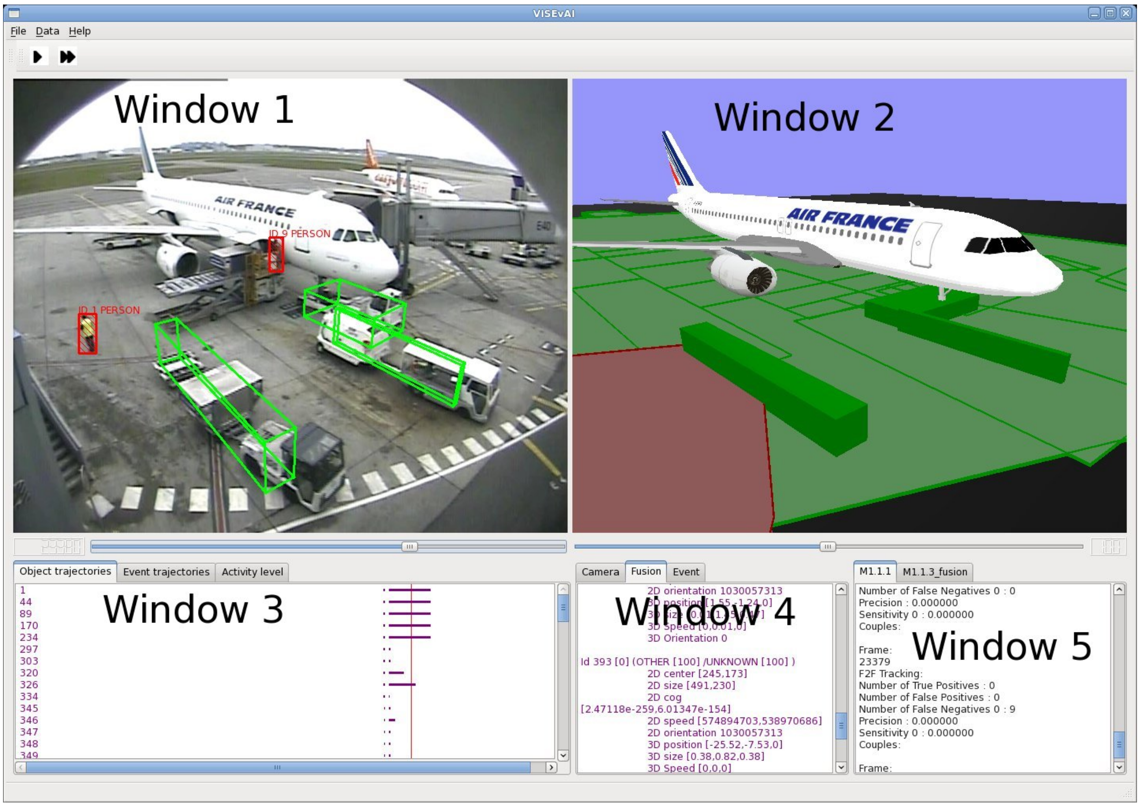


Figure 3. GUI of the ViSEvAl software

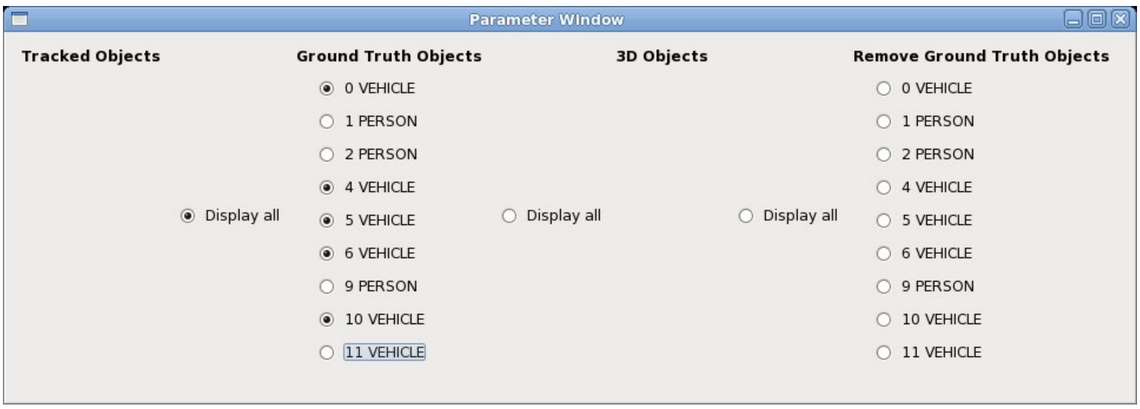


Figure 4. The object windows enables users to choose the object to display

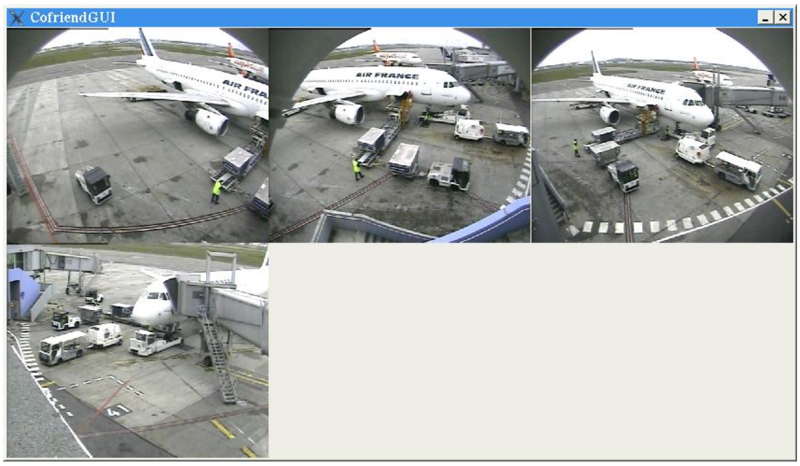


Figure 5. The multi-view window

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. ViSEvAl is under GNU Affero General Public License AGPL (http://www.gnu.org/licenses/) since July 2011. 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* [61](see Figure 6) is a set of tools devoted to design, simulate, verify and generate code for LE [17] [71] 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 [57] in order to perform validation of program behaviors.

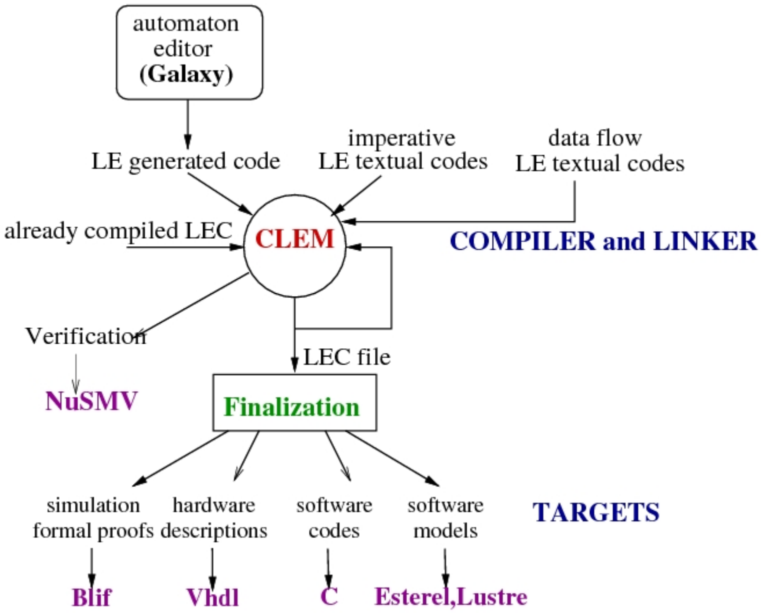


Figure 6. The Clem Toolkit

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, Piotr Bilinski, Bernard Boulay, François Brémond, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Julien Gueytat, Ratnesh Kumar, Vincent Martin, Sabine Moisan, Emmanuel Mulin, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Leonardo Rocha, Rim Romdhame, Silviu Serban, Malik Souded, Monique Thonnat, Sofia Zaidenberg, Daniel Zullo.

This year Pulsar has proposed new algorithms in computer vision (people head and face detection and people re-identification), in reasoning (activity recognition and uncertainty handling). More precisely, the new results for this research axis concern:

- People detection in monocular video sequences (6.2)
- Online Parameter Tuning for Object Tracking Algorithms (6.3)
- Fiber Based Video Segmentation (6.4)
- Multiple Birth and Cut Algorithm for Multiple Object Detection (6.5)
- Exhaustive Family of Energies Minimizable Exactly by a Graph Cut (6.6)
- Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves (6.7)
- Object Tracking Using a Particle Filter based on SIFT Features (6.8)
- Human Re-identification using Riemannian Manifolds (6.9)
- Global Tracking of Multiples Actors (6.10)
- Crowd Data Collection from Video Recordings (6.11)
- Events Recognition and Performance Evaluation (6.12)
- Group interaction and group tracking for video-surveillance in underground railway stations (6.13)
- Action Recognition in Videos (6.14)
- Activity Recognition Applied on Health Care Application (6.15)
- A Cognitive Vision System for Nuclear Fusion Device Monitoring (6.16)
- Scenario Recognition with depth camera (6.17)
- Trajectory Clustering for Activity Learning (6.18)

6.1.2. Software Engineering for Activity Recognition

Participants: François Brémond, Bernard Boulay, Hervé Falciani, Daniel Gaffé, Julien Gueytat, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Sagar Sen, 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 improve our meta-modeling approach to support the development of video surveillance applications based on SUP. We continue the development of a scenario recognition module relying on formal methods to support activity recognition in SUP platform. We also continue to study the definition of multiple services for device adaptive platform for scenario recognition. Finally, we are implementing the new theoretical results obtained last year to improve the Clem toolkit.

The new results related to this research axis concern:

- SUP Software Platform (6.19)
- Model-Driven Engineering and Video-surveillance (6.20)
- Scenario Analysis Module (6.22)
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.23)
- The Clem Toolkit (6.24)

6.2. People detection in monocular video sequences

Participants: Etienne Corvee, François Brémond, Silviu-Tudor Serban, Vasanth Bathrinaryanan.

A video understanding system analyzes human activity by detecting people in video sequences and tracking their displacement and movement throughout the sequences. The better the detection quality, the higher the semantic level of the information is. People activty can differ greatly from one application to another e.g. the presence of a person in one zone can simply be detected from a moving pixel region in a manually specified zone whereas detecting people fighting in a subway requires more complex information. For people activity to be recognized, one needs to detect people accurately in videos and at real time frame rate. Current state of the art algorithms provide generic people detection solutions but with limited accuracy. In the people monitoring domain, although cameras remain mostly fixed, many issues occur in images. For example, outdoor scenes display strong varying lighting conditions (e.g. sunny/cloudy illumination, important shadows), public spaces can be often crowded (e.g. subways, malls) and images can be obtained with a low resolution and can be highly compressed. Hence, detecting and tracking objects in such complex environment remains a delicate task to perform. In addition, detecting people has to face one major difficulty which is caused by occlusion where important information is hidden. When people overlap onto the image plane, their foreground pixels cannot be separated using a standard thresholding operation from a background reference frame. Therefore vision algorithms need to use information held by the underlying pixels and located at specific locations such as body parts.

We have extended our work by implementing and testing a novel people, head and face detection algorithm using Local Binary Pattern based features and Haar like features. The traditional and efficient Adaboost training scheme is adopted to train object features from publicly available databases. This work has been published in the ICVS Workshop [36].

Figure 7. Example of tracked people, head and face

The work has been tested for group tracking in Vanaheim videos (see section 8.2.1.2) and for people tracking in VideoId videos. The VideoId project aims to re-identify people across a network of non overlapping cameras using iris, face and human appearance recognition. An example of tracked people, head and faces in a testing database is shown in figure 7. An example of re-identified face is shown in figure 8 by the VideoId interface in a Paris underground video.

