Activity Report 2012

Team STARS

Spatio-Temporal Activity Recognition Systems

Table of contents

Team STARS

Keywords: Perception, Semantics, Machine Learning, Software Engineering, Cognitive Vision

Creation of the Team: January 01, 2012 *, Updated into Project-Team:* January 01, 2013 .

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

Etienne Corvée [Research engineer at Link Care Services]
Daniel Gaffé [Assistant professor, Faculty Member Nice University and CNRS-LEAT Member, on secondment since September 2012]
Aurelie Gouze [Research engineer at CSTB, since December 2012]
Veronique Joumier [Researcher Engineer, CHU Nice University, upto November 2012]
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]

Engineers

Slawomir Bak [Development Engineer, VideoID, since August 2012]
Vasanth Bathrinarayanan [Development Engineer, VICOMO Project]
Bernard Boulay [Development Engineer, COFRIEND and QUASPER Projects, upto October 2012]
Duc Phu Chau [Development Engineer, VANAHEIM Project, since March 2012]
Hervé Falciani [Development Engineer, [EIT ITC Labs, upto August 2012]
Baptiste Fosty [Development Engineer, since February 2012]
Julien Gueytat [Development Engineer, SWEET HOME Project]
Jihed Joobeur [Development Engineer, PAL AEN, upto September 2012]
Srinidhi Mukanahallipatna [Development Engineer, PAL AEN]
Anh-Tuan Nghiem [Development Engineer, since September 2012]
Jose-Luis Patino Vilchis [Development Engineer, COFRIEND and VANAHEIM Projects, upto June 2012]
Guido-Tomas Pusiol [Development Engineer, since June 2012]
Leonardo Rocha [Development Engineer, CIU Santé, SWEET HOME and VICOMO Projects, upto October 2012]
Silviu-Tudor Serban [Development Engineer, QUASPER Project, upto December 2012]
Sofia Zaidenberg [Development Engineer, VANAHEIM Project]
Salma Zouaoui-Elloumi [Development Engineer, VANAHEIM Project, since December 2012]

PhD Students

Julien Badie [Nice Sophia-Antipolis University, SWEET HOME Grant]
Slawomir Bak [Nice Sophia-Antipolis University, VideoID Grant, upto August 2012]
Piotr Bilinski [Nice Sophia-Antipolis University, Paca Grant]
Duc Phu Chau [Nice Sophia-Antipolis University, Paca Grant, upto March 2012]
Carolina Garate [Nice Sophia-Antipolis University, VANAHEIM Grant]
Ratnesh Kumar [Nice Sophia-Antipolis University, VANAHEIM Grant]
Guido-Tomas Pusiol [Nice Sophia-Antipolis University, CORDIs, upto June 2012]
Rim Romdhame [Nice Sophia-Antipolis University, CIU Santé Project]

Malik Souded [Nice Sophia-Antipolis University, Keeneo CIFRE Grant]

Post-Doctoral Fellow

Carlos-Fernando Crispim Junior [PAL AEN]

Administrative Assistants

Christine Claux [AI Inria, upto may 2012]
Sonia Rousseau [since June 2012 uo to end of July 2012]
Jane Desplanques [since September 2012]

Others

Pierre Aittahar [since April 2012 upto June 2012]
Guillaume Barbe [since April 2012 upto June 2012]
Sorana-Maria Capalnean [EGIDE, since July 2012 upto October 2012]
Cintia Corti [EGIDE, since May 2012 upto November 2012]
Eben Freeman [EGIDE, since June 2012 upto September 2012]
Vaibhav Katiyar [ACET, since July 2012 upto December 2012]
Vannara Loch [since April 2012 upto June 2012]
Qiao Ma [China, EGIDE, since July 2012 upto October 2012]
Firat Ozemir [since June 2012 upto September 2012]
Luis-Emiliano Sanchez [EGIDE, since September 2012 upto end of December 2012]
Bertrand Simon [ENS Lyon, since June 2012 upto mid-July 2012]
Abhineshwar Tomar [ACET, since November 2012]
Swaminathan Sankaranarayanan [EGIDE, upto June 2012]

2. Overall Objectives

2.1. Presentation

2.1.1. Research Themes

STARS (Spatio-Temporal Activity Recognition Systems) is focused on the design of cognitive systems for Activity Recognition. We aim at endowing cognitive systems with perceptual capabilities to reason about an observed environment, to provide a variety of services to people living in this environment while preserving their privacy. In today world, a huge amount of new sensors and new hardware devices are currently available, addressing potentially new needs of the modern society. However the lack of automated processes (with no human interaction) able to extract a meaningful and accurate information (i.e. a correct understanding of the situation) has often generated frustrations among the society and especially among older people. Therefore, Stars objective is to propose novel autonomous systems for the **real-time semantic interpretation of dynamic scenes** observed by sensors. We study long-term spatio-temporal activities performed by several interacting agents such as human beings, animals and vehicles in the physical world. Such systems also raise fundamental software engineering problems to specify them as well as to adapt them at run time.

We propose new techniques at the frontier between computer vision, knowledge engineering, machine learning and software engineering. The major challenge in semantic interpretation of dynamic scenes is to bridge the gap between the task dependent interpretation of data and the flood of measures provided by sensors. The problems we address range from physical object detection, activity understanding, activity learning to vision system design and evaluation. The two principal classes of human activities we focus on, are assistance to older adults and video analytics.

A typical example of a complex activity is shown in Figure 1 and Figure 2 for a homecare application. In this example, the duration of the monitoring of an older person apartment could last several months. The activities involve interactions between the observed person and several pieces of equipment. The application goal is to recognize the everyday activities at home through formal activity models (as shown in Figure 3) and data captured by a network of sensors embedded in the apartment. Here typical services include an objective assessment of the frailty level of the observed person to be able to provide a more personalized care and to monitor the effectiveness of a prescribed therapy. The assessment of the frailty level is performed by an Activity Recognition System which transmits a textual report (containing only meta-data) to the general practitioner who follows the older person. Thanks to the recognized activities, the quality of life of the observed people can thus be improved and their personal information can be preserved.
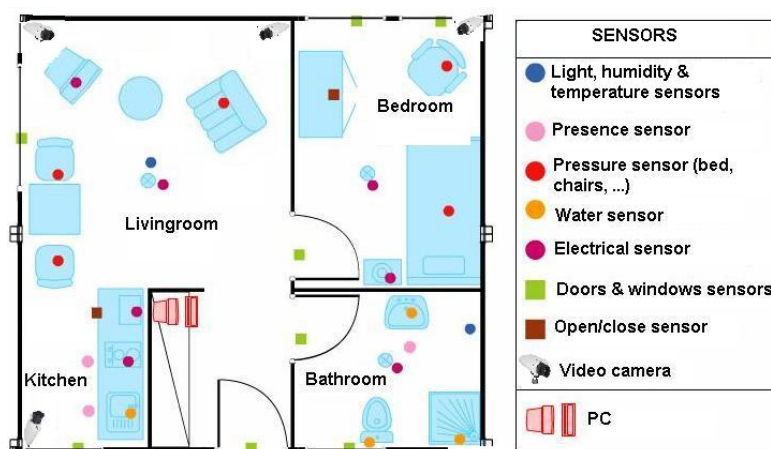
Figure 1. Homecare monitoring: the set of sensors embedded in an apartment
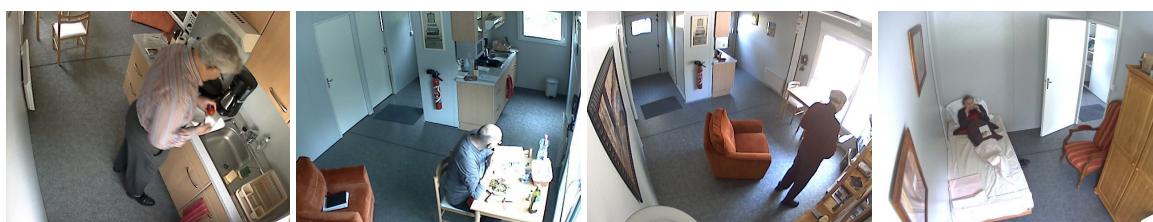
Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

The ultimate goal is for cognitive systems to perceive and understand their environment to be able to provide appropriate services to a potential user. An important step is to propose a computational representation of people activities to adapt these services to them. Up to now, the most effective sensors have been video cameras due to the rich information they can provide on the observed environment. These sensors are currently perceived as intrusive ones. A key issue is to capture the pertinent raw data for adapting the services to the people while preserving their privacy. We plan to study different solutions including of course the local processing of the data without transmission of images and the utilisation of new compact sensors developed

Activity (*PrepareMeal*,
   PhysicalObjects((p : Person), (z : Zone), (eq : Equipment))
   Components((s_inside : InsideKitchen(p, z))
                                           (s_close : CloseToCountertop(p, eq))
                                         (s_stand : PersonStandingInKitchen(p, z)))
   Constraints((z->Name = Kitchen)
                                         (eq->Name = Countertop)
                                       (s_close->Duration >= 100)
                                       (s_stand->Duration >= 100))
   Annotation(AText("prepare meal")
                                         AType("not urgent")))

Figure 3. Homecare monitoring: example of an activity model describing a scenario related to the preparation of a meal with a high-level language

for interaction (also called RGB-Depth sensors, an example being the Kinect) or networks of small non visual sensors.

2.1.2. *International and Industrial Cooperation*

Our work has been applied in the context of more than 10 European projects such as COFRIEND, ADVISOR, SERKET, CARETAKER, VANAHEIM, SUPPORT, DEM@CARE, VICOMO. We had or have industrial collaborations in several domains: *transportation* (CCI Airport Toulouse Blagnac, SNCF, Inrets, Alstom, Ratp, GTT (Italy), Turin GTT (Italy)), *banking* (Crédit Agricole Bank Corporation, Eurotelis and Ciel), *security* (Thales R&T FR, Thales Security Syst, EADS, Sagem, Bertin, Alcatel, Keeneo), *multimedia* (Multitel (Belgium), Thales Communications, Idiap (Switzerland)), *civil engineering* (Centre Scientifique et Technique du Bâtiment (CSTB)), *computer industry* (BULL), *software industry* (AKKA), *hardware industry* (ST-Microelectronics) and *health industry* (Philips, Link Care Services, Vistek).

We have international cooperations with research centers such as Reading University (UK), ENSI Tunis (Tunisia), National Cheng Kung University, National Taiwan University (Taiwan), MICA (Vietnam), IPAL, I2R (Singapore), University of Southern California, University of South Florida, University of Maryland (USA).