We have evaluated our people detection algorithm on the test human dataset provided by INRIA against state of the art algorithms which we refer as HOG [59] and LBP-HOG [77]. The INRIA human dataset is composed of 1132 human images and 453 images of background scenes containing no human. The results are displayed in figure 9 which shows that we obtain slightly better performances than the HOG-LBP technique in terms of missed detection rate vs. FPPI i.e. False Positive Per Image. In this figure, two extreme functioning modes

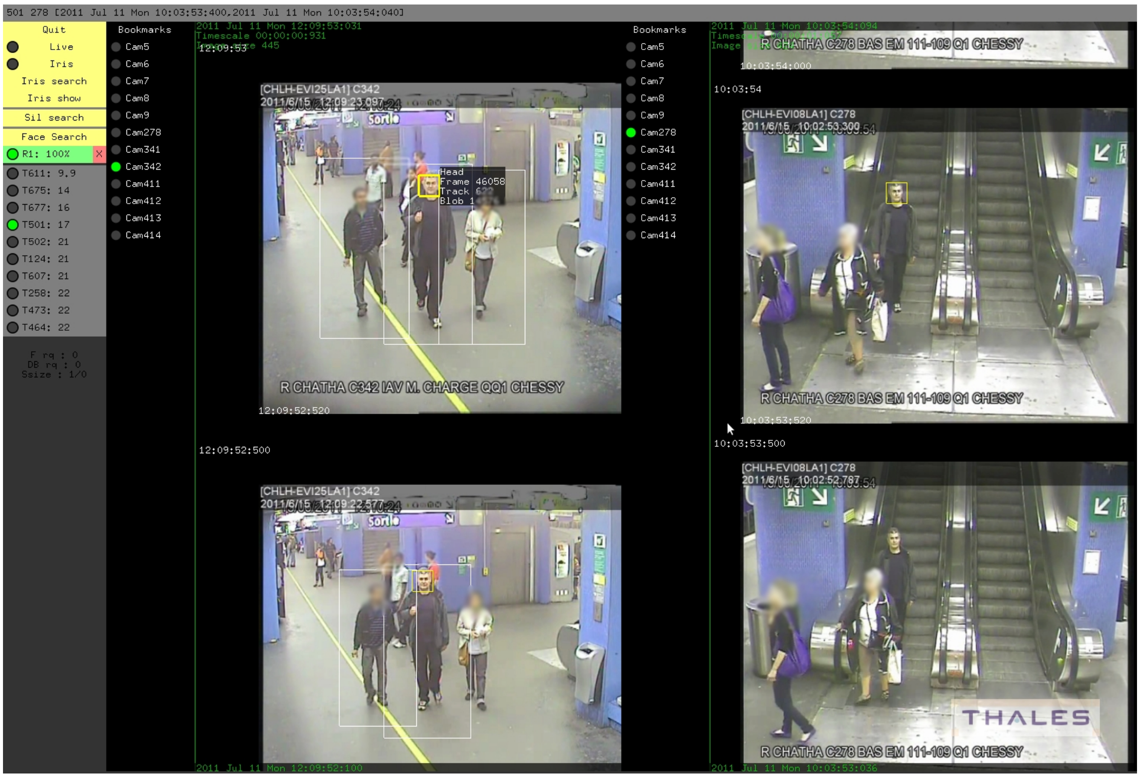


Figure 8. Face recognition in Paris underground

could be chosen: approximately 2 noisy detections are obtained every 1000 background images for 50% true positive detections or 1 noisy detection every 2 frames for a detection rate of approximately 88%.

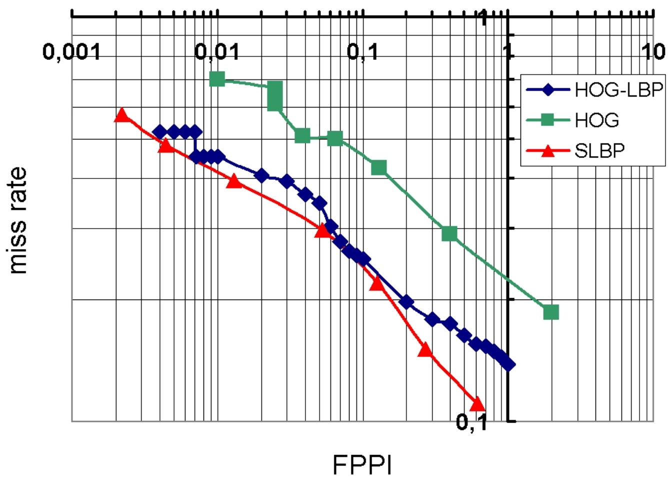


Figure 9. People detection evaluation: False Positive Per Image vs. miss detection rate for the INRIA test database

The same evaluation scheme of people detection above is used for face detection evaluation. The FPPI rates are obtained on 997 NICTA [66] background images of 720x576 pixels. 180 faces provided by a CMU test face image database are used to evaluate true positive rates. We have compared our results with the 2 versions of Haar feature provided by the OpenCv library i.e. the standard 'default' and alternative 'alt' training parameters. The results in table 1 show that the Haar 'alt' technique performs better than the traditional Haar one. And our haar based technique called CCR provides similar face detection rates while giving a less false alarm rate. The proposed approach is approximately 1% less successful in detecting faces than the Haar technique while this latter is 32% more noisier than our CCR technique.

Table 1. Face detection evaluation

technique	TP(%)	FPPI
Haar (default)	91.57	4.132
Haar (alt)	92.13	1.685
CCR	91.01	1.274

6.3. Online Parameter Tuning for Object Tracking Algorithms

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

Many approaches have been proposed to track mobile objects in a scene. However the quality of tracking algorithms always depends on scene properties such as: mobile object density, contrast intensity, scene depth and object size. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has appropriately selected, it is difficult to tune online its parameters to get the best performance.

Therefore we propose a new control approach for mobile object tracking. More precisely in order to cope with the tracking context variations, this approach learns how to tune the parameters of object appearance-based tracking algorithms. The tracking context of a video sequence is defined as a set of features: density of mobile objects, their occlusion level, their contrasts with regard to the background and their 2D areas. In an offline supervised learning phase, satisfactory tracking parameters are searched for each training video sequence. Then these video sequences are classified by clustering their contextual features. Each context cluster is associated with the learned tracking parameters. In the online control phase, two approaches are proposed. In the first one, once a context change is detected, the tracking parameters are tuned for the new context using the learned values. In the second approach, the parameter tuning is performed when the context changes and the tracking quality (computed by an online performance evaluation algorithm [56]) is not good enough. An online learning process enables to update the context/parameter relations.

We have also proposed two new tracking algorithms to experiment the proposed control method. The first tracker relies on a Kalman filter and a global tracking which aims at fusing trajectories belonging to the same mobile object. This work has been published in [35]. The second tracker relies on the similarities of eight object descriptors (2D and 3D positions, area, shape ratio, HOG, color histogram, color covariance and dominant color) to build object trajectories. This work has been published in [34].

(a) CARETEKER (b) CAVIAR

Figure 10. (a) CARETEKER: Illustration of the Caretaker video; (b) CAVIAR: Illustration of the Caviar video

The proposed controller has been experimented on a long, complex video belonging to the Caretaker European project [1] (see figure 10(a)) and 26 videos of Caviar dataset [2] (see figure 10(b)). For the Caretaker video, when the controller is used, the tracking quality increases from 52% to 78%. For the Caviar dataset, the experimental results show that the tracking performance increases from 78.3% to 84.4% when using the controller. The tracking results on Caviar videos with the proposed controller are as good as the ones obtained with manual parameter tuning.

6.4. Fiber Based Video Segmentation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (space$\times$time). Figure 11 shows 2D slices of a video volume. Bottom right corner of each figure shows the current temporal depth in the volume, while top right shows the X-time slice and bottom left shows Y-time slice. In this 3D representation of videos, points of static background form straight lines of homogeneous intensity over time, while points of moving objects form curved lines. Analogous to the fibers in MRI images of human brains, we name fibers, these straight and curved lines of homogeneous intensity. So, in our case, to segment the whole video volume data, we are interested in a dense estimation of fibers involving all pixels.

[1] http://cordis.europa.eu/ist/kct/caretaker_synopsis.htm
[2] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

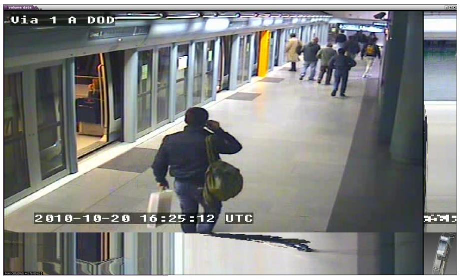


Figure 11.

For the detection of these fibers, we use motion flow vectors and intensity correlation of 2D patches over time. As these techniques are not reliable everywhere in the image domain, we sort the fibers based on the reliability of the detections from these techniques. The subsequent goal is then to pick high ranked fibers to propagate motion information and boundary fronts to other regions of the 3D volume.

To reliably propagate information from a fiber, we express the reliability of detection of a fiber and the cost of propagation of information from it. The later can be based on a distance measure of a pixel from a fiber, while reliability of a fiber involves motion coherency, color homogeneity, duration along time axis etc.

Our work closely relates to [72]. A video is represented by a set of particles (trajectory of an image point sample). The algorithm then extends and truncates particle trajectories to model motion near occlusion boundaries.

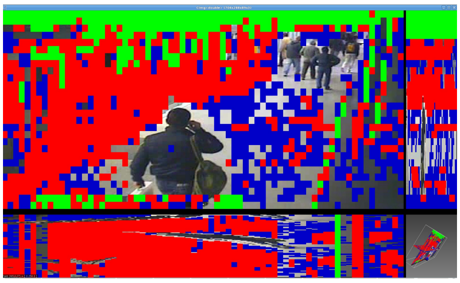


Figure 12. Video Volume and Colored Straight Fibers : Green - Most reliable, Blue - Least reliable

Figure 12 shows some straight fibers found in a video volume. The reliability of these fibers is based on temporal length. Fibers which have temporal span same as that of the video are colored in green, while fibers which have temporal span of less than 10% of the video are colored blue. Red colored fibers have temporal length in between green and blue colored fibers.

6.5. Multiple Birth and Cut Algorithm for Multiple Object Detection

Participant: Guillaume Charpiat.

In collaboration with the Ariana team (Ahmed Gamal-Eldin, Xavier Descombes and Josiane Zerubia), we developed a new optimization method which we call Multiple Birth and Cut (MBC). It combines the recently proposed Multiple Birth and Death (MBD) algorithm and the Graph-Cut algorithm. MBD and MBC optimization methods are applied to energy minimization of an object based model, the marked point process. The most important advantage of the MBC over MBD is the reduction of number of parameters. By proposing good candidates throughout the selection phase in the birth step, the speed of convergence is increased. In this selection phase, the best candidates are chosen from object sets by a belief propagation algorithm. The algorithm is applied on the flamingo counting problem in a colony [37], [26].

6.6. Exhaustive Family of Energies Minimizable Exactly by a Graph Cut

Participant: Guillaume Charpiat.

Graph cuts are widely used in many fields of computer vision in order to minimize in small polynomial time complexity certain classes of energies. These specific classes depend on the way chosen to build the graphs representing the problems to solve. We study here all possible ways of building graphs and the associated energies minimized, leading to the exhaustive family of energies minimizable exactly by a graph cut. To do this, we consider the issue of coding pixel labels as states of the graph, i.e. the choice of state interpretations. The family obtained comprises many new classes, in particular energies that do not satisfy the submodularity condition, including energies that are even not permuted-submodular.

We studied in details a generating subfamily, in particular we proposed a canonical form to represent Markov random fields, which proves useful to recognize energies in this subfamily in linear complexity almost surely, and then to build the associated graph in quasilinear time. We performed a few experiments to illustrate the new possibilities offered [33]. We have also started to use this technique to minimize exactly approximations of Markov random field energies instead of minimizing approximately the exact energies, by projecting energies on the family we know to solve globally efficiently.

6.7. Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves

Participant: Guillaume Charpiat.

This is joint work with Gabriel Peyré (CNRS, Ceremade, Université Paris-Dauphine). We intend to favor piecewise-rigid motions, i.e. articulated movements, during shape evolutions, especially when computing morphings or image segmentation with shape prior. To do this, we first need a dissimilarity measure between shapes, whose gradient is meaningful. We formulate one using kernels and bistochastization.

The parameters of these kernels are automatically estimated in a fixed-point scheme that guarantees physical relevance, and the notion of bistochastization is extended to continuous distributions. Finally, piecewise rigidity is ensured during gradient descents by a change of the norm from which the gradient is derived. This norm is formulated so as to favor sparse second derivatives, which produces articulated movements without knowing by advance the location of the articulations.

The formula of the norm is actually elegantly simple, involving simple geometric quantities, derivatives, and the L_1 norm. Note that this norm does not derive from an inner product but defines a gradient in the sense of [5] as the minimizer of an energy. It turns out that in our case the energy defining the gradient is actually convex, and efficient minimization follows.

6.8. Object Tracking Using a Particle Filter based on SIFT Features

Participants: Malik Souded, François Brémond.

The approach consists in detecting SIFT points of interest on the objects to track, calculating their SIFT descriptors, tracking these points with a particle filter, and finally achieving tracking process by linking them along the time with links which are weighted by measures on SIFT descriptors and reliability.

The main contributions in this work are on three points.

The first point consists in techniques of detection and selection of SIFT points, allowing a better distribution of points of interest on the target and allowing better management of partial occlusions, and secondly an optimized computing time thanks to the parallelization of these SIFT computation on modern processors (see figure 13).

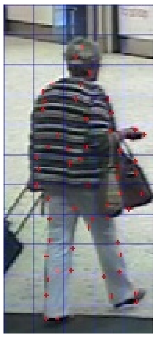


Figure 13. SIFT Points detection and selection

The second point concerns the weighting of the particles during tracking. This is done with a combination of two kinds of information: the similarity measure of the SIFT descriptor and the state of motion of pixels corresponding to the particles. This allows more robust tracking of SIFT points regardless of the quality of the background subtraction providing the detected objects.