2.2. Highlights of the Year

Stars designs cognitive vision systems for activity recognition based on sound software engineering paradigms.

This year, we have designed several novel algorithms for activity recognition systems. In particular, we have extended 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. This algorithm outperforms state-of-the-art approaches on several benchmarking datasets (e.g. Ilids). We have realized a new algorithm for the recognition of short actions and validated also its performance on several benchmarking databases (e.g. ADL). We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. We have extended an original work on learning techniques such as data mining in large multimedia databases based on offline trajectory clustering. We have designed a generic controller algorithm, which is able to automatically tune the parameters of tracking algorithms.

We have also continued a large clinical trial with Nice Hospital to characterize the behaviour profile of Alzheimer patients compared to healthy older people.

We have organized a summer school which was held at Inria in October 2012, entitled "Human Activity and Vision Summer School", with many prestigious researchers (e.g. M. Shah).

3. Scientific Foundations

3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved**: the software architecture direction provides new methodologies for building safe activity recognition systems and the perception and the semantic activity recognition directions provide new activity recognition techniques which are designed and validated for concrete video analytics and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversally, we consider a new research axis in machine learning, combining a priori knowledge and learning techniques, to set up the various models of an activity recognition system. A major objective is to automate model building or model enrichment at the perception level and at the understanding level.

3.2. Perception for Activity Recognition

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

Computer Vision; Cognitive Systems; Learning; Activity Recognition.

3.2.1. Introduction

Our main goal in perception is to develop vision algorithms able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives. We have also several issues related to perception which combine machine learning and perception techniques: learning people appearance, parameters for system control and shape statistics.

3.2.2. Appearance models and people tracking

An important issue is to detect in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios.

Appearance models. In particular, we study the temporal variation of the features characterizing the appearance of a human. This task could be achieved by clustering potential candidates depending on their position and their reliability. This task can provide any people tracking algorithms with reliable features allowing for instance to (1) better track people or their body parts during occlusion, or to (2) model people appearance for re-identification purposes in mono and multi-camera networks, which is still an open issue. The underlying challenge of the person re-identification problem arises from significant differences in illumination, pose and camera parameters. The re-identification approaches have two aspects: (1) establishing correspondences between body parts and (2) generating signatures that are invariant to different color responses. As we have already several descriptors which are color invariant, we now focus more on aligning two people detections and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. As the results of the re-identification are always given as the ranking list, re-identification focuses on learning to rank. "Learning to rank" is a type of machine learning problem, in which the goal is to automatically construct a ranking model from a training data.

Therefore, 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). New 3D sensors (e.g. Kinect) are also investigated, to help in getting an accurate segmentation for specific scene conditions.

Long term tracking. For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modelling and managing the uncertainty of these processes is also an open issue. In Stars we propose to add a reasoning layer to a classical Bayesian framework1pt modelling the uncertainty of the tracked objects. This reasoning layer can take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

Controling system parameters. Another 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. Learning shape and motion

Another approach, to improve jointly segmentation and tracking, is to consider videos as 3D volumetric data and to search for trajectories of points that are statistically coherent both spatially and temporally. This point of view enables new kinds of statistical segmentation criteria and ways to learn them.

We are also using the shape statistics developed in [5] for the segmentation of images or videos with shape prior, by learning local segmentation criteria that are suitable for parts of shapes. This unifies patch-based detection methods and active-contour-based segmentation methods in a single framework. These shape statistics can be used also for a fine classification of postures and gestures, in order to extract more precise information from videos for further activity recognition. In particular, the notion of shape dynamics has to be studied.

More generally, to improve segmentation quality and speed, different optimization tools such as graph-cuts can be used, extended or improved.

3.3. Semantic Activity Recognition

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

Activity Recognition, Scene Understanding,Computer Vision

3.3.1. Introduction

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

Thus we work along the two following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models) and learning (how to learn the models needed for activity recognition).

3.3.2. High Level Understanding

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and 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 modelling of activities, we are working towards two directions: the uncertainty management for representing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. We have also built a language for video event modelling 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 ...).

3.3.3. 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 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 to learn 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.

3.3.4. Activity Recognition and Discrete Event Systems

The previous research axes are unavoidable to cope with the semantic interpretations. However they tend to let aside the pure event driven aspects of scenario recognition. These aspects have been studied for a long time at a theoretical level and led to methods and tools that may bring extra value to activity recognition, the most important being the possibility of formal analysis, verification and validation.

We have thus started to specify a formal model to define, analyze, simulate, and prove scenarios. This model deals with both absolute time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension, and verification). Our purpose is to offer a generic tool to express and recognize activities associated with a concrete language to specify activities in the form of a set of scenarios with temporal constraints. The theoretical foundations and the tools being shared with Software Engineering aspects, they will be detailed in section 3.4.

The results of the research performed in perception and semantic activity recognition (first and second research directions) produce new techniques for scene understanding and contribute to specify the needs for new software architectures (third research direction).

3.4. Software Engineering for Activity Recognition

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

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

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 research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. We work on the following three research axes: models (adapted to the activity recognition domain), platform architecture (to cope with deployment constraints and run time adaptation), and system verification (to generate dependable systems). For all these tasks we follow state of the art Software Engineering practices and, if needed, we attempt to set up new ones.

3.4.1. *Platform Architecture for Activity Recognition*
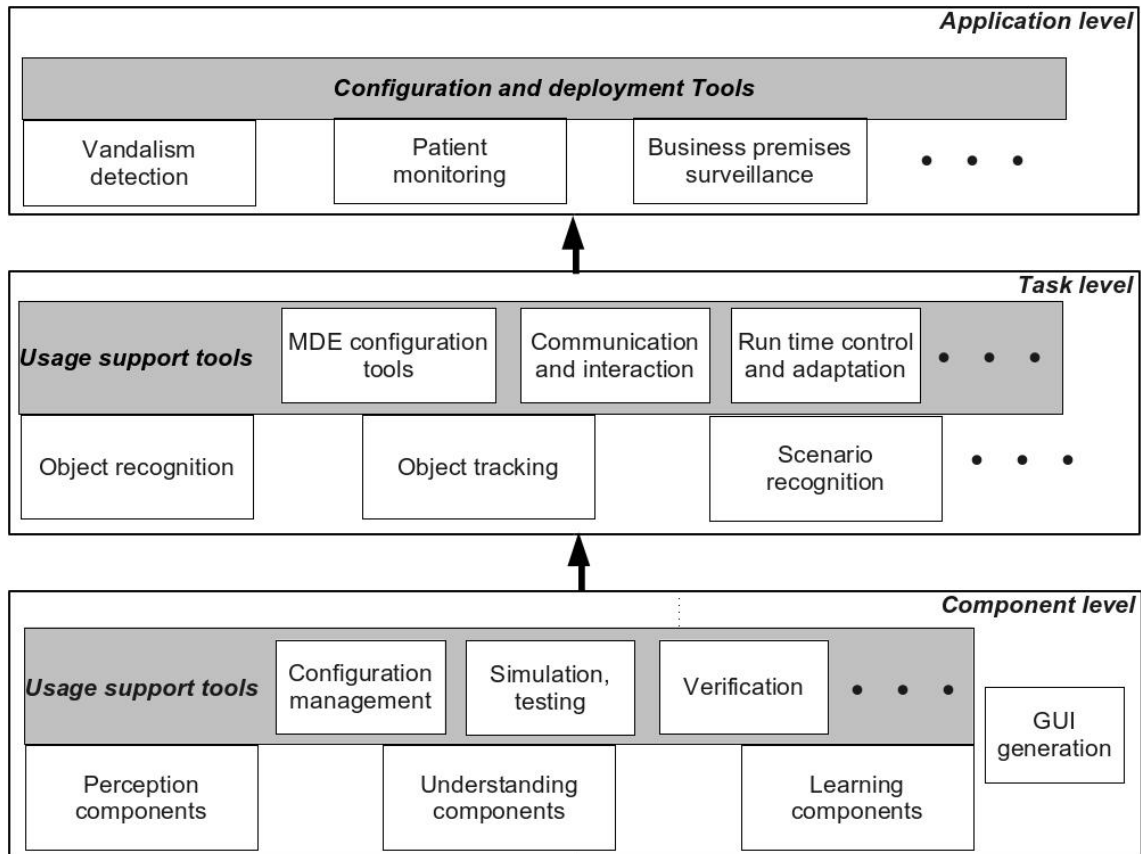
Figure 4. Global Architecture of an Activity Recognition The grey areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [15], 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.

Figure 4 presents our conceptual vision for the architecture of an activity recognition platform. It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.

– Perception components contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.

– Understanding components provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.

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

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

• The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate in a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Object Recognition, Tracking, Scenario Recognition... In the future, other tasks will probably enrich this level.

For these tasks to nicely collaborate, communication and interaction facilities are needed. We shall also add MDE-enhanced tools for configuration and run-time adaptation.

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

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

To cope with real application requirements, we shall also investigate distributed architecture, real time implementation, and user interfaces.

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving, etc. Note that, in Figure 4, some of the boxes can be naturally adapted from SUP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

3.4.2. Discrete Event Models of Activities

As mentioned in the previous section (3.3) we have started to specify a formal model of scenario dealing with both absolute time and logical time. Our scenario and time models as well as the platform verification tools rely on a formal basis, namely the synchronous paradigm. To recognize scenarios, we consider activity descriptions as synchronous reactive systems and we apply general modelling methods to express scenario behaviour.

Activity recognition systems usually exhibit many safeness issues. From the software engineering point of view we only consider software security. Our previous work on verification and validation has to be pursued; in particular, we need to test its scalability and to develop associated tools. Model-checking is an appealing technique since it can be automatized and helps to produce a code that has been formally proved. Our verification method follows a compositional approach, a well-known way to cope with scalability problems in model-checking.

Moreover, recognizing real scenarios is not a purely deterministic process. Sensor performance, precision of image analysis, scenario descriptions may induce various kinds of uncertainty. While taking into account this uncertainty, we should still keep our model of time deterministic, modular, and formally verifiable. To formally describe probabilistic timed systems, the most popular approach involves probabilistic extension of timed automata. New model checking techniques can be used as verification means, but relying on model checking techniques is not sufficient. Model checking is a powerful tool to prove decidable properties but introducing uncertainty may lead to infinite state or even undecidable properties. Thus model checking validation has to be completed with non exhaustive methods such as abstract interpretation.

3.4.3. *Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems*

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [54]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify models to represent transformations. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [70]. In particular, the constraints can be analysed as a SAT problem.