The last point concerns the selection of temporal links between tracked objects and detected ones. These links are selected according to their weight. The weight of each link is based on the proportion of common SIFT points to both objects (two successive images) potentially linked, and the reliability of each of these SIFT point. This reliability is calculated for each point by measuring the variation of the SIFT descriptor during the tracking time.

The occlusion management is performed using three types of information: SIFT descriptors used for tracking (matching after reappearance) the dominant colors of the object of interest and finally the width, height, and speed (in real world) of the object, which are learned in Gaussian models during the tracking (tracking being used with video cameras which have been calibrated), see figure 14.

The approach was tested on 121 sequences of four different datasets: 80 sequences from CAVIAR , 34 sequences from ETISEO, 3 sequences from PETS2001 and 2 sequences from VS_PETS2003. The obtained results are satisfying. The comparison of these results with the state of the art shows improvements for the benchmarking dataset (ETISEO) The following table compares the proposed approach and state of the art results on ETISEO data base:

	Metrics	ETI-VS1-BE-18-C4	ETI-VS1-BE-16-C4	ETI-VS1-MO-7-C1
Proposed tracker	M1	0.68	0.54	0.90
Chau et al. (VISAPP 2011)	M1	0.64	0.36	0.87
Best team in ETISEO Project (2005)	M1	0.48	0.44	0.77

This work was published in [49]

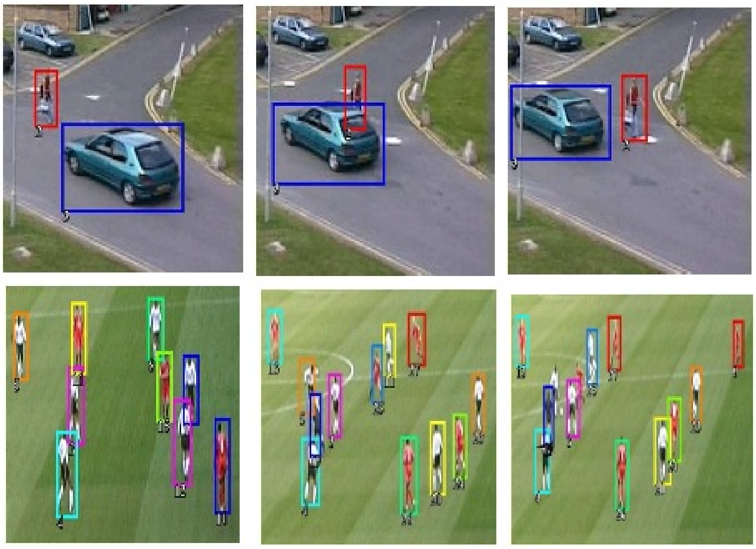


Figure 14. Tracking and Partial/full occlusion management results

6.9. Human Re-identification using Riemannian Manifolds

Participants: Sławomir Bąk, Etienne Corvée, François Brémond, Monique Thonnat.

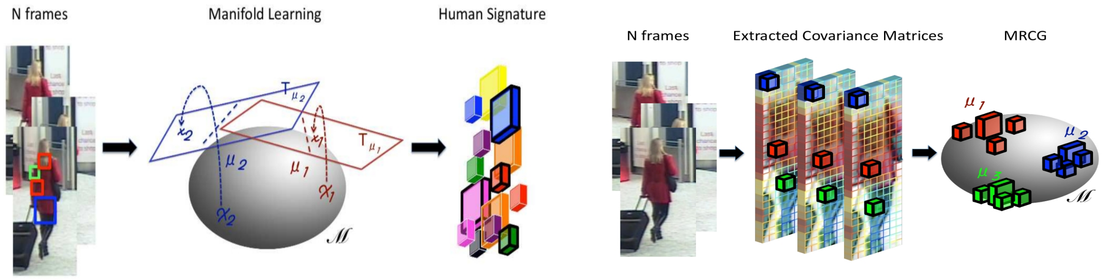
This work addresses the *human re-identification* problem, which is defined as a requirement to determine whether a given individual has already appeared over a network of cameras. This problem is particularly hard by significant appearance changes across different camera views. In order to re-identify people, a human *signature* should handle difference in illumination, pose and camera parameters.

We propose new appearance models based on the *mean riemannian covariance* (MRC) matrices combining the appearance information from multiple images. These mean covariance matrices not only keep information on feature distribution but also carry out essential cues about temporal changes of an appearance. Using MRC-s, we propose two methods for an appearance representation:

- *Learned Covariance Patches* (LCP) [25] - a distinctive representation is extracted by a boosting scheme. The structure of MRC patches (size, position) is learned using boosting algorithm based on confidence-rated predictions (see Fig. 15 (a)). These confidence-rated coefficients are employed to weight appearance characteristics of a specific individual *w.r.t.* the reference (training) dataset of humans.
- *Mean Riemannian Covariance Grid* (MRCG) [30] - less computationally demanding technique than LCP. We represent a human appearance by a grid of MRC *cells* (see Fig. 15 (b)). Relevant *cells* are identified by an efficient discriminant analysis. This analysis takes into account variance of MRC patch in the class of humans (reference dataset). MRC is assumed to be more significant when its variance is larger in the class of humans: (1) the most common patterns belong to the background (the variance is small); (2) the patterns which are far from the rest are at the same time the most discriminative (the variance is large). All operations, such as *mean* or *variance*, are performed on covariance manifold specified as Riemannian.

Both methods are evaluated and compared with the state of the art using publicly available datasets. We demonstrate that the proposed approaches outperform state of the art methods. Further, we extract new sets of individuals from i-LIDS data to investigate more carefully advantages of using many images for human re-identification.

The computation complexity is analyzed in the context of distance operator between two signatures. Comparing two human signatures, it is necessary to compute distance between covariance matrices, which requires solving the generalized eigenvalues problem. This operation is computationally heavy. In [31], we propose an implementation for finding generalized eigenvalues and eigenvectors for distance operator, using NVIDIA GPU architecture. We improve significantly the performance, reaching 66 speedup using Tesla S1070.



(a) LCP (b) MRCG

Figure 15. Computation of the human signature using Riemannian manifold (depicted with the surface of the sphere): (a) LCP - the structure of patches is learned by a boosting algorithm; (b) MRCG - covariances gathered from tracking results are used to compute the MRC-s.

6.10. Global Tracking of Multiples Actors

Participants: Julien Badie, François Brémond.

We propose a new approach for long term tracking of individuals. Our main objective is to design a tracking algorithm for people reidentification [30] that can track people even if they come back in the scene after leaving it. This algorithm is based on covariance matrix and we have also added some contextual information of the scene (for instance, zones where people can enter the scene) to improve tracking performance. In addition, a basic noise detection system and a tracking correction system are proposed in order to handle short-term tracking errors such as multiplication of IDs corresponding to only one individual. The noise detection system is designed to find and remove objects that are detected in a very small number of consecutive frames (for instance 4) and disappear afterward. The tracking correction system associates IDs recently lost with IDs that have just started to be tracked based on geometrical features and 3D distance criteria.

As a result, the tracking quality is significantly improved on 5 video sequences tested from the ETISEO dataset [3]. The people reidentification algorithm gives encouraging results for future work. The number of IDs associated to one individual is reduced (on average 50% less) and the tracking quality improves due to the IDs stability. This algorithm can detect not only people re-entering the scene but also trajectory interruptions due to occlusions or misdetections.
This approach could enable the detection of new kinds of events on video sequences such as long range people tracking on a camera network.

6.11. Crowd Data Collection from Video Recordings

Participants: Jihed Joobeur, François Brémond.

[3] http://www-sop.inria.fr/orion/ETISEO/

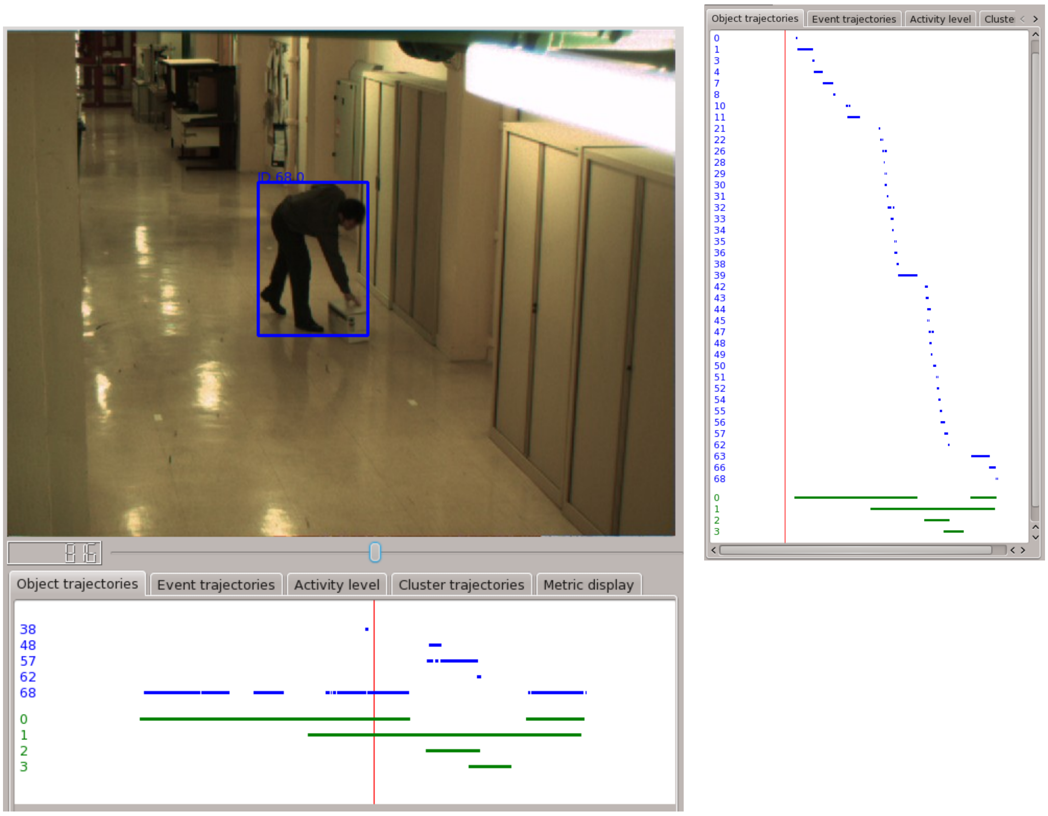


Figure 16. The green lines are the ground-truth and the blue lines are the tracking results using both reidentification and noise removal (left) and with only noise removal (right). The man is tracked throughout the whole video with a single ID (68).

The aim of this work is to analyze crowd behaviors by detecting specific situations : panic, congestion, fighting etc. We validate our work with subway station videos from VANAHEIM project. We use Mixture of Gaussian based segmentation to extract moving point and then detecting moving objects. Subsequently inside these moving objects we detect FAST feature points and compute HOG descriptors for tracking these points. We compute different features based on these points like speed and orientation. To estimate the crowd density we use features based on Grey-Level Co-occurrence Matrix. As these features depend on the distance of people from the camera, we divide the scene into different zones which have each zone same distance from the camera. In each area, compiling all the information on speed, direction and learned over a threshold density of the crowd, we can learn and detect different situations. For example, if the density increases and the average speed decreases in a pre-defined zone, that may correspond to a congestion situation.

On figure 17 the FAST feature points are shown in blue points, while the tracking of these points is shown in yellow.

Figure 17. Feature points detection and tracking in different zones.

6.12. Events Recognition and Performance Evaluation

Participants: Ricardo Cezar Bonfim Rodrigues, François Brémond.

The goal of this work is to evaluate the accuracy and performance of events detection, see workflow in Figure 18.The experiments will be performed using the tools developed in Pulsar team, such as Scene Understand Platform (SUP) [4] a plugin for events detection [50] and ViseVal [5].

The experiments were performed using video sequences of a subway station (VANAHEIN dataset) where the goal was to detect events such as people waiting, entering, buying tickets and so on. Preliminary results showed a very low accuracy and demonstrated that the scenario configuration parameters are very sensitive in this problem. It means many of the expected events were missed or misclassified, specially composite events (when more than one activity recognition is required) see some issues on Figure 19.

Based on the issues, a second experiment was configured using 3 different video sequences. In this new experiment, the scenario was adjusted to give more tolerances to people detection issues, the camera calibration was refined and some events were remodeled. After these changes the results were significantly improved. This last experiments showed that the engine proposed by Pulsar team is able to detect events accurately however events modeling can be very sensitive to the scenario configuration, see the results in Table 3.