An additional challenge is to manage the possible run-time changes of implementation due to context variations (e.g., lighting conditions, changes in the reference scene, etc.). Video surveillance systems have to dynamically adapt to a changing environment. The use of models at run-time is a solution. We are defining adaptation rules corresponding to the dependency constraints between specification elements in one model and software variants in the other [51], [80], [72].

4. Application Domains

4.1. Introduction

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 introduce. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, Stars focuses on two main domains: **video analytics** and **healthcare monitoring**.

4.2. Video Analytics

Our experience in video analytics [7], [1], [9] (also referred to as visual surveillance) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

For instance, the Keeneo start-up was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in video analytics (VSIP library, which was a previous version of SUP). Keeneo has been bought by Digital Barriers in August 2011 and is now independent from Inria. However, Stars continues to maintain a close cooperation with Keeneo for impact analysis of VSIP and for exploitation of new results.

Moreover new challenges are arising from the visual surveillance community. For instance, people detection and tracking in a crowded environment are still open issues despite the high competition on these topics. Also detecting abnormal activities may require to discover rare events from very large video data bases often characterized by noise or incomplete data.

4.3. Healthcare Monitoring

We have initiated a new strategic partnership (called CobTek) with Nice hospital [62], [81] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longer analysis. Finer activity description is needed for instance, to discriminate the activities (e.g. sitting, walking, eating) of Alzheimer patients from the ones of healthy older people. It is essential to be able to pre-diagnose dementia and to provide a better and more specialised care. Longer analysis is required when people monitoring aims at measuring the evolution of patient behavioural disorders. Setting up such long experimentation with dementia people has never been tried before but is necessary to have real-world validation. This is one of the challenge of the European FP7 project Dem@Care where several patient homes should be monitored over several months.

For this domain, a goal for Stars is to allow people with dementia to continue living in a self-sufficient manner in their own homes or residential centers, away from a hospital, as well as to allow clinicians and caregivers remotely proffer effective care and management. For all this to become possible, comprehensive monitoring of the daily life of the person with dementia is deemed necessary, since caregivers and clinicians will need a comprehensive view of the person's daily activities, behavioural patterns, lifestyle, as well as changes in them, indicating the progression of their condition.

The development and ultimate use of novel assistive technologies by a vulnerable user group such as individuals with dementia, and the assessment methodologies planned by Stars are not free of ethical, or even legal concerns, even if many studies have shown how these Information and Communication Technologies (ICT) can be useful and well accepted by older people with or without impairments. Thus one goal of Stars team is to design the right technologies that can provide the appropriate information to the medical carers while preserving people privacy. Moreover, Stars will pay particular attention to ethical, acceptability, legal and privacy concerns that may arise, addressing them in a professional way following the corresponding established EU and national laws and regulations, especially when outside France.

As presented in 3.1, Stars aims at designing cognitive vision systems with perceptual capabilities to monitor efficiently people activities. As a matter of fact, vision sensors can be seen as intrusive ones, even if no images are acquired or transmitted (only meta-data describing activities need to be collected). Therefore new communication paradigms and other sensors (e.g. accelerometers, RFID, and new sensors to come in the future) are also envisaged to provide the most appropriate services to the observed people, while preserving their privacy. To better understand ethical issues, Stars members are already involved in several ethical organizations. For instance, F. Bremond has been a member of the ODEGAM - "Commission Ethique et Droit" (a local association in Nice area for ethical issues related to older people) from 2010 to 2011 and a member of the French scientific council for the national seminar on "La maladie d'Alzheimer et les nouvelles technologies - Enjeux éthiques et questions de société" in 2011. This council has in particular proposed a chart and guidelines for conducting researches with dementia patients.

For addressing the acceptability issues, focus groups and HMI (Human Machine Interaction) experts, will be consulted on the most adequate range of mechanisms to interact and display information to older people.

5. Software

5.1. SUP

Figure 5. Tasks of the Scene Understanding Platform (SUP).

SUP is a Scene Understanding Software Platform written in C and C++ (see Figure 5). 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., up to activity recognition, to achieve the tasks (detection, classification, etc.) the platform supplies. Each module has a specific 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 Stars 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.

Goals of SUP are twofold:

1. From a video understanding point of view, to allow the Stars researchers sharing the implementation of their work through this platform.
2. From a software engineering point of view, to integrate the results of the dynamic management of vision applications when applied to video analytics.

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, classification and 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 modify or 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 for 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 (e.g. 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 provides 4 metrics.

Figure 6. GUI of the ViSEvAl software
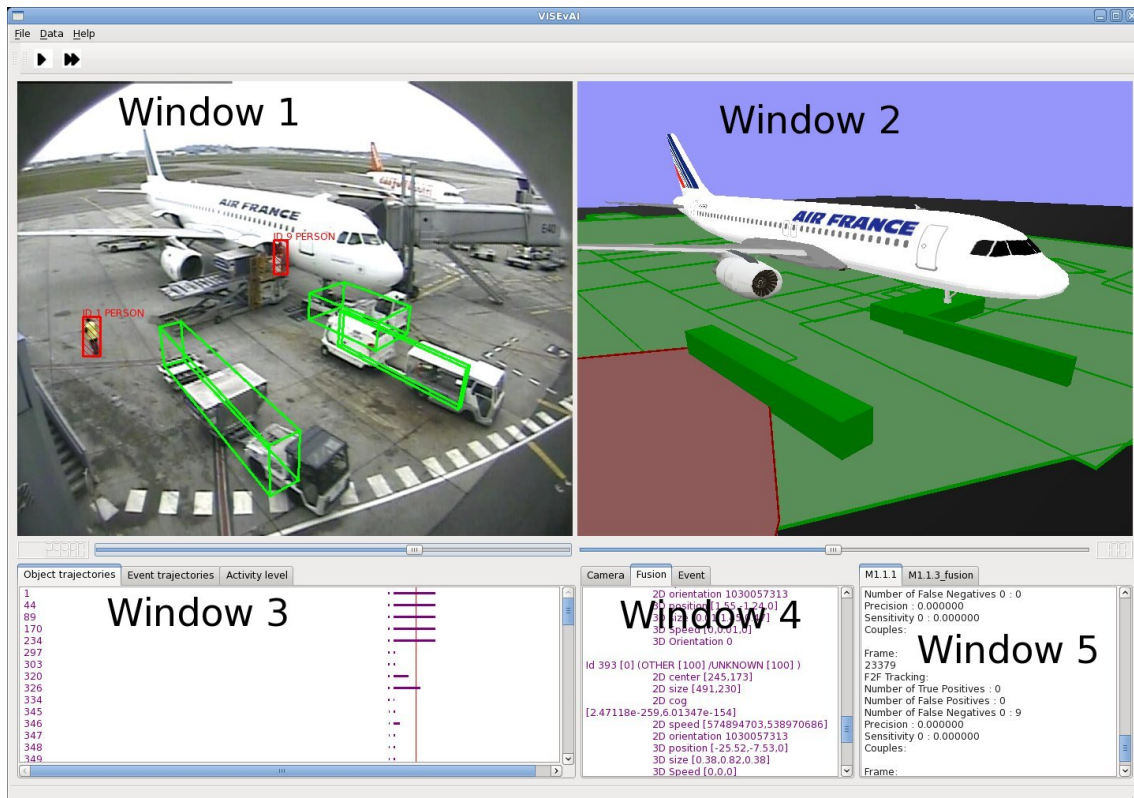
The GUI is composed of 3 different parts:

1. The widows dedicated to result visualization (see Figure 6):
   – Window 1: the video window displays the current image and information about the detected and ground-truth objects (bounding-boxes, identifier, type,...).
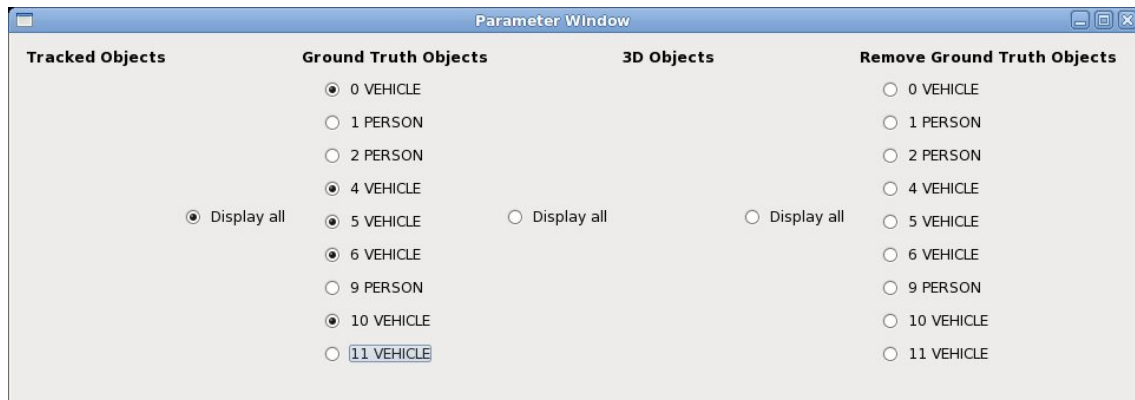
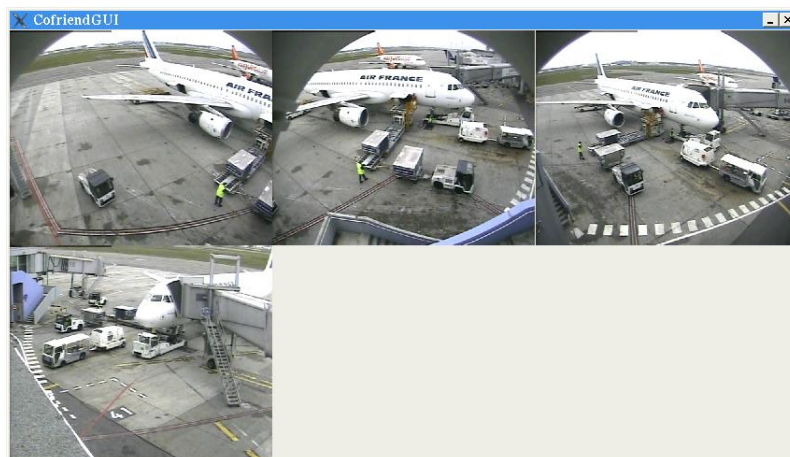*Figure 7. The object window enables users to choose the object to display*

Figure 8. The multi-view window

– 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.