[4] http://raweb.inria.fr/rapportsactivite/RA2010/pulsar/uid27.html
[5] http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl_Description.html

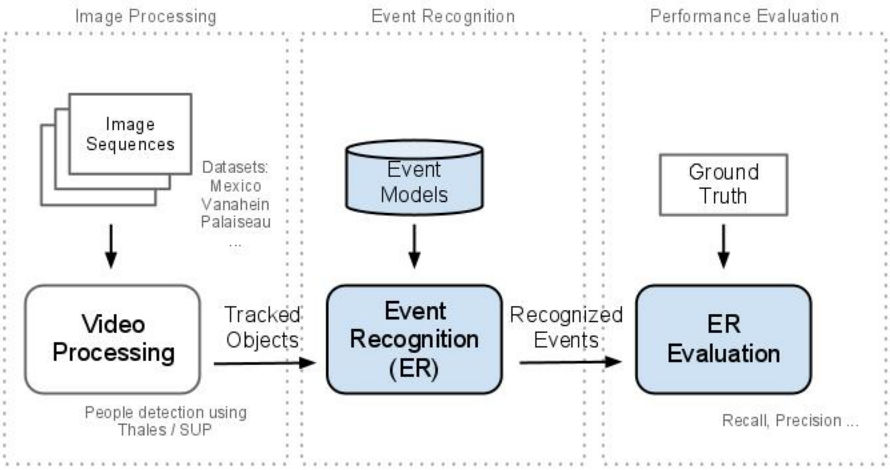


Figure 18. Events detection workflow

Figure 19. Blue bounding boxes correspond to Annotations (Reference data) and red detected objects. Its possible to observe if there is a detection issue on the left image.

Table 3. Results of events detection based on the people detection performed in experiment 2 using an IDIAP algorithm . The global Precision and sensibility of people detection are respectively 0.91 and 0.95.

	Sequence 1	Sequence 2	Sequence 2
Precision (global)	0.73	1.00	0.88
Sensitivity (global)	0.82	0.90	0.85

6.13. Group interaction and group tracking for video-surveillance in underground railway stations

Participants: Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc-Phu Chau, Etienne Corvée, François Brémond.

One goal in the European project VANAHEIM is the tracking of groups of people. Based on frame to frame mobile object tracking, 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 people being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less erratic: 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. We use the event-description language described in [50] to define events, described using basic group properties such as size, type of trajectory or number and density of people and perform the recognition of events and behaviors such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events). Two approaches to this problem have been implemented. The first approach takes as input the frame-to-frame tracking results of individual mobiles and tries to gather them into groups based on their trajectories through the temporal window. 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. The update of a group consists in re-calculating the group coherence with new mobiles from the current frame. If adding the mobile does not put the coherence under a defined threshold, the mobiles are added to the group. A pre-selection is 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 if possible.

A first improvement has been done by integrating the use of the LBP-based people detector described in [36]. This makes the algorithm more robust to false detections such as train doors closing. But on the other hand, it also introduces false negatives as, among other things, people are only detected if fully visible in the image. The group tracking algorithm has been tested both with the original, background subtraction-based mobile object detection (noted S hereafter) and the LBP-based people detection (noted LBP hereafter).

For evaluating the detection, we used 3 annotated sequences: Sequence 1 is a short sequence of 128 frames with just one ground truth object (one group), Sequence 2 has 1373 frames and 9 ground truth objects, and Sequence 3 is 17992 frames long and 25 ground truth objects were annotated. Detection and tracking results are shown in table4.

The whole algorithm chain has been integrated into the common VANAHEIM platform and sent to partners for pre-integration.

We also used videos from the ViCoMo project, recorded in the Eindhoven airport to test our approach. No formal evaluation has been done yet on these sequences due to the lack of ground truth. Nevertheless, these videos contain several acted scenes which could be successfully recognized: groups merging, splitting and entering a forbidden zone.

This work has been published in [50].

In parallel, a new approach is being developed, making use of a long-term tracker described in [35]. This tracker provides more robust individual trajectories to the group tracker, containing less confusions in cases where people cross each other. We apply the Mean Shift clustering algorithm on trajectories of people through a sliding time window (*e.g.* 10 frames). If the target is lost in one or several frames, we interpolate its positions. The clustering brings together mobiles having similar trajectories, which is our definition of a group. At each frame, clusters are calculated and then a matching is done to associate clusters to existing groups in the previous frame, and thus track groups. Looking backwards (within a window) on the trajectory of a mobile we might find a mobile on that trajectory that belongs to a group. If such a group is found, it is called the *probable*

Table 4. Segmentation (S) and Human Detector (HD) Results

	Sequence 1		Sequence 2		Sequence 3	
	S	HD	S	HD	S	HD
True Positives (TP)	72	67	1395	1079	5635	3679
False Positives (FP)	0	0	11	111	1213	642
False Negatives (FN)	6	11	269	585	3686	5642
Precision (global)	1	1	0.99	0.90	0.82	0.85
Sensitivity (global)	0.92	0.84	0.83	0.65	0.60	0.40
Tracking confusion	1	1	1	0.99	0.92	0.96

group of the current mobile. Each trajectory cluster is associated to the group that is the probable group of most mobiles in the cluster. Several clusters may be associated to the same group. This cluster association makes the algorithm robust to cases where one or several mobiles temporarily separate from the group. If the separation is longer than the time window, the probable group of these mobiles will be empty and a split will be detected.

Additionally, we work on improving the people detection by combining both methods: background subtraction-based and LBP-based. We compare overlapping mobiles from both methods and choose the best one based on their respective confidence values and their sizes. If a target was detected by only one of the two methods, we keep the target given that the confidence is high enough. If a mobile from the background subtraction method is big enough to cover several LBP-detected people (the LBP-based people detection output targets have the size of a human, whereas the background subtraction can detect a bigger mobile with the size of a GROUP_OF_PEOPLE), we attach the LBP-people as sub-mobiles of the group mobile so no information is lost. This method is a work in progress and no evaluation have been done yet.

Figure 20 shows two examples of group and event detection.

6.14. Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

The aim of this work is to learn and recognize short human actions in videos. We perform an extensive evaluation of local spatio-temporal descriptors, then we propose a new action recognition approach for RGB camera videos. We also propose a new approach for RGB-D cameras. For all our experiments, we develop an evaluation framework based on the bag-of-words model, SVM and cross-validation technique. We use the bag-of-words model to represent actions in videos and we use non-linear multi-class Support Vector Machines together with leave-one-person-out cross-validation technique to perform action classification.

Local spatio-temporal descriptors have shown to obtain very good performance for action recognition in videos. Over the last years, many different descriptors have been proposed. They are usually evaluated using too specific experimental methods and using different datasets. Moreover, existing evaluations make assumptions that do not allow to fully compare descriptors. In order to explore capabilities of descriptors, we perform an extensive evaluation of local spatio-temporal descriptors for action recognition in videos. Four widely used state-of-the-art descriptors (HOG, HOF, HOG-HOF and HOG3D) and four video datasets (Weizmann, KTH, ADL and KECK) have been selected. In contrast to other evaluations, we test all the computed descriptors, we perform experiments on several codebook sizes and use several datasets, differing in difficulty. Our results show how the recognition rate depends on the codebook size and the dataset. We

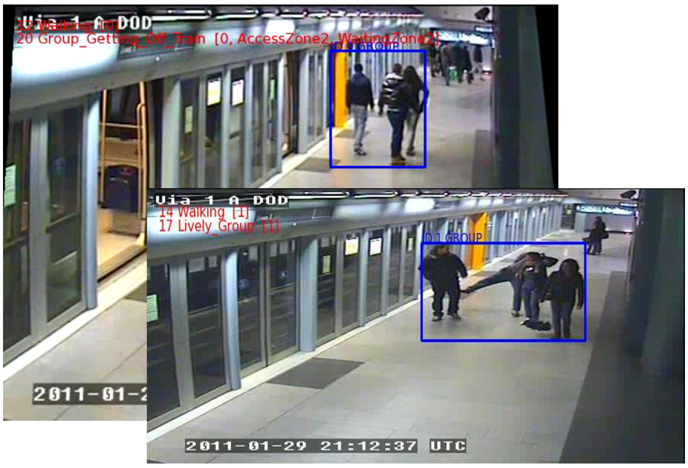


Figure 20. Example of detected groups and events: a group getting off the train (top) and a group having a lively behavior (bottom).

observe that usually the HOG descriptor alone performs the worst but outperforms other descriptors when it is combined with the HOF descriptor. Also, we observe that smaller codebook sizes lead to consistently good performance across different datasets. This work has been published in [32].

We also propose a new action recognition method for RGB camera videos based on feature point tracking and a new head estimation algorithm. We track feature points along a video and compute appearance features (HOG-HOF) for each trajectory. Additionally, we estimate a head position for each visible human in the video, using the following chain: segmentation, person, head and face detectors. Finally, we create an action descriptor based on the combination of all these sources of information. Our approach has been evaluated on several datasets, including two benchmarking datasets: KTH and ADL, and our new action recognition dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It refers to people performing daily living activities like: standing up, sitting down, walking, reading a magazine etc.

We also study the usefulness of low-cost RGB-D camera for action recognition task. We propose a new action recognition method using both RGB and depth information. We track feature points using RGB videos and represent trajectories in a four-dimensional space using additionally depth information. Experiments have been successfully performed on our new RGB-D action recognition dataset, recorded using Microsoft's Kinect device.

6.15. Activity Recognition Applied on Health Care Application

Participants: Rim Romdhane, Veronique Joumier, François Brémond.

The aim of this work is to propose a constraint-based approach for video event recognition with probabilistic reasoning for handling uncertainty. This work was validated on health care applications.

6.15.1. Event Recognition

We propose an activity recognition framework which is able to recognize composite events with complex temporal relationships. We consider different aspects of the uncertainty of the recognition during the event modeling and the event recognition process to overcome the noise or missing observations which characterize real world applications.

To reach this goal, we manage the uncertainty in the event modeling and event recognition processes by a combination of logical and probabilistic reasoning for handling uncertainty. We improve the event description language developed in Pulsar team and introduce a new probabilistic description based approach to gain in flexibility for event modeling by adding the notion of utility. Utility expresses the importance of sub-events to the recognition of the whole event. We compute the probability of recognition for both primitive (i.e. elementary) events and composite events based on Bayesian theory.

We compute the probability that the event e is recognized given a sequence of observations O as described in [48]. The observations consist of the set of the physical objects po_e moving in the scene. If the probability of an event is over a predefined threshold, the event is recognized.

6.15.2. Health Care Application

The proposed event recognition approach is validated using the videos from the health care application SWEETHOME (http://cmrr-nice.fr/sweethome/) and CIUSante (https://extranet.chu-nice.fr/ciu-sante). We have worked in close collaboration with clinicians from Nice hospital to evaluate the behaviours of Alzheimer patients. We have first model 69 event models for health care application using our event modeling formalism. 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). Experiments have been performed in a room of Nice hospital equipped with 2 video cameras where 45 elderly volunteers have spent between 15 min to 1 hour.Volunteers include Alzheimer patients, MCI (mild cognitive impairment) and healthy elderly.

The study described in [38] and [27] shows the ability of the proposed automatic video activity recognition system to detect activity changes between elderly subjects with and without dementia during a clinical experimentation. A total of 28 volunteers (11 healthy elderly subjects, 17 Alzheimer's disease patients (AD)) participate to the experimentation. The proposed study shows that we could differentiate the two profiles of participants based on motor activity parameters, such as the duration of the recognized activities, the strike length and the walking speed, computed from the proposed automatic video activity recognition system. These primary results are promising and validating the interest of automatic analysis of video as an objective evaluation tool providing comparative results between participants and over the time.

6.16. A Cognitive Vision System for Nuclear Fusion Device Monitoring

Participants: Guillaume Charpiat, Vincent Martin, François Brémond, Monique Thonnat.

In collaboration with Victor Moncada, Jean-Marcel Travere and Thierry Loarer (CEA Cadarache), we propose a cognitive vision-based system for the intelligent monitoring of tokamaks during plasma operation, based on multi-sensor data analysis and symbolic reasoning. The practical purpose is to detect and characterize in real time abnormal events such as hot spots measured through infrared images of the in-vessel components in order to take adequate decisions. Our system is made intelligent by the use of a priori knowledge of both contextual and perceptual information for ontology-driven event modeling and task-oriented event recognition. The system is made original by combining both physics-based and perceptual information during the recognition process. Real time reasoning is achieved thanks to task-level software optimizations. The framework is generic and can be easily adapted to different fusion device environments. The developed system and its achievements on real data of the Tore Supra tokamak imaging system can be found in [39].

6.17. Scenario Recognition with depth camera

Participants: Bernard Boulay, Daniel Zullo, Swaminathan Sankaranarayanan, François Brémond.

Thanks to Microsoft and its kinect sensor, RGB-depth camera becomes popular and accessible. The basic idea of depth camera is to combine a visible camera, with an IR camera associated to a laser to determine the depth of each image pixel. This kind of sensor is well adapted for applications which monitor people (e.g. monitoring Alzeihmer patient in hospital): because the people are in a predefined area and near the camera.

The depth cameras have two main advantages: first, the output image contains depth information and second, the sensor is independent from the light changes (IR sensor).

In our work, we propose to use the kinect sensor to acquire 3D images, detect the people and recognize interesting activities (see Figure 21).

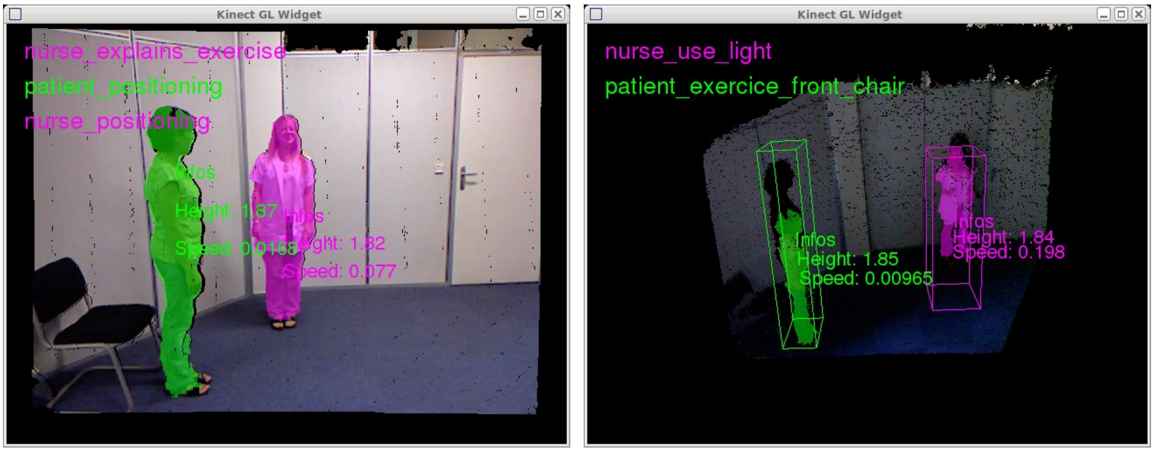


Figure 21. People detection and activity recognition in day and night conditions with kinect sensor.

The nestk library is used to manage the kinect sensor. This library is based on OpenNI framework (an open source driver) to acquire the images. Moreover the library is able to compute some treatments (e.g. people detection) and to provide a true 3D map of the scene in the referential of the kinect.

Basic attributes are computed for each detected person: 3D position, 3D height,... These attributes are then used to compute more complex information: speed, global posture to recognize interesting activities thanks to the ScreKs framework (scenario recognition based on expert knowledge) of the SUP platform as walking, stopping, standing, sitting,... (following a protocol delivered by doctors). Then we have a plug and play system able to recognize basic activities associated to a person.

Moreover, if information on the scene, as interesting zones, or equipment are available, complex activities can be recognized as nurse explaining exercise or nurse switching off the light.

The next step, is to use the human skeleton detection to recognize precisely the posture of the patient in order to understand more precise activities and infer a behaviour model.

6.18. Trajectory Clustering for Activity Learning

Participants: Jose Luis Patino, Guido Pusiol, Hervé Falciani, Nedra Nefzi, 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 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 analize the scene occupancy and its topology and also to identify activities [67], [68], [70], [55]. 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. Three 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]); 3)Monitoring activities in subway/street surveillance systems.

6.18.1. Monitoring of elderly people at home

We propose a novel framework to understand daily activities in home-care applications; the framework is capable of discovering, modeling and recognizing long-term activities (e.g. "Cooking", "Eating") occurring in unstructured scenes (i.e. "an apartment").

The framework links visual information (i.e., tracked objects) to the discovery and recognition of activities by constructing an intermediate layer of primitive events automatically.

The primitive events characterize the global spatial movements of a person in the scene ("in the kitchen"), and also the local movements of the person body parts ("opening the oven"). The primitive events are built from interesting regions, which are learned at multiple semantic resolutions (e.g. the "oven" is inside the "kitchen"). An example of the regions and possible activities for a single resolution is displayed in Fig. 22.

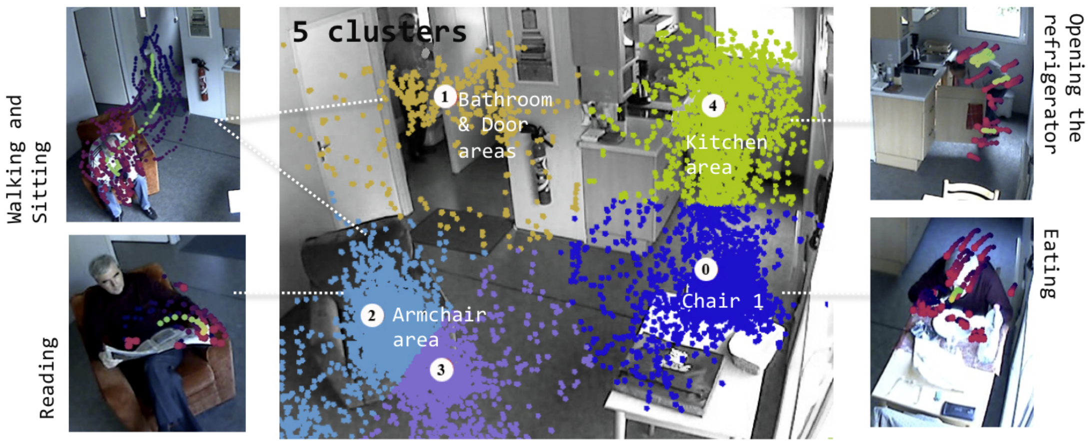


Figure 22. Example of some learned regions and possible activities.

A probabilistic model is learned to characterize each discovered activity. The modeled activities are automatically recognized in new unseen videos where a pop-up with a semantic description appears when an activity is detected. Examples of semantic labels are illustrated in Fig. 23 (a, b, c).

Recently we introduced 3D (MS. Kinect) information to the system. The preliminary results show an improvement superior to the 30% of the recognition quality. Also, the system can recognize activities in challenging situations as the lack of light. -See Fig.2 (b) and Fig.2 (c) -.

The approach can be used to recognize most of the interesting activities in a home-care application and has been published in [43]. Other examples and applications are available online in http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php.

6.18.2. 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 at an airport dock-station. Our contribution is a novel approach for discovering, in a unsupervised manner, the significant activities from observed videos. Spatial and temporal properties from detected mobile objects are modeled employing soft computing relations, that is, spatio-temporal relations graded with

[6] http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php

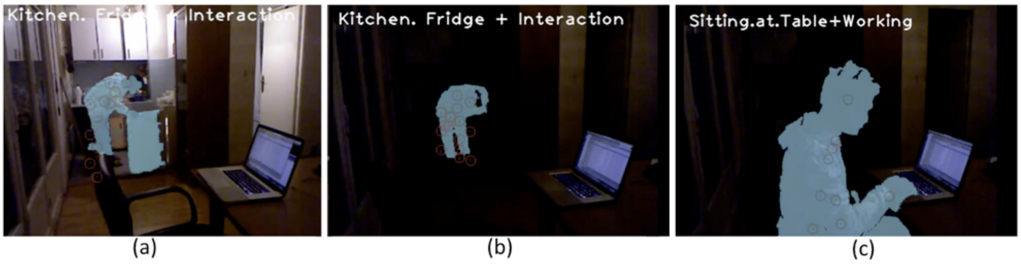


Figure 23. Examples of the recognized activities in 3D and under different light conditions.

different strengths. Our system works off-line and is composed of three modules: The trajectory speed analysis module, The trajectory clustering module, and the activity analysis module. The first module is aimed at segmenting the trajectory into segments of fairly similar speed (tracklets). The second aims at obtaining behavioural displacement patterns indicating the origin and destination of mobile objects observed in the scene. We achieve this by clustering the mobile tracklets and also by discovering the topology of the scene. The latter module aims 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. 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 below. This approach has been applied to a database containing near to 25 hours of recording of dock-station monitoring at the Toulouse airport. The discovered activities are: 'GPU positioning', 'Handler deposits chocks', 'Frontal unloading operation', 'Frontal loading operation', 'Rear loading operation', 'Push back vehicle positioning'. An example of discovered activity (Frontal loading) is given on the figure 24. When comparing our results with explicit ground-truth given by a domain expert, we were able to identify the events in general with a temporal overlap of at least 50%. The comparison with a supervised method on the same data indicates that our approach is able to extract the interesting activities signalled in the ground-truth with a higher True Positive Rate (74% TPR for the supervised approach against 80% TPR with our unsupervised method). This work has been published in [42]

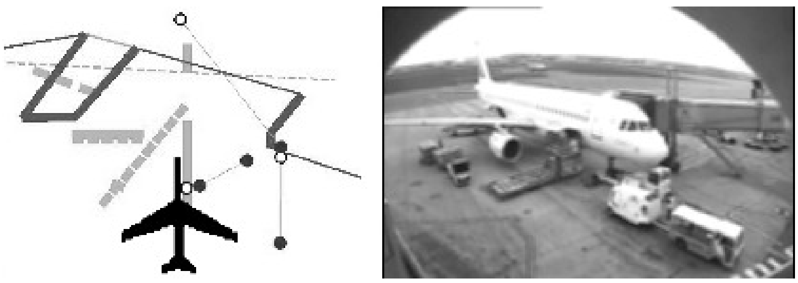


Figure 24. Example of an activity cluster obtained. The left panel presents the tracklets of the mobiles participating in the Frontal Loading activity. Filled circles indicate the beginning of a tracklet. Empty circles indicate the end of a tracklet. The right panel presents the start frame of the activity.

6.18.3. Monitoring activities in subway/street surveillance systems

In this work we have built a system to extract from video and in an unsupervised manner the main activities that can be observed from the a subway scene. We have setup a processing chain broadly working on three steps: The system starts in a first step by the unsupervised learning of the main activity areas of the scene. In a second step, mobile objects are then characterized in relation to the learned activity areas: either as ‘staying in a given activity zone’ or ‘transferring from an activity zone to another’ or a sequence of the previous two behaviours if the tracking persists long enough. In a third step we employ a high-level relational clustering algorithm to group mobiles according to their behaviours and discover other characteristics from mobile objects which are strongly correlated. We have applied this algorithm to two domains. First, monitoring two hours of activities in the hall entrance of an underground station and showed what are the most active areas of the scene and how rare/abnormal (going to low occupied activity zones) and frequent activities (e.g. buying tickets) are characterized. In the second application, monitoring one hour of a bus street lane, we were able again to learn the topology of the scene and separate normal from abnormal activities. When comparing with the available ground-truth for this application, we obtained a high recall measure (0.93) with an acceptable precision (0.65). This precision value is mostly due to the different levels of abstraction between the discovered activities and the gound-truth. The incremental learning procedure employed in this work is published in [52] while the full activity extraction approach was published in [41]

6.19. SUP Software Platform

Participants: Julien Gueytat, Leonardo Rocha, 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.

During 2011 several tasks have been accomplished:

- A stable packaged release is available
- 3D simulation from a scenario description
- Existing algorithms have been improved in performance and accuracy
- Kinect sensor has been added to the hardware supported

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.
- 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.
7. 3D simulation from description, it generates a virtual 3D animation from information provided from a text file with the description of the scenario.

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.

6.20. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean Paul Rigault, Sagar Sen, François Brémond.

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on the runtime adaptation of such systems.

Video-surveillance systems are complex and exhibit high degrees of variability along several dimensions. At the specification level, the number of possible applications and type of scenarios is large. On the software architecture side, the number of components, their variations due to possible choices among different algorithms, the number of tunable parameters... make the processing chain configuration rather challenging. Moreover, the context of an application may change in real time, requiring dynamic reconfiguration of the chain. This huge variability raises problems at design time (finding the configurations needed by the chain, foreseeing the different possible contexts), at deployment time (selecting the initial configuration), and at run time (switching configurations to react to context changes).

The first step was to formalize in an unified way all the necessary concerns —at the specification as well as at the component level— and their relations. To this end, we rely on *Feature Models* and (semi) automatic model transformations [7]. Feature Models are widely used to represent systems with many possible variation points. Moreover they are liable to formal analysis (using propositional logic and satisfiability techniques) and thus lead to valid configurations, by construction. We have developed two feature models, one for the specification of the application (type of application, context of execution, expected quality of service,etc.) and one for the implementation representation (components and their assembly). Each model has its own internal constraints. Moreover, the two models are not independent: they are connected by cross model transformation rules that formalize the bridge between application requirements and component assemblies that realize them [40].

Second, we propose a framework to derive valid possible system configurations and to adapt running configurations to context changes. Users can select features describing their application in the specification model, through a simple graphic interface. The outcome is a sub-model of the specification model. Based on the cross model transformation rules, our framework automatically transforms this sub-model into a sub-model of the component model. The latter represents all possible component configurations of the target video-surveillance system that satisfy the specifications. Both sub-models will be kept throughout the system life. They are used while the system is running to adjust its configuration in response to context changes.