2. The object window enables the user to choose the object to be displayed (see Figure 7).

3. The multi-view window displays the different points of view of the scene (see Figure 8).

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

The ViSEvAl software was tested and validated into the context 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. Clem

The *Clem Toolkit* [63](see Figure 9) is a set of tools devoted to design, simulate, verify and generate code for LE [19] [77] programs. LE is a synchronous language supporting a modular compilation. It also supports automata possibly designed with a dedicated graphical editor.

Each LE program is compiled later into lec and lea files. Then when we want to generate code for different backends, depending on their nature, we can either expand the lec code of programs in order to resolve all abstracted variables and get a single lec file, or we can keep the set of lec files where all the variables of the main program are defined. Then, the *finalization* will simplify the final equations and code is generated for simulation, safety proofs, hardware description or software code. Hardware description (Vhdl) and software code (C) are supplied for LE programs as well as simulation. Moreover, we also generate files to feed the NuSMV model checker [61] in order to perform validation of program behaviors.

6. New Results

6.1. Introduction

This year Stars has proposed new algorithms related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

6.1.1. *Perception for Activity Recognition*

Participants: Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, Bernard Boulay, François Brémond, Sorana Capalnean, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Eben Freeman, Carolina Garate, Jihed Joober, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Sabine Moisan, Silviu Serban, Malik Souded, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.
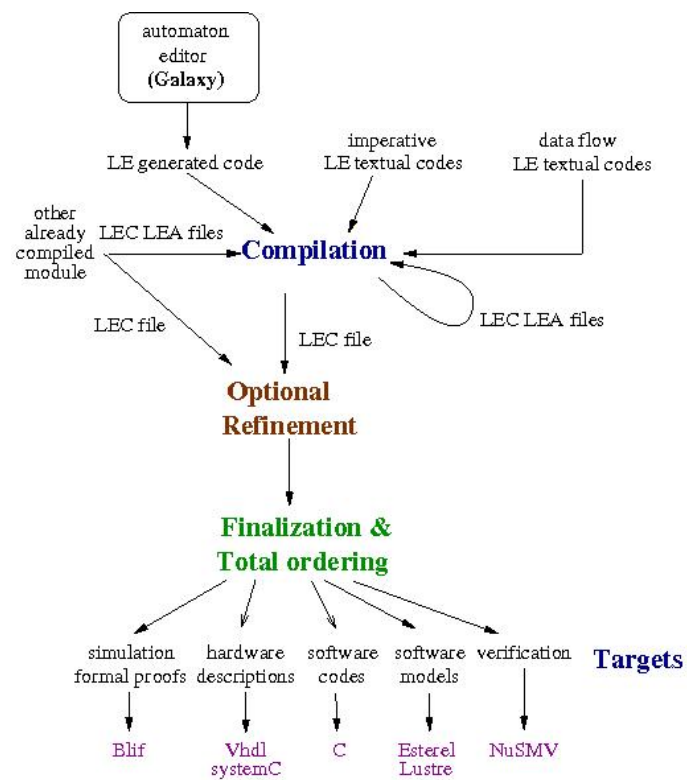
Figure 9. The Clem Toolkit

This year Stars has extended an efficient algorithm for detecting people. 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 validated also its performance on several benchmarking databases (e.g. ADL). We have improved a generic event recognition algorithm by handling event uncertainty at several processing levels. More precisely, the new results for perception for activity recognition concern:

- Image Compression and Modelization (6.2)
- Background Subtraction (6.3)
- Fiber Based Video Segmentation (6.4)
- Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation (6.5)
- Dynamic and Robust Object Tracking in a Single Camera View (6.6)
- Optimized Cascade of Classifiers for People Detection Using Covariance Features (6.7)
- Learning to Match Appearances by Correlations in a Covariance Metric Space (6.8)
- Recovering Tracking Errors with Human Re-identification (6.9)
- Human Action Recognition in Videos (6.10)
- Group Interaction and Group Tracking for Video-surveillance in Underground Railway Stations (6.11)
- Crowd Event Monitoring Using Texture and Motion Analysis (6.12)
- Detecting Falling People (6.13)
- People Detection Framework (6.14)

6.1.2. Semantic Activity Recognition

Participants: Sorana Capalnean, Guillaume Charpiat, Cintia Corti, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qioa Ma, Firat Ozemir, Jose-Luis Patino Vilchis, Guido-Tomas Pusiol, Rim Romdhame, Bertrand Simon, Abhineshwar Tomar.

Concerning semantic activity recognition, the contributions are :

- A Model-based Framework for Activity Recognition of Older People using Multiple sensors (6.15)
- Activity Recognition for Older People using Kinect (6.16)
- Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection (6.17)
- Online Activity Learning from Subway Surveillance Videos (6.18)
- Automatic Activity Detection Modeling and Recognition: ADMR (6.19)

6.1.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Julien Gueytat, Baptiste Fosty, Sabine Moisan, Anh tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Leonardo Rocha, Luis-Emiliano Sanchez, Swaminathan Sankaranarayanan.

This year Stars has continued the development of the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms. We continue to improve our meta-modelling approach to support the development of video surveillance applications based on SUP. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes. We continue the development of a scenario analysis module (SAM) relying on formal methods to support activity recognition in SUP platform. We improve the theoretical foundations of CLEM toolkit and we rely on it to build SAM. Finally, we are improving the way we perform adaptation in the definition of a multiple services for device adaptive platform for scenario recognition.

The contributions for this research axis are:

- SUP Software Platform (6.20)
- Qualitative Evaluation of Detection and Tracking Performance (6.21)
- Model-Driven Engineering and Video-surveillance (6.22)
- Synchronous Modelling and Activity Recognition (6.23)

6.2. Image Compression and Modelization

Participants: Guillaume Charpiat, Eben Freeman.

Recent results in statistical learning have established the best strategy to combine several advices from different experts, for the problem of sequential prediction of times series. The notions of prediction and compression are tightly linked, in that a good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding), based on the predicted probabilities of the events to come : the more predictable an event E is, the easier to compress it will be, with coding cost $-\log(p(E))$ with such techniques.

The initial idea here, by Yann Ollivier (TAO team), within a collaboration with G. Charpiat and Jamal Atif (TAO team), was to adapt these results to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors. The main difference is that there is no canonical physically-relevant 1D ordering of the pixels in an image, so that a sequential order (of the pixels to predict their colors) had to be defined first. Preliminary results with a hierarchical ordering scheme already competed with standard techniques in lossless compression (png, lossless jpeg2000).

During his internship in the Stars team, Eben Freeman developed this approach, by building relevant experts able to predict a variety of image features (regions of homogeneous color, edges, noise, . . .). We also considered random orderings of pixels, using kernels to express probabilities in a spatially-coherent manner. Using such modellings of images with experts, we were also able to generate new images, that are typical of these models, and show more structure than the ones associated to standard compression schemes (typical images highly compressed).

6.3. Background Subtraction

Participants: Vasanth Bathrinarayanan, Anh-Tuan Nghiem, Duc-Phu CHAU, François Brémond.

Keywords: Gaussian Mixture Model, Shadow removal, Parameter controller, Codebook model, Context based information

6.3.1. *Statistical Background Subtraction for Video Surveillance Platform*

Anh-Tuan Nghiem work on background subtraction is an extended version of Gaussian Mixture Models [73]. The algorithm compares each pixel of current frame to background representation which is developed based on the pixel information from previous frames. It includes shadow and highlight removal to give better results. Selective background updating method based on the feedback from the object detection helps to better model background and remove noise and ghosts.

Figure 10 shows a sample illustration of the output of the background subtraction, where blue are foreground pixels and red are shadow or illumination change pixels and a green bounding box is a foreground blob. Also we have compared our algorithm with few other such as OpenCV and also IDIAP's background subtraction(not tuned perfectly, used default parameters) and the results are shown in Figure 11 where the green background refers to best performance of the comparisons. This evaluation is done on PETS 2009 data-set with our obtained foreground blobs to the manually annotated bounding boxes of people.

6.3.2. *Parameter controller using Contextual features*

The above method has some parameters that has to be tuned every time for each video, which is a time consuming work. The work of Chau et al [59] learns the contextual information from the video and controls object tracking algorithm parameters during the run-time of the algorithm. This approach is at preliminary stage for background subtraction algorithm to automatically adapt parameters. These parameters are learned as described in the offline learning process block diagram 12 over several ground truth videos and clustered into a database. The contextual feature which are used presently include object density, occlusion, contrast, 2D area, contrast variance, 2D area variance. Figure 13 shows a sample of video chunks based on contextual feature similarity for a video from caviar data-set.

The controller's preliminary results are promising and we are experimenting and evaluating with different features to learn the parameters. The results will be published in upcoming top computer vision conferences.

Figure 10. Sample illustration of output of EGMM

PETS dataset Evaluation on Background subtraction blob with GT of people--			
Metrics	**IDIAP**	**SUP**	**OPENCV**
Metric M1.1			
Global results:			
Number of True Positives	3761	4046	4057
Number of False Positives	1807	317	1099
Number of False Negatives 0	1070	793	774
Precision (mean by frame)	0.71	0.93	0.81
Sensitivity 0 (mean by frame)	0.79	0.84	0.85
Precision (global)	0.68	0.93	0.79
Sensitivity 0 (global)	0.78	0.84	0.84

Figure 11. Evaluation with some background subtraction algorithms
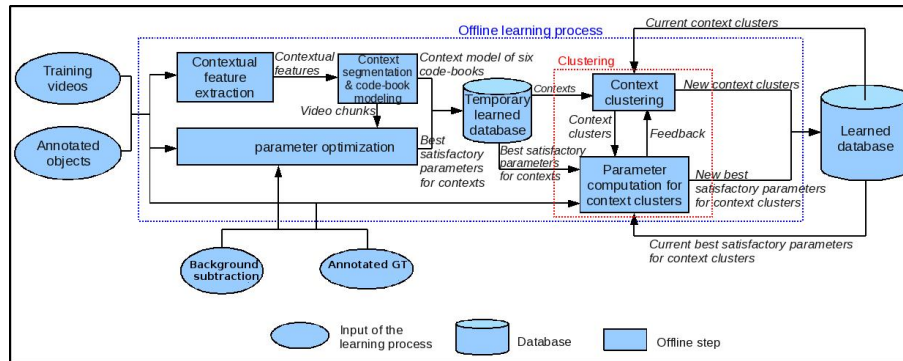
6.4. Fiber Based Video Segmentation