To achieve this dynamic adaptation, our framework sets up three collaborating modules as shown in Figure 25:

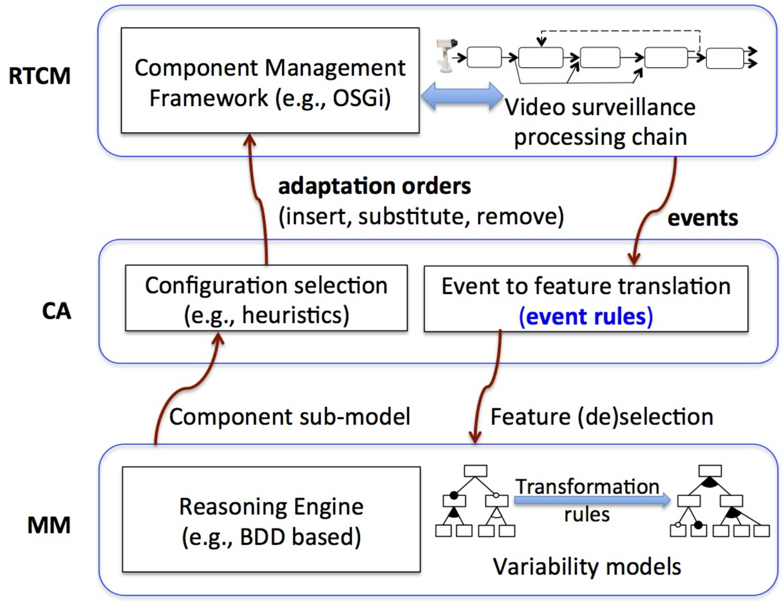


Figure 25. Run Time Adaptation Architecture

- the **Run Time Component Manager** (RTCM) captures low level events manifesting context changes (e.g., lighting changes); it forwards them to the Configuration Adapter which returns a new component configuration; the RTCM is then responsible for applying this configuration, that is

[7] Here the "features" correspond to selectable concepts of the systems; they can be at any abstraction level (a feature may correspond to an specification entity such as "Intrusion detection" or to a more concrete element such as "High frame rate"). The features are organized along a tree, with logical selection relations (optional, mandatory features, exclusive choices...) and some constraints that restrict the valid combinations of features (i.e., *configurations*)

to tune, add, remove, or replace components, and possibly to change the workflow itself.

- the **Configuration Adapter** (CA) receives change events from the RTCM, translates them into the feature formalism, and forwards the result to the Model Manager; in return, it obtains a sub-model of component configurations compatible with the change; this sub-model is a compact representation of a set of valid configurations and the CA is responsible to select one and to instruct the RTCM to apply it; this selection uses some heuristics, possibly based on a cost function such as minimizing the number of component changes in the processing chain or maximizing the quality of service (e.g., accuracy, responsiveness).
- the **Model Manager** (MM) manages the representation of the two specialized Feature Models corresponding to the specification and possible component assemblies of the current application together with their constraints; its role is to enforce configuration validity. It is also responsible of the set of rules relating run time events and (de)selection of features in both models. From the CA, the Model Manager receives information about incoming events; it uses the rules to select or deselect the corresponding features; it then applies constraints, rules, and model transformations to infer a component sub-model that represents a subset of valid component configurations and that it returns to the CA.

This year, we have tested our approach on simple applications using well-known libraries (OpenCV) on different scenarios. At the moment, 77 features and 10^8 configurations are present in the specification model while 51 features and 10^6 configurations are present in the component model. Once the video surveillance designer has selected the features required by an application, before deployment, the average number of features to consider at runtime in the component model is less than 10^4. The configuration spaces is reduced by several orders of magnitude and enables the use of the other tools in the end-to-end engineering process, whereas it would not have been possible without. Our experiments show the feasibility of such an approach with a limited performance overhead (if any) compared to traditional run time control where *ad hoc* adaptation code is hardwired and does not rely on the run time availability of an abstract representation of the application and its context evolution [29]. The next step will be to test our approach on our SUP platform and to study intelligent configuration selection heuristics.

6.21. The Girgit Software

Participants: Leonardo Rocha, Sabine Moisan, Jean-Paul Rigault, Sagar Sen.

Girgit is a Python based framework to build context-aware self-adaptive software systems.

Girgit is a simple and small [1] framework that allows dynamic reconfiguration of data processing chains and accepts any set of components for the configuration as long as they have the corresponding Python wrapper.

The basic Idea of Girgit is to provide a platform to be able to make dynamic adaptive systems. It provides a dynamic adaptive engine that can deal with event/action pairs called rules and also provides an application programming interface to be able to use it.

As the system is designed to be interactive, any user can interact with it, be a human operator or a reasoning engine. Rules can be pre-loaded with the configuration, and components can launch events, this allows pre-configured rules to be loaded at launch time of the Girgit.

Girgit has been used to evaluate the performance on real time video applications and show the architecture. Three publications [47], [45], [46] where based on the framework.

6.22. Scenario Analysis Module

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

To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or others sensors. In this framework, this year we focus on recognition algorithm improvement in order to face the problem of large number of scenario instances recognition.

The purpose of this research axis 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 [69] and we apply general modeling methods [62] 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” [63] expressed in the scenario language.

Last 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. This year, to face this challenge we first replace some operators of the language, by others having a more strict semantics. For instance, we replace the *before* operator whose semantics allowed that events can meet, by two operators , a strict *before* and a *meet*. Hence, the number of events a scenario instance reacts to decreases. Second, we now generate within the recognition engine, the expected events of the next step. This avoids to run the engine automatically with events that are not relevant for the recognition process.

Presently, we still rely on the existing synchronous language (Lustre [62]) 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 [57].

Now the challenge is to take into account some uncertainty on the primitive events due to input sensor errors. In the family of synchronous languages, the *Lutin* language [8] could be able to automate the generation of realistic input sequences of events, taking into account probabilistic distributions over primitive events. In other words, it could generate a set of input events for which a set a constraints can be verified. In complement, it offers also means to compute the real values verifying these constraints. Thus, we think to rely on Lutin to express uncertainty on primitive events and get input events to feed scenario recognition engines.

6.23. Multiple Services for Device Adaptive Platform for Scenario Recognition

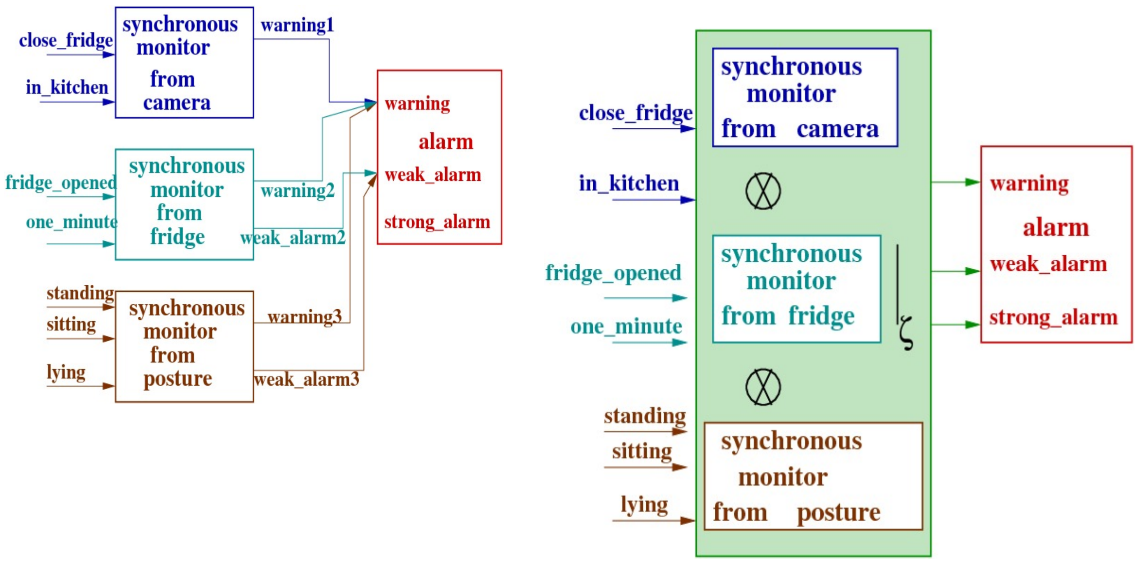
Participants: Annie Ressouche, Jean-Yves Tigli.

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.19) 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 [76], [75], [64]) 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.

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 concerning this critical component. Thus, we consider that the critical component is validated.

[8] http://www-verimag.imag.fr/Lutin.html

When a critical component has multiple synchronous monitors corresponding to several concern managements in the application, we want to build an only synchronous model component which agrees with all these primitive synchronous monitors To specify how output events sent by different synchronous monitors and connected to a critical component, we introduce a sound (with respect to our mathematical formalism) operation of *composition under constraints* of synchronous models (see figure 26). 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 [62]. Then the validation of properties and the creation of the validated synchronous component is automatic [44], [53].



(a) SM (b) CSM

Figure 26. Example of multiple access to alarm critical component: (a) SM - the synchronous monitors; (b): CSM the composition under constraints operation

This year we focus on the main challenge of this approach which is to deal with the possibly very large number of constraints a user must specify. Indeed, each synchronous monitor has to tell how it combines with others, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To be adaptive with efficiency, we must face this problem. We first introduced some default rules to avoid the user to express a large number of constraints. We also studied how Abstract Interpretation technique can help us to reduce this complexity. This approach works if we forbid some “non monotonic” constraints, but this is a strong limitation. Thus, it is still a challenge for us. This drawback is a popular challenge in adaptive middleware and some results exist relying on controller synthesis methods. We are not in the exact framework where these techniques apply, but we plan to study if we can rely on some extension of these techniques.

On another hand, we also want to complement our preservation result in studying how the proof of a global property can be decomposed into the proof of local ones. In general, this decomposition (known as assume-guarantee paradigm) is difficult to apply but there is no communication between synchronous monitors and so the decomposition could be tractable. Moreover, some works address this problem and we can rely on them.

6.24. 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 (see figure 6) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. Indeed, only few approaches consider a modular compilation because there is a deep incompatibility between *causality* and modularity. 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. We defined three different approaches 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 monotonic increasing function.

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).

To be modular, we defined 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 adjustment;
2. fix point characterization starting with the vectors of already computed dates and considering only the variables in the dependencies (upstream and downstream) of common variables

This year we began the implementation of a separated compilation of LE programs, according to these theoretical results. We define a new intermediate format (lea) to record partially compiled module, i.e module whose Boolean equation systems may be composed of non defined variables (we called them abstract). Then we are implementing a refinement operation which replaces these abstract variables by their definition and performs adjustment of the dates. According to our theoretical results, we know that the resulting sorting is the same as with a global approach. After the termination of this separated compilation of LE programs, the challenge will be to use Clem to design a large application in the domain of smart cards. The application needs more than forty LE automata in parallel and the compiled code will have more than 500 registers and thousands variables. Only a separated compilation will work.

The Clem toolkit is completely described in [28]

7. Contracts and Grants with Industry

7.1. Contracts with Industry

The Pulsar team has strong collaborations with industrial partners through European projects and national grants. In addition we have also:

- a contract with Thales ThereSIS to support two studies on video event recognition and on event discovery;
- a contract with Link Care Services to study fall detection for older people with dementia.

7.2. Grants with Industry

contract with Keeneo (bought by Digital Barrier) for the PhD fellowship CIFRE (Conventions Industrielles de Formation par la Recherche) of Malik Souded on people tracking though a camera network.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. National Projects

8.1.1.1. CIU-Santé

Program: DGCIS-FUI

Project acronym: CIU-Santé

Project title: Centre Innovation et Usage pour la Santé

Duration: November 2008- May 2012

Coordinator: Pôle de compétitivité SCS

Partners: CSTB, Sophia Antipolis (FR); Nice City Hospital, Nice (FR); Actis Ingenierie (FR), Movea (FR), CEA (FR); UNSA (FR).

See also: http://www.ciusante.org/

Abstract: CIU-Santé (Centre d'Innovation et d'Usages en Santé) is a DGCIS project to develop experiments in the domain of health. Two experimental rooms have been equipped in Nice-Cimiez Hospital for monitoring Alzheimer patients.

8.1.1.2. VIDEO-ID

Program: ANR Sécurité

Project acronym: VIDEO-ID

Project title: VideoSurveillance and Biometrics

Duration: February 2008-February 2012

Coordinator: Thales Security Systems and Solutions S.A.S

Other partners: INRIA; EURECOM; TELECOM and Management Sud Paris; CREDOF ; RATP

See also: http://www-sop.inria.fr/pulsar/projects/videoid/

Abstract: Using video surveillance, the VIDEO-ID project aims at achieving real time human activity detection including the prediction of suspect or abnormal activities. This project also aims at performing identification using face and iris recognition. Thanks to such identification, a detected person will be tracked throughout a network of distant cameras, allowing to draw a person's route and his destination. Without being systematic, a logic set of identification procedures is established: event and abnormal behaviour situation and people face recognition.

8.1.1.3. SWEET-HOME

Program: ANRTecsan

Project acronym: SWEET-HOME

Project title: Monitoring Alzheimer Patients at Nice Hospital

Duration: November 2009-November 2012

Coordinator: CHU Nice Hospiteal (FR)

Other partners: INRIA (FR); LCS (FR); CNRS unit - UMI 2954, MICA Center in Hanoi (VN); SMILE Lab , National Cheng Kung University (TW); National Cheng Kung University Hospital (TW).

Abstract: 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.

8.1.1.4. QUASPER

Program: FUI

Project acronym: QUASPER

Project title: QUAlification et certification des Systèmes de PERception

Duration: June 2010 - May 2012

Coordinator: THALES There SIS

Other partners: AFNOR; AKKA; DURAN; INRETS; Sagem Securité; ST Microelectronics; Thales RT; Valeo Vision SAS; CEA; CITILOG; Institut d'Optique; CIVITEC; SOPEMEA; ERTE; HGH.

See also: http://www.systematic-paris-region.org/fr/projets/quasper-rd

Abstract: QUASPER project 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.

8.1.1.5. moniTORE

Program: FR-FC (Fédération nationale de Recherche Fusion par Confinement Magnétique - ITER)

Project acronym: moniTORE

Project title: Real Time Monitoring of Imaging Diagnostic Applied to Tore Plasma Operation.

Duration: 1 year

Coordinator: Imaging and Diagnostics Group of the CEA Cadarache

Other partners: INRIA Pulsar team

Abstract: moniTORE is 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.

Le projet monitore est une action soutenue et financee par la (FR-FCM Fédération nationale de Recherche Fusion par Confinement Magnétique - ITER) dont Pulsar est un laboratoire membre. pour la description : monitore = real time monitoring of imaging diagnostic applied to tore plasma operation.

8.1.2. Large Scale INRIA Initiative

8.1.2.1. PAL

Program: INRIA

Project acronym: PAL

Project title: Personally Assisted Living

Duration: 2010 -2014

Coordinator: COPRIN team

Others partners: AROBAS, DEMAR, E-MOTION, PULSAR, PRIMA, MAIA, TRIO, and LAGADIC INRIA teams

See also: http://www-sop.inria.fr/coprin/aen/

Abstract: The objective of this project is to create a research infrastructure that will enable experiments with technologies for improving the quality of life for persons who have suffered a loss of autonomy through age, illness or accident. In particular, the project seeks to enable development of technologies that can provide services for elderly and fragile persons, as well as their immediate family, caregivers and social groups.

8.2. European Initiatives

8.2.1. FP7 Projects

8.2.1.1. Co-FRIEND

Title: Cognitive Vision System able to adapt itself to unexpected situations

Type: COOPERATION (ICT)

Defi: Cognitive systems, interaction, robotics

Instrument: Specific Targeted Research Project (STREP)

Duration: February 2008 - January 2011

Coordinator: AKKA (FR) (France)

Others partners: Akka, Toulouse airport (FR); University of Reading and of Leeds (UK); Cognitive System Laboratory, University of Hamburg (G).

See also: http://www.co-friend.net

Abstract: The main objectives of this project are to develop techniques to recognize and learn automatically all servicing operations around aircraft parked on aprons.

8.2.1.2. VANAHEIM

Title: Autonomous Monitoring of Underground Transportation Environment

Type: COOPERATION (ICT)

Defi: Cognitive Systems and Robotics

Instrument: Integrated Project (IP)

Duration: February 2010 - July 2013

Coordinator: Multitel (Belgium)

Others partners: INRIA Sophia-Antipolis (FR); Thales Communications (FR); IDIAP (CH); Torino GTT (Italy); Régie Autonome des Transports Parisiens RATP (France); Ludwig Boltzmann Institute for Urban Ethology (Austria); Thales Communications (Italy).

See also: http://www.vanaheim-project.eu/

Abstract: 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.

8.2.1.3. SUPPORT

Title: Security UPgrade for PORTs

Type: COOPERATION (SECURITE)

Instrument: IP

Duration: July 2010 - June 2014

Coordinator: BMT Group (UK)

Others partners: INRIA Sophia-Antipolis (FR); Swedish Defence Research Agency (SE); Securitas (SE); Technical Research Centre of Finland (FI); MARLO (NO); INLECOM Systems (UK).

Abstract: 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.

8.2.1.4. Dem@Care

Title: Dementia Ambient Care: Multi-Sensing Monitoring for Intelligent Remote Management and Decision Support

Type: COOPERATION (ICT)

Defi: Cognitive Systems and Robotics

Instrument: Collaborative Project (CP)

Duration: November 2011-November 2015

Coordinator: Centre for Research and Technology Hellas (G)

Others partners: INRIA Sophia-Antipolis (FR); University of Bordeaux 1(FR); Cassidian (FR), Nice Hospital (FR), LinkCareServices (FR), Lulea Tekniska Universitet (SE); Dublin City University (IE); IBM Israel (IL); Philips (NL); Vistek ISRA Vision (TR).

Abstract: The objective of Dem@Care is the development of a complete system providing personal health services to persons with dementia, as well as medical professionals, by using a multitude of sensors, for context-aware, multiparametric monitoring of lifestyle, ambient environment, and health parameters. Multisensor data analysis, combined with intelligent decision making mechanisms, will allow an accurate representation of the person's current status and will provide the appropriate feedback, both to the person and the associated medical professionals. Multi-parametric monitoring of daily activities, lifestyle, behaviour, in combination with medical data, can provide clinicians with a comprehensive image of the person's condition and its progression, without their being physically present, allowing remote care of their condition.

8.2.2. Collaborations in European Programs, except FP7

8.2.2.1. ViCoMo

Program: ITEA 2

Project acronym: ViCoMo

Project title: Visual Context Modeling

Duration: October 2009 - October 2012

Coordinator: International Consortium (Philips, Acciona, Thales, CycloMedia, VDG Security)

Other partners: TU Eindhoven; University of Catalonia; Free University of Brussels; INRIA; CEA List;

Abstract: 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 performed.

8.3. International Initiatives

8.3.1. INRIA International Partners

8.3.1.1. 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.

8.3.1.2. Collaboration with U.S.

Pulsar collaborates with the University of Southern California.

8.3.1.3. Collaboration with Europe

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

8.3.2. Visits of International Scientists

8.3.2.1. Internships

This year Pulsar has hosted seven international internships (see section 1).

8.3.3. Participation In International Programs

8.3.3.1. EIT ICT Labs

EIT ICT Labs is one of the first three Knowledge and Innovation Communities (KICs) selected by the European Institute of Innovation & Technology (EIT) to accelerate innovation in Europe. EIT is a new independent community body set up to address Europe's innovation gap. It aims to rapidly emerge as a key driver of EU's sustainable growth and competitiveness through the stimulation of world-leading innovation. Among the partners, there are strong technical universities (U Berlin, 3TU / NIRICT, Aalto University, UPMC - Université Pierre et Marie Curie, Université Paris-Sud 11, Institut Telecom, The Royal Institute of Technology); excellent research centres (DFKI, INRIA, Novay , VTT, SICS) and leading companies (Deutsche Telekom Laboratories, SAP, Siemens, Philips, Nokia, Alcatel-Lucent, France Telecom, Ericsson). This project is largely described at http://eit.ictlabs.eu.

We are involved in the EIT ICT Labs - Health and Wellbeing .

9. Dissemination

9.1. Animation of the scientific community

9.1.1. Journals

- G. Charpiat reviewed this year for TPAMI (Transactions on Pattern Analysis and Machine Intelligence), TIP (Transactions on Image Processing), IVC (Image and Vision Computing), IJIG (International Journal of Image and Graphics)

- 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 CVIU Computer Vision and Image Understanding, PATREC and gerontechnology Journal.

9.1.2. Conferences and Workshops

- F. Brémond was reviewer for the conferences: CVPR 2011, ICCV 2011, BMVC 2011, WCC-AAISS 2011, Workshop on Human behaviour Analysis in Open or Public Spaces 2011, AVSS 2011, RIVF 2012, CVPR 2012
- F. Brémond was Conference Area Chair of ICVS 2011 and AVSS 2011
- F. Brémond was Program Committee member of the conferences: VS2011, ICDP-11, ARTEMIS 2011, IF 2012, HBU 2012
- F. Brémond organized the workshop Behaviour 2011
- G. Charpiat reviewed this year for CVPR, Eurographics, Muscle, RFIA, Eusipco.
- M. Thonnat was general chair of the 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

- F. Brémond was invited t by Rita Cucchiara to give a talk at the SafeCity@Euromed Summit on the 21st May 2011 in Genoa.
- F. Brémond was invited for summer school by Florence SEDES : Ambiance Intelligence, Lille, 4-10 July 2011.
- F. Brémond was invited by Viorel Negru as Short-Term Visiting Scientist on scene understanding, at West University Timisoara, Romania, July 27, 2011.
- F. Brémond was invited to make a demonstration by Emmanuel HIRSCH, Éthique, technologie et maladie d'Alzheimer ou apparentée EREMA, April 2011.
- F. Brémond was invited by Eric Castelli to give talks on video understanding, at International Research Center MICA, Hanoi University of Science and Technology, October 12, 2011.
- F. Brémond was invited to give a talk by Jenny Benois-Pineau, LABRI Université Bordeaux, December 5 2011.
- F. Brémond made a demonstration at the Rencontre Inria-Industries Santé à domicile et de l'autonomie in Paris on October 20th 2011
- F. Brémond was invited to give a talk at the RivieraDEV, conference on software design, at INRIA Sophia-Antipolis on October 21th 2011
- G . Charpiat was invited by Yann Ollivier to give a talk on his works in vision, learning and optimization in the INRIA TAO team.
- G. Charpiat was invited to give a talk about his works on graph-cut optimization in the "Semaine conjointe GdR MOA et GdR MSPC Optimisation et traitement des images" (La Londe les Maures, June 2011).
- G. Charpiat was invited to give a talk about his works on learning shape metrics in the INRIA workshop on Statistical Learning (IHP, Paris, December 2011), and to present a poster in the Symposium on Empirical Inference (Tübingen, Germany, December 2011).

9.1.4. HDR & PhD

- F. Brémond was jury member for the following PhD defenses:
  - Dora Luz Almanza, Université de Toulouse III Paul Sabatier, January 7, 2011
  - Cédric ROSE, Université de Henri Poincaré - Nancy 1, mai 27, 2011
  - Nadeem Salamat, Université de la Rochelle, October 7, 2011
  - Juan Carlos San Miguel Avedillo, Universidad Autonoma de Madrid, October 27, 2011
  - Anna-Louise Ellis, Reading University, November 7, 2011
  - Vladislavs Dovgalecs, Université Bordeaux 1, December 5, 2011
- M. Thonnat was member of the jury for the PhD defense of Carolina Venagas, Telecom ParisTech, 13 January 2011
- M. Thonnat was reviewer for the HDR of Nicolas Lomenie, Université Paris Descartes, May 2011

9.1.5. 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.

9.1.6. Avisory Boards

- F. Brémond is Scientific Board member of WISG 2011 of the French ANR
- F. Brémond is Scientific Board member of The “Espace Éthique et maladie d’Alzheimer “
- M. Thonnat is Scientific Advisory Board member 2010 -2013 European project Fish4Knowledge on Intelligent Information Management, Challenge 4: Digital Libraries and Content (http://www.fish4knowledge.eu).
- M. Thonnat is Scientific Board member of the National Reference Center Sante, Dependance et Autonomie since 2010
- M. Thonnat is Scientific Board member of Ecole Nationale des Ponts since 2008

9.1.7. Expertise

- F. Brémond is EC INFSO Expert in the framework of Ambient Assisted Living European FP7 (VitalMind projects) and for the CALL7 - Objectives 5.4 & 5.5 in 2011

9.1.8. Miscellanous

- F. Brémond was reviewer for the “dossier d’appellation de Professeur” of Sonia GARCIA, Télécom SudParis, 8 Juin 2011
- F. Brémond was member of the jury of Farah Arab, for obtaining the “prix de thèse de la fondation Médéric Alzheimer”, August 20, 2011.
- G. Charpiat presented the team to ENS Lyon students.
- G. Charpiat takes part in Mastic, a local scientific animation committee (Médiation et Animation Scientifique dans les MAthémathiques et dans les Sciences et Techniques Informatiques et des Communications).
- J-P Rigault is a member of the Administration Board of the Polytechnic Institute of Nice University.
- 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

Teaching:

Master : Video Understanding Techniques, 3h, M2, EURECOM, Sophia Antipolis (FR) (François Brémond)

Licence : Verification with Scade Suite, 6h, L3, Ecole des Mines de Paris (FR) (Annie Ressouche)

Master : Synchronous Languages and Verification, M2, Polytechnic School of Nice Sophia Antipolis University (FR) (Annie Ressouche)

PhD & HdR :

PhD in progress: Julien Badie, People tracking and video understanding, October 2011, François Brémond

PhD in progress : Slawomir Bak , People Detection in Temporal Video Sequences by Defining a Generic Visual Signature of Individuals, February 2009, François Brémond

PhD in progress : Piotr Bilinski, Gesture Recognition in Videos, March 2010, François Brémond

PhD in progress : Duc Phu Chau, Object Tracking for Activity Recognition, December 2008, François Brémond and Monique Thonnat

PhD in progress : Carolina Garate, Video Understanding for Group Behavior Analysis, August 2011, François Brémond

PhD in progress : Ratnesh Kumar, Fiber-based segmentation of videos for activity recognition, January 2011, Guillaume Charpiat and Monique Thonnat

PhD in progress : Emmanuel Mulin, Utilisation de l'Outil Vidéo dans l'Analyse des Troubles du Comportement chez les Patients Atteints de la Maladie d'Alzheimer, January 2010, François Brémond

PhD in progress : Guido-Tomas Pusiol, Learning Techniques for Video Understanding, September 2008, François Brémond

PhD in progress : Rim Romdhame, Event Recognition in Video Scenes with Uncertain Knowledge, March 2009, François Brémond and Monique Thonnat

PhD in progress : Malik Souded : Suivi d'Individu à travers un Réseau de Caméras Vidéo, February 2010, François Brémond

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. BRÉMOND, M. THONNAT. *Data Mining on Large Video Recordings*, in "5eme Colloque Veille Stratégique Scientifique et Technologique VSST 2007", Marrakech, Marrocco, 21st - 25th October 2007.