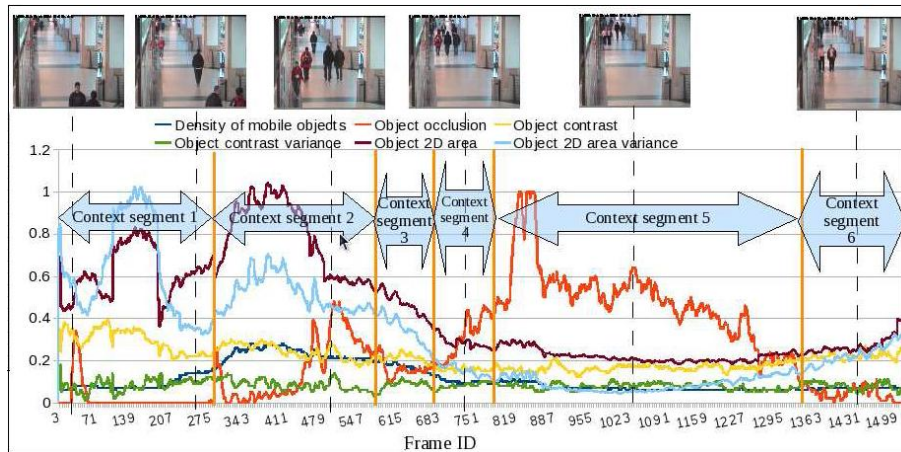*Figure 12. Block representation of the offline learning scheme to build a database*

Figure 13. Context segmentation of the sequence ThreePastShop2cor (belonging to the Caviar dataset). The context segments are separated by the vertical orange lines. The control parameters are then learned for each context segment

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

Keywords: Video Volume, Fibers, Trajectory

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data (2D×time). Figure 14 shows an input video and its corresponding partition in terms of fiber at a particular hierarchy level. Particularly, it 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. Analogically to the fibers in MRI images of human brains, we term *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.

Figure 14. **Left:** *Input Video and Spatio-Temporal Slices.* **Right:** *Segmented Results at a Particular Hierarchy Level*

Initial fibers are built using correspondences computing algorithms like optical flow and descriptor matching. As these algorithms are reliable near corners and edges, we build fibers at these locations for a video. Our subsequent goal is to partition this video in terms of fibers built, by extending them (both spatially and temporally) to the rest of the video.

To extend fibers, we compute geodesics from pixels (not belonging to the initially built fibers) to fibers. For a reliable extension, the cost of moving along a geodesic is proportional to the trajectory similarity of a pixel *wrt* a fiber, wherein a pixel trajectory is similar to the fiber trajectory. This cost function quantifies the color homogeneity of a pixel trajectory along with its color similarity *wrt* a fiber. A pixel is then associated to a fiber for which this cost is minimum.

With the above mentioned steps we obtain a partition of a video in terms of fibers wherein we have a trajectory associated with each pixel. This hierarchical partition provides a mid-level representation of a video, which can be seen as a facilitator or a pre-processing step towards higher level video understanding systems *eg* activity recognition.

6.5. Enforcement of Monotonous Shape Growth/Shrinkage in Video Segmentation

**Participant:** Guillaume Charpiat.

keywords: graph cuts, video segmentation, shape growth

The segmentation of noisy videos or time series is a difficult problem, not to say an impossible or ill-posed task when the noise level is very high. While individual frames can be analysed independently, time coherence in image sequences provides a lot of information not available for a single image. Most of the state-of-art works explored short-term temporal continuity for object segmentation in image sequences, *i.e.*, each next frame is segmented by using information from one or several images at previous time points. It is, however, more advantageous to simultaneously segment many frames in the data set, so that segmentation of the entire image set supports each of the individual segmentations.

In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process. Examples of growing shapes are forest fires in satellite images and organ development in medical imaging. We propose a segmentation framework based on graph cuts for the joint segmentation of a multi-dimensional image set. By minimizing an energy computed on the resulting spatio-temporal graph of the image sequence, the proposed method yields a *globally optimal solution*, and runs in practice in linear complexity in the total number of pixels.

Two applications are performed. First, with Yuliya Tarabalka (Ayin team), we segment multiyear sea ice floes in a set of satellite images acquired through different satellite sensors, after rigid alignment (see Figure 15). The method returns accurate melting profiles of sea ice, which is important for building climate models. The second application, with Bjoern Menze (ETH Zurich, also MIT and collaborator of Asclepios team), deals with the segmentation of brain tumors from longitudinal sets of multimodal MRI volumes. In this task we impose an additional inter-modal inclusion constraint for joint segmentation of different image sequences, finally also returning highly sensitive time-volume plots of tumor growth.

(a)
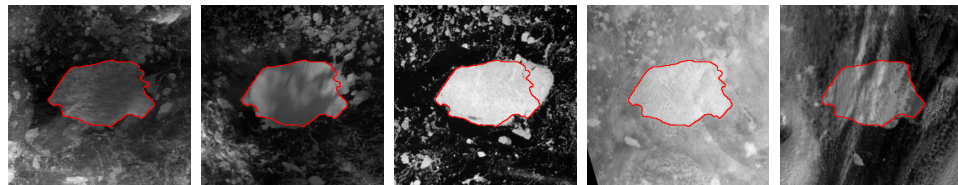
(b)
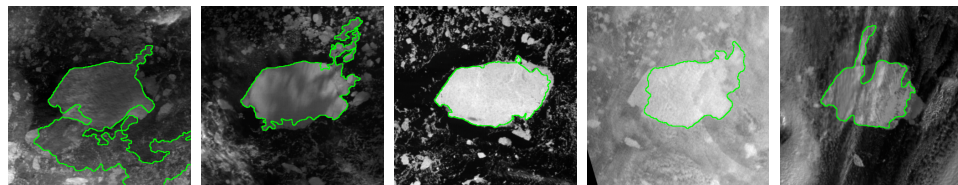
Figure 15. (a) Aligned satellite images captured each four days superposed with segmentation contours computed by our approach. (b) Segmentation contours for images (a) obtained by applying graph cut segmentation to each image at a single time moment. Note that the segmentations (a) are pixelwise precise, and that the white regions surrounding sometimes the boundaries are other ice blocks, agglomerating temporarily only, thus correctly labelled. Hence the importance of enforcing time coherence.

6.6. Dynamic and Robust Object Tracking in a Single Camera View

Participants: Duc-Phu Chau, Julien Badie, François Brémond, Monique Thonnat.

Keywords: Object tracking, online parameter tuning, controller, self-adaptation and machine learning

Object tracking quality usually depends on video scene conditions (e.g. illumination, density of objects, object occlusion level). In order to overcome this limitation, we present a new control approach to adapt the object tracking process to the scene condition variations. The proposed approach is composed of two tasks.

The objective of the first task is to select a convenient tracker for each mobile object among a Kanade-Lucas-Tomasi-based (KLT) tracker and a discriminative appearance-based tracker. The KLT feature tracker is used to decide whether an object is correctly detected. For badly detected objects, the KLT feature tracking is performed to correct object detection. A decision task is then performed using a Dynamic Bayesian Network (DBN) to select the best tracker among the discriminative appearance and KLT trackers.

The objective of the second task is to tune online the tracker parameters to cope with the tracking context variations. The tracking context, or context, of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. Each contextual feature is represented by a code-book model. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated with satisfactory tracking parameters. In the online control phase, once a context change is detected, the tracking parameters are tuned using the learned values. This work has been published in [29], [35].

We have tested the proposed approach on several public datasets such as Caviar and PETS. Figure 16 illustrates the results of the object detection correction using the KLT feature tracker.

Figure 16. Illustration of the object detection correction for a Caviar video. The green bounding box is the output of the object detection process. The red bounding boxes are the results of the detection correction task.

Figure 17 illustrates the tracking output for a Caviar video (on the left image) and for a PETS video (on the right image). The experimental results show that our method gets the best performance compared to some recent state of the art trackers.

Table 1 presents the tracking results for 20 videos from the Caviar dataset. The proposed approach obtains the best MT value (i.e. mostly tracked trajectories) compared to some recent state of the art trackers.

Figure 17. Tracking results for Caviar and PETS videos

Table 1. Tracking results on the Caviar dataset. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed bold.

Method	MT (%)	PT (%)	ML (%)
Zhang et al., CVPR 2008 [89]	85.7	10.7	3.6
Li et al., CVPR 2009 [71]	84.6	14.0	1.4
Kuo et al., CVPR 2010 [69]	84.6	14.7	**0.7**
Proposed approach	**86.4**	10.6	3.0

Table 2 presents the tracking results of the proposed approach and three recent approaches [56], [82], [67] for a PETS video. With the proposed approach, we obtain the best values in both metrics MOTA (i.e. Multi-object tracking accuracy) and MOTP (i.e. Multi-object tracking precision). The authors in [56], [82], [67] do not present the tracking results with the MT, PT and ML metrics.

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. MOTA: Multi-object tracking accuracy, higher is better. MOTP: Multi-object tracking precision, higher is better. The best values are printed bold.

Method	MOTA	MOTP	MT (%)	PT (%)	ML (%)
Berclaz et al., PAMI 2011 [56]	0.80	0.58	-	-	-
Shitrit et al., ICCV 2011 [82]	0.81	0.58	-	-	-
Henriques et al., ICCV 2011 [67]	0.85	0.69	-	-	-
Proposed approach	**0.86**	**0.72**	71.43	19.05	9.52

6.7. Optimized Cascade of Classifiers for People Detection Using Covariance Features

Participants: Malik Souded, François Brémond.

keywords: People detection, Covariance descriptor, LogitBoost.

We propose a new method to optimize a state of the art approach for people detection, which is based on classification on Riemannian manifolds using covariance matrices in a boosting scheme. Our approach makes training and detection faster while maintaining equivalent or better performances. This optimisation is achieved by clustering negative samples before training, providing a smaller number of cascade levels and less weak classifiers in most levels in comparison with the original approach.

Our approach is based on Tuzel et al. [86] work which was improved by Yao et al. [87]. We keep the same scheme to achieve our people detector: train a cascade of classifiers based on covariance descriptors, using a Logitboost training algorithm which was modified by Tuzel et al. to deal with the Riemannian manifolds metrics and using the operators which were presented in [75]. In fact, Covariance matrices do not belong to vector space but to the Riemannian manifold of (d x d) symmetric positive definite matrices. The trained cascade of classifiers is applied for detection after training.

We propose an additional step to speed up training and detection process. We propose to apply a clustering step on negative training dataset before training the classifiers. This clustering step is performed both in Riemannian manifold and in the vector space of mapped covariance matrices, using the operators and metrics previously cited.