[3] B. BOULAY, F. BRÉMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System.*, in "Pattern Recognition Letter.", 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] G. CHARPIAT, P. MAUREL, J.-P. PONS, R. KERIVEN, O. FAUGERAS. *Generalized Gradients: Priors on Minimization Flows*, in "International Journal of Computer Vision", 2007.

[6] N. CHLEQ, F. BRÉMOND, M. THONNAT. *Advanced Video-based Surveillance Systems*, Kluwer A.P. , Hangham, MA, USA, November 1998, p. 108-118.

[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. BRÉMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", 2007, vol. 18, p. 189-205.

[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] 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.

[11] V. MARTIN, J.-M. TRAVERE, F. BRÉMOND, 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.

[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. BRÉMOND, M. THONNAT. *Video-Data Modelling and Discovery*, in "4th IET International Conference on Visual Information Engineering VIE 2007", London, UK, 25th - 27th July 2007.

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

[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, 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] N. ZOUBA, F. BRÉMOND, 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.

Publications of the year

Articles in International Peer-Reviewed Journal

[25] S. BAK, E. CORVEE, F. BRÉMOND, M. THONNAT. *Boosted human re-identification using Riemannian manifolds*, in "Image and Vision Computing", August 2011 [*DOI :* 10.1016/J.IMAVIS.2011.08.008], http://hal.inria.fr/hal-00645588/en/.

[26] A. GAMAL ELDIN, X. DESCOMBES, G. CHARPIAT, J. ZERUBIA. *Multiple Birth and Cut Algorithm for Multiple Object Detection*, in "Journal of Multimedia Processing and Technologies", 2012, http://hal.inria.fr/hal-00616371/en.

[27] R. ROMDHANE, E. MULIN, A. DERREUMEAUX, N. ZOUBA, J. PIANO, L. LEE, P. MALLEA, M. THONNAT, F. BRÉMOND, P. ROBERT, I. LEROI, R. DAVID. *Automatic Video Monitoring system for assessment of Alzheimer's Disease symptoms*, in "JNHA 2011", January 2011, http://hal.inria.fr/inria-00616747/en.

Articles in National Peer-Reviewed Journal

[28] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Revue des sciences et technologies de l'information, série Théorie et Science Informatique", June 2011, vol. 4, n^o 30, p. 441-471, http://hal.inria.fr/inria-00524499/en.

International Conferences with Proceedings

[29] M. Acher, P. Collet, P. Lahire, S. Moisan, J.-P. Rigault. *Modeling Variability from Requirements to Runtime*, in "IEEE CS proceedings of the 16th IEEE International Conference on Engineering of Complex Computer Systems, ICECCS 2011", Las Vegas, United States, 2011, http://hal.inria.fr/inria-00617273/en.

[30] S. Bak, E. Corvee, F. Brémond, M. Thonnat. *Multiple-shot Human Re-Identification by Mean Riemannian Covariance Grid*, in "Advanced Video and Signal-Based Surveillance", Klagenfurt, Austria, August 2011, http://hal.inria.fr/inria-00620496/en.

[31] S. Bak, K. Kurowski, K. Napierala. *Human Re-identification System On Highly Parallel GPU and CPU Architectures*, in "Multimedia Communications, Services and Security", Krakow, Poland, A. Dziech, A. Czyżewski (editors), Springer Berlin Heidelberg, June 2011, vol. 149 [*DOI :* 10.1007/978-3-642-21512-4_35], http://hal.inria.fr/hal-00645938/en/.

[32] P. Bilinski, F. Brémond. *Evaluation of local descriptors for action recognition in videos*, in "International Conference on Computer Vision Systems", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00619091/en.

[33] G. Charpiat. *Exhaustive Family of Energies Minimizable Exactly by a Graph Cut*, in "Computer Vision and Pattern Recognition", Colorado Springs, États-Unis, June 2011, http://hal.inria.fr/inria-00616370/en/.

[34] D. P. Chau, F. Brémond, M. Thonnat. *A multi-feature tracking algorithm enabling adaptation to context variations*, in "The International Conference on Imaging for Crime Detection and Prevention (ICDP)", London, Royaume-Uni, November 2011, http://hal.inria.fr/inria-00632245/en/.

[35] D. P. Chau, F. Brémond, M. Thonnat, E. Corvee. *Robust Mobile Object Tracking Based on Multiple Feature Similarity and Trajectory Filtering*, in "The International Conference on Computer Vision Theory and Applications (VISAPP)", Algarve, Portugal, March 2011, http://hal.inria.fr/inria-00599734/en.

[36] E. Corvee, F. Brémond. *Haar like and LBP based features for face, head and people detection in video sequences*, in "International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00624360/en.

[37] A. Gamal Eldin, X. Descombes, G. Charpiat, J. Zerubia. *A Fast Multiple Birth and Cut Algorithm using Belief Propagation*, in "ICIP 2011 (International Conference on Image Processing)", Bruxelles, Belgium, September 2011, http://hal.inria.fr/inria-00592446/en.

[38] V. Joumier, R. Romdhane, F. Brémond, M. Thonnat, E. Mulin, P. H. Robert, A. Derreumaux, J. Piano, J. Lee. *Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients*, in "International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00624369/en.

[39] V. Martin, V. Moncada, J.-M. Travere, T. Loarer, F. Brémond, G. Charpiat, M. Thonnat. *A Cognitive Vision System for Nuclear Fusion Device Monitoring*, in "8th International Conference, ICVS 2011", Sophia Antipolis, France, Crowley, J. L. Draper, B. Thonnat, Monique (editors), Lecture Notes in Computer Science, Springer, September 2011, vol. 6962, p. 163-172, http://hal.inria.fr/hal-00609876/en.

[40] S. MOISAN, J.-P. RIGAULT, M. ACHER, P. COLLET, P. LAHIRE. *Run Time Adaptation of Video-Surveillance Systems: A software Modeling Approach*, in "ICVS, 8th International Conference on Computer Vision Systems", Sophia Antipolis, France, France, 2011, http://hal.inria.fr/inria-00617279/en.

[41] L. PATINO, H. BENHADDA, N. NEFZI, B. BOULAY, F. BRÉMOND, M. THONNAT. *Abnormal behavior detection in video protection systems*, in "International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00624363/en.

[42] L. PATINO, M. EVANS, J. FERRYMAN, F. BRÉMOND, M. THONNAT. *Unsupervised Activity Extraction on Long-Term Video Recordings employing Soft Computing Relations*, in "8th International Conference on Computer Vision Systems, ICVS 2011", Sophia Antipolis, France, September 2011, http://hal.inria.fr/hal-00650048/en/.

[43] G. PUSIOL, F. BRÉMOND, M. THONNAT. *Unsupervised Discovery, Modeling and Analysis of long term Activities*, in "international Conference in Vision Systems", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00621201/en.

[44] A. RESSOUCHE, J.-Y. TIGLI, O. CARRILLO. *Toward Validated Composition in Component-Based Adaptive Middleware*, in "SC 2011", Zurich, Suisse, E. J. SVEN APEL (editor), LNCS, Springer, July 2011, vol. 6708, p. 165-180, http://hal.inria.fr/inria-00605915/en/.

[45] L. M. ROCHA, S. MOISAN, J.-P. RIGAULT, S. SAGAR. *Girgit: A Dynamically Adaptive Vision System for Scene Understanding*, in "ICVS", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00616642/en/.

[46] L. M. ROCHA, S. MOISAN, J.-P. RIGAULT, S. SAGAR. *Towards Lightweight Dynamic Adaptation A Framework and its Evaluation*, in "JAIIO - ASSE", Córdoba, Argentina, September 2011, http://hal.inria.fr/inria-00616644/en.

[47] L. M. ROCHA, S. SAGAR, S. MOISAN, J.-P. RIGAULT. *Girgit: A Framework for Component-based Dynamic Adaptation in Python*, in "EuroSciPy 2011", Paris, France, August 2011, http://hal.inria.fr/hal-00650026/en/.

[48] R. ROMDHANE, F. BRÉMOND, B. BOULAY, M. THONNAT. *Probabilistic Recognition of Complex Event*, in "ICVS 2011", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00616742/en.

[49] M. SOUDED, F. BRÉMOND. *An Object Tracking in Particle Filtering and Data Association Framework, Using SIFT Features*, in "International Conference on Imaging for Crime Detection and Prevention (ICDP)", London, United Kingdom, November 2011, http://hal.inria.fr/hal-00647256/en/.

[50] S. ZAIDENBERG, B. BOULAY, C. GARATE, D. P. CHAU, E. CORVEE, F. BRÉMOND. *Group interaction and group tracking for video-surveillance in underground railway stations*, in "International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00624356/en.

Scientific Books (or Scientific Book chapters)

[51] G. CHARPIAT, I. BEZRUKOV, Y. ALTUN, M. HOFMANN, B. SCHÖLKOPF. *Machine Learning Methods for Automatic Image Colorization*, in "Computational Photography: Methods and Applications", R. L. EDITOR (editor), CRC Press, January 2011, http://hal.inria.fr/inria-00481561/en.

[52] L. PATINO, F. BRÉMOND, M. THONNAT. *Incremental learning on trajectory clustering*, in "Intelligent Paradigms in Safety and Security", D. P. REMAGNINO (editor), Springer-Verlag, 2011, http://hal.inria.fr/inria-00503059/en/.

Research Reports

[53] A. RESSOUCHE, J.-Y. TIGLI, O. CARILLO. *Composition and Formal Validation in Reactive Adaptive Middleware*, INRIA, February 2011, n° RR-7541, http://hal.inria.fr/inria-00565860/en.

References in notes

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

[55] 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.

[56] D. P. CHAU, F. BRÉMOND, M. THONNAT. *Online evaluation of tracking algorithm performance*, in "The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP)", London, Royaume-Uni, December 2009, p. SBN: 978-1-84919-207-1, P38, http://hal.inria.fr/inria-00486479/en/.

[57] 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° 2404, p. 359-364, http://nusmv.fbk.eu/NuSMV/papers/cav02/ps/cav02.ps.

[58] 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° 2, p. 143-169.

[59] N. DALAL, B. TRIGGS. *Histograms of Oriented Gradients for Human Detection*, in "International Conference on Computer Vision & Pattern Recognition (CVPR '05)", San Diego, États-Unis, C. SCHMID, S. SOATTO, C. TOMASI (editors), IEEE Computer Society, 2005, vol. 1, p. 886–893 [*DOI :* 10.1109/CVPR.2005.177], http://hal.inria.fr/inria-00548512/en/.

[60] ECVISION. *A Research Roadmap of Cognitive Vision*, IST Project, 2005, n° IST-2001-35454.

[61] 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.

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

[63] 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.

[64] 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 (IC-ITST)", London UK, November 2009.

[65] 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.

[66] G. OVERETT, L. PETERSSON, N. BREWER, L. ANDERSSON, N. PETTERSSON. *A New Pedestrian Dataset for Supervised Learning*, in "In IEEE Intelligent Vehicles Symposium", 2008.

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

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

[69] 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, p. 477-498.

[70] G. PUSIOL, L. PATINO, E. CORVEE, F. BRÉMOND, 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.

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

[72] P. SAND, S. TELLER. *Particle Video: Long-Range Motion Estimation using Point Trajectories*, in "Proceedings of the 2006 CVPR", IEEE Computer Society, 2006, p. 2195–2202.

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

[74] 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.

[75] 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).

[76] 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.

[77] X. WANG, T. HAN, S. YAN. *An HOG-LBP human detector with partial occlusion handling*, in "ICCV09 - International Conference on Computer Vision", 2009.

[78] J. YAO, J.-M. ODOBEZ. *Fast Human Detection from Videos Using Covariance Features*, in "The Eighth International Workshop on Visual Surveillance - VS2008", Marseille, France, Graeme Jones and Tieniu Tan and Steve Maybank and Dimitrios Makris, 2008, http://hal.inria.fr/inria-00325628.