The idea consists in regrouping all similar negative samples, with regard to their covariance information, into decreasing size clusters. Each classifier of the cascade is trained on one cluster, specializing this classifier for a given kind of covariance information, and then, speeding up the training step and providing shorter classifier, which accelerate its response when applied on image. In the same time, the specialization of each cascade classifier shortens the cascade too, speeding up the detection (see Figure 18 and Figure 19).

A paper describing this approach has been accepted in VISAPP 2013 conference [50].

6.8. Learning to Match Appearances by Correlations in a Covariance Metric Space

Participants: Sławomir Bąk, Guillaume Charpiat, Etienne Corvée, Francois Brémond, Monique Thonnat.
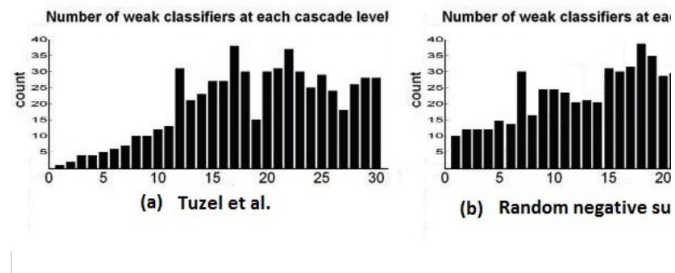
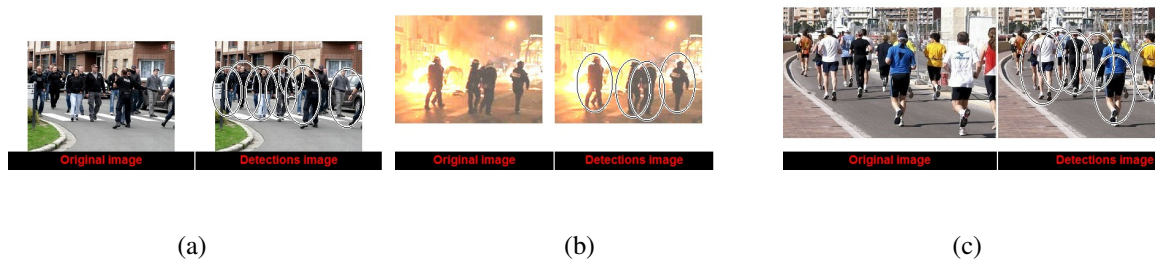*Figure 18. Comparison of classifiers cascade structure*

(a)   (b)   (c)

Figure 19. Cascade classifiers: some detection results

keywords: covariance matrix, re-identification, appearance matching

This work addresses the problem of appearance matching across disjoint camera views. Significant appearance changes, caused by variations in view angle, illumination and object pose, make the problem challenging.

We propose to formulate the appearance matching problem as the task of learning a model that selects the most descriptive features for a specific class of objects. Our main idea is that different regions of the object appearance ought to be matched using various strategies to obtain a distinctive representation. Extracting region-dependent features allows us to characterize the appearance of a given object class (*e.g.* class of humans) in a more efficient and informative way. Different kinds of features characterizing various regions of an object is fundamental to our appearance matching method.

We propose to model the object appearance using covariance descriptor yielding rotation and illumination invariance. Covariance descriptor has already been successfully used in the literature for appearance matching. In contrast to state of the art approaches, we do not define *a priori* feature vector for extracting covariance, but we learn which features are the most descriptive and distinctive depending on their localization in the object appearance (see figure 20). Learning is performed in a covariance metric space using an entropy-driven criterion. Characterizing a specific class of objects, we select only essential features for this class, removing irrelevant redundancy from covariance feature vectors and ensuring low computational cost.

The proposed technique has been successfully applied to the person re-identification problem, in which a human appearance has to be matched across non-overlapping cameras [34]. We demonstrated that: (1) by using different kinds of covariance features *w.r.t.* the region of an object, we obtain clear improvement in appearance matching performance; (2) our method outperforms state of the art methods in the context of pedestrian recognition on publicly available datasets (i-LIDS-119, i-LIDS-MA and i-LIDS-AA); (3) using 4×4 covariance matrices we significantly speed-up the processing time offering an efficient and distinctive representation of the object appearance.
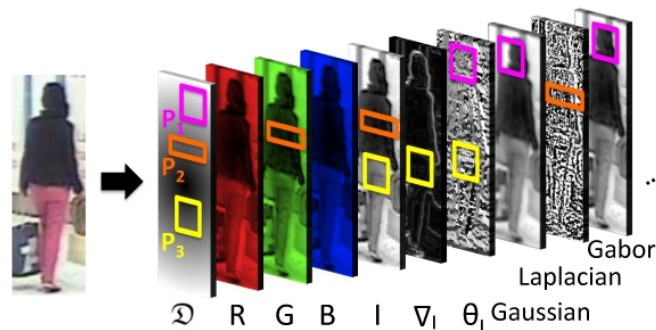
Figure 20. Example of three different covariance features. Every covariance is extracted from a region (P), distance layer (D) and three channel functions (e.g. bottom covariance feature is extracted from region P_3 using layers: D, I-intensity, ∇_I-gradient magnitude and θ_I-gradient orientation).

6.9. Recovering Tracking Errors with Human Re-identification

Participants: Julien Badie, Slawomir Bak, Duc-Phu Chau, François Brémond, Monique Thonnat.

keywords: tracking error correction, re-identification

This work addresses the problem of people tracking at long range even if the target people are lost several times by the tracking algorithm. We have identified two main reasons for tracking interruption. The first one concerns interruptions that can be quickly recovered, which includes short mis-detections, occlusions with other persons or static obstacles. The second one occurs when a person is occluded or mis-detected for a long time or when the person leaves the scene and comes back latter. Our main objective is to design a framework that can track people even if their trajectory is very segmented and/or associated with different IDs. We called this problem the global tracking challenge (see Figure 21).
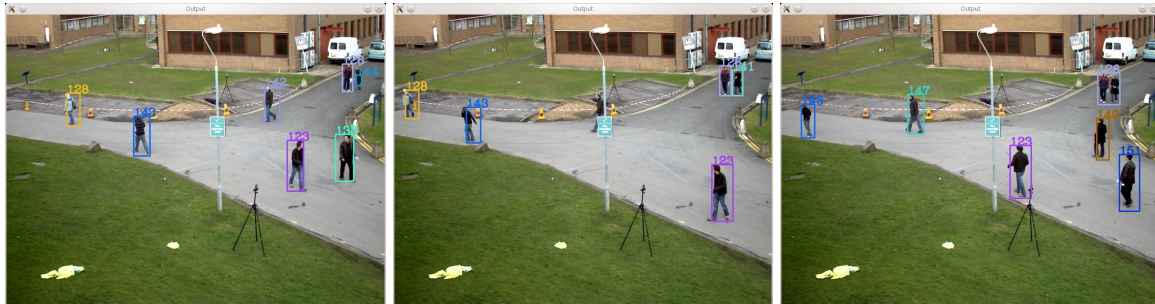

Figure 21. The global tracking challenge : correcting errors due to occlusions (ID 142 on the first frame becomes 147 on the last frame) and tracking people that are leaving the scene and reentering (ID 133 on the first frame becomes 151 on the last frame).

In order to describe a person's tracklet (segment of trajectory), we use a visual signature called Mean Riemannian Covariance Grid and a discriminative method to emphasize the main differences between each tracklet. This step improves the reliability and the accuracy of the results. By computing the distance between the visual signatures, we are able to link tracklets belonging to the same person into a tracklet cluster. Only tuples of tracklets that are not overlapping each other are used as initial candidates. Then, we use Mean Shift to create the clusters.

We evaluated this method on several datasets (i-LIDS, Caviar, PETS 2009). We have shown that our approach can perform as well as the other state of the art methods on Caviar and can perform better on i-LIDS. On PETS 2009 dataset, our approach performs better than standard tracker but cannot be compared with the best state of the art methods due to unadapted metrics.

This approach is described in detail in two articles : one published in ICIP 2012 [35], which is focused on computing the covariance signature and the way to discriminate it and the other one published in PETS 2012 workshop (part of AVSS 2012 conference) [33], which is focused on the method to link the tracklets.

This work will be added to a more general tracking controller that should be able to detect several kinds of detection and tracking errors and try to correct them.

6.10. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition, Contextual Features, Pairwise Features, Relative Tracklets, Spatio-Temporal Interest Points, Tracklets, Head Estimation.

The goal of this work is to automatically recognize human actions and activities in diverse and realistic video settings.

Over the last few years, the bag-of-words approach has become a popular method to represent video actions. However, it only represents a global distribution of features and thus might not be discriminative enough. In particular, the bag-of-words model does not use information about: local density of features, pairwise relations among the features, relative position of features and space-time order of features. Therefore, we propose three new, higher-level feature representations that are based on commonly extracted features (e.g. spatio-temporal interest points used to evaluate the first two feature representations or tracklets used to evaluate the last approach). Our representations are designed to capture information not taken into account by the model, and thus to overcome its limitations.

In the first method, we propose new and complex contextual features that encode spatio-temporal distribution of commonly extracted features. Our feature representation captures not only global statistics of features but also local density of features, pairwise relations among the features and space-time order of local features. Using two benchmark datasets for human action recognition, we demonstrate that our representation enhances the discriminative power of commonly extracted features and improves action recognition performance, achieving 96.16% recognition rate on popular KTH action dataset and 93.33% on challenging ADL dataset. This work has been published in [36].

In the second approach, we design new representation of features encoding statistics of pairwise co-occurring local spatio-temporal features. This representation focuses on pairwise relations among the features. In particular, we introduce the geometric information to the model and associate geometric relations among the features with appearance relations among the features. Despite that local density of features and space-time order of local features are not captured, we are able to achieve similar results on the KTH dataset (96.30% recognition rate) and 82.05% recognition rate on UCF-ARG dataset. An additional advantage of this method is to reduce the processing time of training the model from one week on a PC cluster to one day. This work has been published in [37].

In the third approach, we propose a new feature representation based on point tracklets and a new head estimation algorithm. Our representation captures a global distribution of tracklets and relative positions of tracklet points according to the estimated head position. Our approach has been evaluated on three datasets, including KTH, ADL, and our locally collected Hospital dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc*. Sample frames with extracted tracklets from video sequences of the ADL and Hospital datasets are illustrated on Figure 22. Consistently, experiments show that our representation enhances the discriminative power of tracklet features and improves action recognition performance. This work has been accepted for publication in [38].

Figure 22. Sample frames with extracted tracklets from video sequences of the ADL (left column) and Hospital (right column) datasets.

6.11. 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.

Keywords: events detection, behaviour recognition, automatic video understanding, tracking

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). Analysing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [88] 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 behaviours such as violence or vandalism (alarming events) or a queue at the vending machine (non-alarming events).

The group tracking approach uses Mean-Shift clustering of trajectories to create groups. Two or more individuals are associated in a group if their trajectories have been clustered together by the Mean-Shift algorithm. The trajectories are given by the long-term tracker described in [60]. Each trajectory is composed of a person's positions (x, y) on the ground plane (in 3D) over the time window, and of their speed at each frame in the time window. Positions and speed are normalized using the minimum and maximum possible values (0 and $10m/s$ for the speed and the field of view of the camera for the position). The Mean-Shift algorithm requires a *tolerance* parameter which is set to 0.1, meaning that trajectories need to be distant by less than 10% of the maximum to be grouped.

Figure 23. Example of a group composed of non-similar individual trajectories.

As shown in Figure 23, people in a group might not always have similar trajectories. For this reason, a group is also created when people are very close. A group is described by its coherence, a value calculated from the average distances of group members, their speed similarity and direction similarity. The update phase of the group uses the coherence value. A member will be kept in a group as long as the group coherence is above a threshold. This way, a member can temporarily move apart (for instance to buy a ticket at the vending machine) without being separated from the group.

This work has been applied to the benchmark CAVIAR dataset for testing, using the provided ground truth for evaluation. This dataset is composed of two parts: acted scenes in the Inria hall (9 sequences of 665 frames in average) and not acted recordings from a shopping mall corridor (7 sequences processed of 1722 frames in average). The following scenarios have been defined using the event-description language of [88]: *fighting*, *split up*, *joining*, *shop enter*, *shop exit*, *browsing*. These scenarios have been recognized in the videos with a high success rate (94%). The results of this evaluation and the above described method have been published in [45].

The group tracking algorithm is integrated at both Torino and Paris testing sites and runs in real time on live video streams. The global VANAHEIM system has been presented as a demonstration at the ECCV 2012 conference. A demonstration video has been compiled from the results of the group tracking on 60 sequences from the Paris subway showing interesting groups with various activities such as *waiting*, *walking*, *lost*, *kids* and *lively*.

6.12. Crowd Event Monitoring Using Texture and Motion Analysis

Participants: Vaibhav Katiyar, Jihed Joober, François Brémond.

keywords: Crowd Event, Texture Analysis, GLCM, Optical Flow

The aim of this work is to monitor crowd event using crowd density, change of speed and orientation of group of people. For reducing complexity we are using human density rather than individual human detection and tracking. In this study Human density is quantified mainly into three groups- (1) Empty (2) Sparse (3) Dense. These are approximated by calculating Haralick features from Grey Level Co-occurrence Matrix (GLCM).

We use Optical flow for getting motion information like current speed and orientation of selected FAST feature points. Subsequently we used this information for classifying crowd behaviour into normal or abnormal categories wherein we seek for sudden change in speed or orientation heterogeneity for abnormal behaviour.

In future work this abnormal behaviour may further be classified into different events like Running, Collecting, Dispersion, Stopping/Blocking.

6.13. Detecting Falling People

Participants: Etienne Corvee, Francois Bremond.

keywords: fall, tracking, event

We have developed a people falling algorithm based on our object detection and tracking algorithm [58] and using our ontology based event detector [57]. These algorithms extract moving object trajectories from videos and triggers alarms whenever the people activity fits event models. Most surveillance systems use a multi Gaussian technique [83] to model background scene pixels. This technique is very efficient in detecting in real-time moving objects in scenes captured by a static camera, with low level of shadows, few persons interacting in the scene and with as few as possible illumination changes. This technique does not analyse the content of the moving pixels but simply assign them as foreground or background pixels.

Many state of the art algorithms exist that can recognize objects such as a person human shape, a head, a face or a couch. However, these algorithms are quite time consuming or the database used for training the system is not well adapted to our application domain. For example, people detection algorithms use databases containing thousands of image instances of standing or walking persons taken by camera from a certain distance from the persons and from a facing position. In our indoor monitoring application, cameras are located on the roof with high tilt angle so that most of the scene (e.g.rooms) is viewed. With such camera spatial configuration, the image of a person on the screen rarely corresponds to the person images in the training database. In addition, people are often occluded by the image border (the image of the full body is not available), image distortion needs to be corrected and people often have poses that are not present in the database (e.g. a person bending or sitting).

Using our multi Gaussian technique [74], after having calibrated a camera scene, a detected object is associated with a 3D width and height in two positions : the standing and lying positions. This 3D information is checked against 3D human model and any object is then labelled as either a standing person, a lying person or unknown. Many 3D filtering thresholds are used ; for example, object speed should not be greater than a human possible running speed. Second, we use an ontology based event detector to build a hierarchy of event model complexity. We detect people to have fallen on the floor if the object has been detected as a person on the floor and outside the bed and couch for at least several seconds consecutively. An example of a fallen person is shown in Figure 24.

Figure 24. Detection of a fallen person.

6.14. People Detection Framework

Participants: Srinidhi Mukanahallipatna, Silviu-Tudor Serban, François Brémond.

keywords: LBP, Adaboost, Cascades

We present a new framework called COFROD (Comprehensive Optimization Framework for Real-time Object Detection) for object detection that focuses on improving state of the art accuracy, while maintaining real-time detection speed. The general idea behind our work is to create an efficient environment for developing and analyzing novel or optimized approaches in terms of classification, features, usage of prior knowledge and custom strategies for training and detection. In our approach we opt for a standard linear classifier such as Adaboost. Inspired by the integral channel feature approach, we compute variants of LBP and Haar-like features on multiple channels of the input image. Thus, we obtain an elevated number of computationally inexpensive features that capture substantial information. We use an extensive training technique in order to obtain optimal classifier.

We propose a comprehensive framework for object detection with an intuitive modular design and high emphasis on performance and flexibility. Its components are organized by parent-modules, child-modules and auxiliary-modules. The parent-modules contain several child-modules and focus on a general task such as Training or Detection. Child-modules solve more specific tasks, such as feature extraction, training or testing and in most cases require auxiliary-modules. The later have precise intents, for instance computing a color channel transformation or a feature response.

We present two detection configurations. One relies on a single intensively trained detector and the other as a set of specialist detectors.

Our baseline detector uses cascades in order to speed up the classifier. By removing most false positive at first stages, computation time is significantly reduced. Classifier for each cascade is generated using the training approach.

Our contribution is in the form of a hierarchical design of specialized detectors. At first level we use a version of the baseline detector in order to remove irrelevant candidates. At the second level, specialist detectors are defined. These detectors can be either independent or can use third level detectors and cumulate their output. A specialist detector can take the role of solving an exact classification issue, such as sitting pose. In that case it is trained only with data relevant to that task. In some applications, a specialist detector can be trained to perform exceptionally on a specific situation. In this case training samples are adapted to the particularity of the testing, and possibly parts of the testing sets are used for training.

This is a versatile system for object detection that excels in both accuracy and speed. We present a valuable strategy for training and a hierarchy of specialized people detectors for dealing with difficult scenarios. We also propose an interesting feature channel and a method for loosing less detection speed-up. In our approach we build upon the ideas of feature scaling instead of resizing images and of transferring most computations from detection to training, thus achieving real-time performance on VGA resolution.

Figure 25 and Figure 26 illustrate our detections results. Figure 27 shows the performance of our system compared to other. IDIAP detector was used without tuning the parameters.
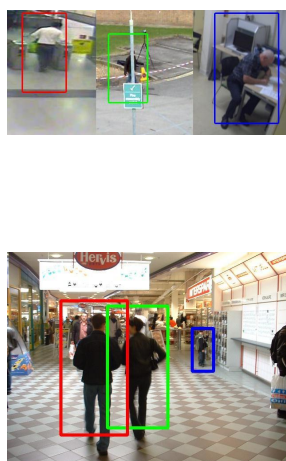

Figure 25. Detection Results

6.15. A Model-based Framework for Activity Recognition of Older People using Multiple sensors

Participants: Carlos -Fernando Crispim Junior, Qiao Ma, Baptiste Fosty, Cintia Corti, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, Multi-sensor Analysis, Surveillance System, Older people, Frailty assessment

We have been investigating a model-based activity recognition framework for the automatic detection of physical activity tests and instrumental activities of daily living (IADL, *e.g.*, preparing coffee, making a phone call) of older people. The activities are modelled using a constraint-based approach (using spatial, temporal, and *a priori* information of the scene), and a generic ontology based on natural terms which allows medical experts to easily modify the defined activity models. Activity models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). The framework has been tested as a system on the clinical protocol developed by the Memory Center of Nice hospital. This clinical protocol aims at studying how ICTs (Information and Communication Technologies)
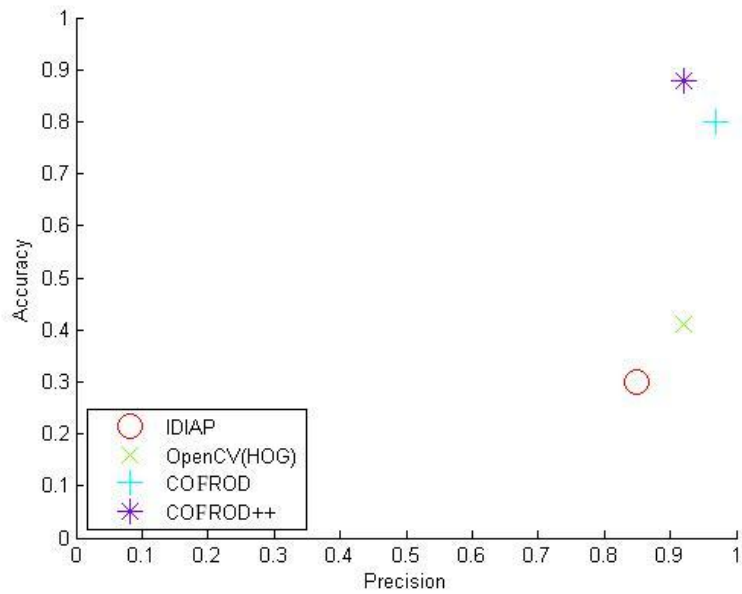
Figure 26. PETS Detection Results

Figure 27. Detection results on the PETS dataset

can provide objective evidence of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI). The Clinical protocol participants are recorded using a RGB video-camera (8 fps), a RGB-D Camera (Kinect - Microsoft), and an inertial sensor (MotionPod) which allows a multi-sensor evaluation of the activities of the participants in an observation room equipped with home appliances. A study of the use of a multi-sensor monitoring for Patient diagnosis using events annotated by experts has been performed in partnership with CHU-Nice and SMILE team of TAIWAN, and it has shown the feasibility of the use of these sensors for patient performance evaluation and differentiation of clinical protocol groups (Alzheimer's disease and healthy participants) [31] and [40]. The multi-sensor evaluation has used the proposed surveillance system prototype and has been able to detect the full set of physical activities of the scenario 1 of the clinical protocol (e.g., Guided a ctivities : Balance test, Repeated Transfer Test), with a true positive rate of 96.9% to 100% for a set of 38 patients (MCI=19, Alzheimer=9) using data of an ambient camera. An extension of the developed framework has been investigated to handle multiple sensors data in the event modeling. In this new scenario, information from the ambient camera and the inertial sensor worn on the participants chest is used (see Figure 28). The prototype using the extended framework has been tested on the automatic detection of IADLs, and preliminary results points to an average sensitivity of 91% and an average precision of 83.5%. This evaluation has been performed for 9 participants videos (15 min each, healthy: 4, MCI: 5). See [39] for more details. Future work will focus on a learning mechanism to automatic fuse events detected by a set of heterogeneous sensors, and at supporting clinicians at the task of studying differences between the activity profile of healthy participants and early to moderate stage Alzheimer's patients.
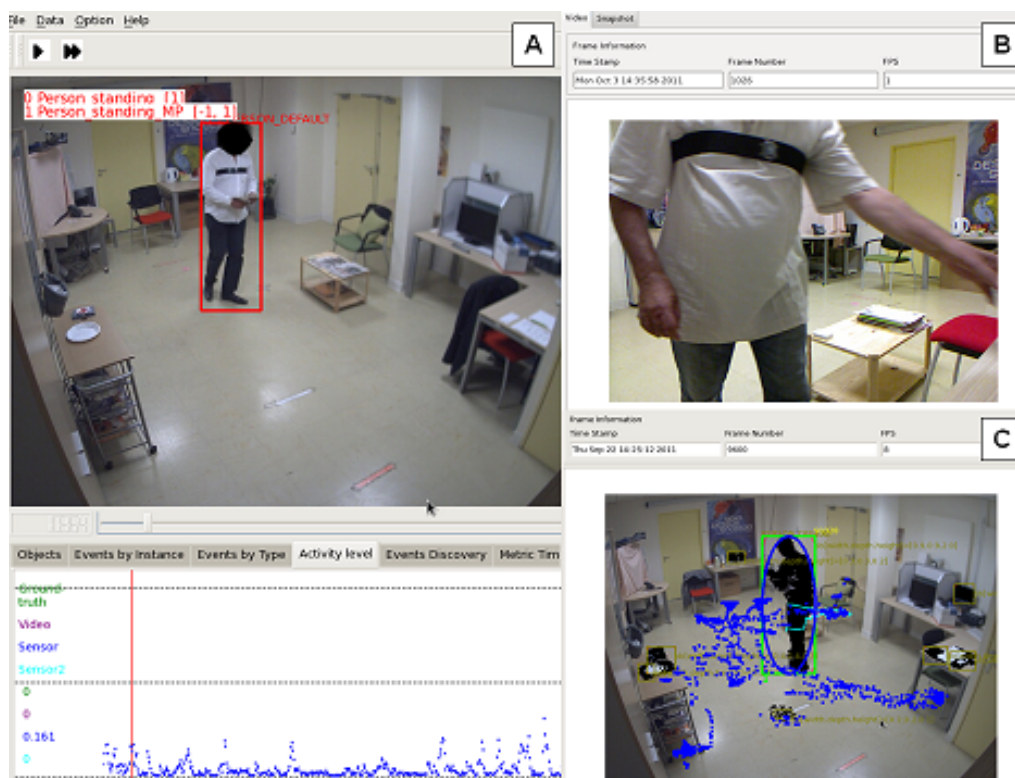

Figure 28. A: Ambient Camera View of Patient Activity. Actimetry captured by the inertial sensor is displayed at the bottom, B: RGB-D Camera View of Patient. The inertial sensor is worn by the patient by an accessory chest strap, C: Trajectory information of Patient Activity during the experimentation.

6.16. Activity Recognition for Older People using Kinect

Participants: Baptiste Fosty, Carlos -Fernando Crispim Junior, Véronique Joumier, Philippe Robert, Alexandra Konig, François Brémond, Monique Thonnat.

keywords: Activity Recognition, RGB-D camera analysis, Surveillance System, Older people, Frailty assessment

Within the context of the Dem@Care project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) from Microsoft (Kinect) for an activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease (AD) and related conditions (like Memory Cognitive Impairment - MCI) for older people. This system is designed on a model-based activity recognition framework. Using a constraint-based approach with contextual and spatio-temporal informations of the scene, we have developed activity models related to the physical activity part of the protocol (Scenario 1, guided activities : balance test, walking test, repeated transfers posture between sitting and standing). These models are organized in a hierarchical structure according to their complexity (Primitive state, Composite State, Primitive Event, and Composite Event). This work is an adaptation of the work performed for multi-sensor analysis [39].

Several steps are needed to adapt the processing. We had for example to generate new ground truth, or we had to design new 3D zones of interest according to Kinect point of view and referential (differing from the 2D camera). Moreover, in order to improve the reliability of the results, we had to solve several issues in the processing chain. For instance, Kinect and the detection algorithm provided by OpenNi and Nestk (free libraries) have several limitations which leads to wrong detection of human. We proposed in these cases several solutions like filtering wrong object detections by size (see Figure29 C) or recomputing the height of older people based on their head when wearing black pants (absorption of infrared) (see Figure 29 D).

For the experimentation, we have processed the data recorded for 30 patients. The results are shown in Figure 30. With a true positive rate of almost 97% and a precision of 94.2%, our system is able to extract most of the activities performed by patients. Then, relevant and objective information can be delivered to clinicians, to assess the patient frailty. For further information on the performance of the detection process, we also generate the results frame by frame, which are shown in Figure 31. We see there that the performance of the event detection in terms of true positive rate is almost as good as by events (94.5%). Nevertheless, if we focus on the precision, it is lower than previously. This means that we still need to improve detection accuracy of the beginning and the end of an event.

Future work will focus on using the human skeleton to extract finest information on the patient activity and to process more scenarios (semi-guided and free).

6.17. Descriptors of Depth-Camera Videos for Alzheimer Symptom Detection

Participants: Guillaume Charpiat, Sorana Capalnean, Bertrand Simon, Baptiste Fosty, Véronique Joumier.

keywords: Kinect, action description, video analysis

In a collaboration with the CHU hospital of Nice, a dataset of videos was recorded, where elderly are asked by doctors to perform a number of predefined exercises (like walking, standing-sitting, equilibrium test), and recorded with an RGBD camera (Kinect). Our task is to analyze the videos and detect automatically early Alzheimer symptoms, through statistical learning. Here we focus on the 3D depth sensor (no use of the RGB image), and aim at providing action descriptors that are accurate enough to be informative.

During her internship in the Stars team, Sorana Capalnean proposed descriptors relying directly on the 3D points of the scene. First, based on trajectory analysis, she proposed a way to recognize the different physical exercises. Then she proposed, for each exercise, specific descriptors aiming at providing the information asked by doctors, such as step length, frequency and asymmetry for the walking exercise, or sitting speed and acceleration for the second exercise, etc. Problems to deal with included the high level of noise in the 3D cloud of points given by the Kinect, as well as an accurate localization of the floor.
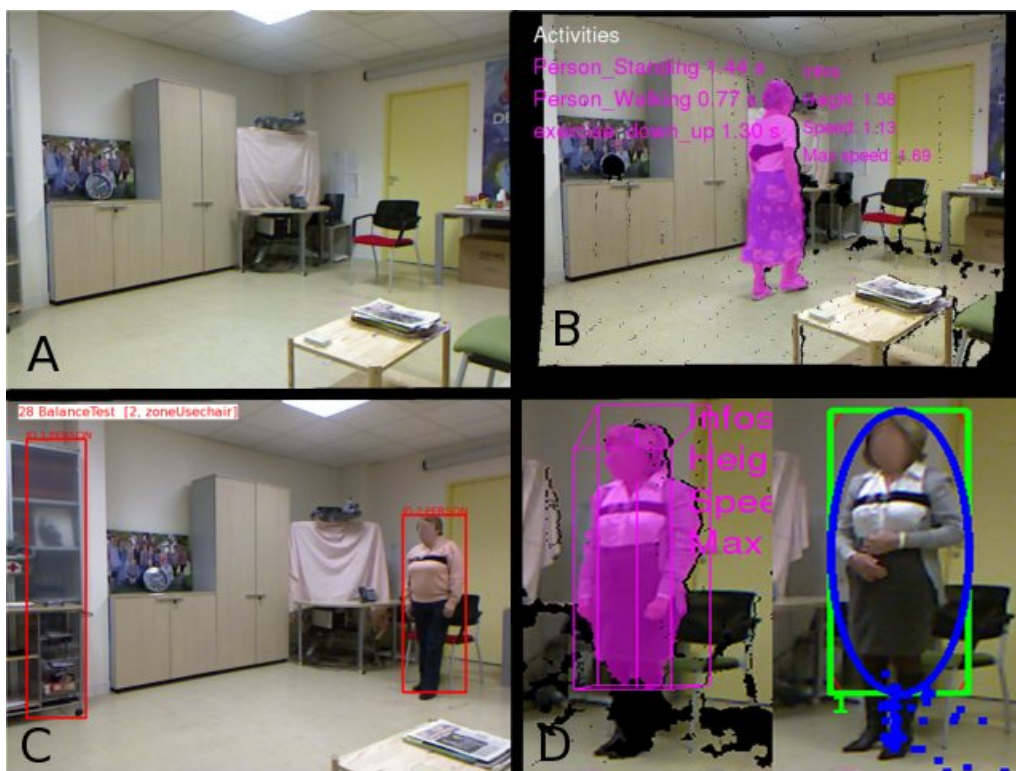
Figure 29. A: RGB-D camera view of the scene, B: 3D representation of the scene with some event detection, C: people detection problem (furniture detected as extra person), D: people detection problem (black clothes not detected).

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	30	VP	30	100	100	100
		FP	0			
		FN	0			
WalkingTest_firstAttempt	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
WalkingTest_secondAttempt	30	TP	27	90	100	94,7
		FP	0			
		FN	3			
RepeatedTransfersTest	30	TP	30	100	90,9	95,2
		FP	3			
		FN	0			
UpAndGoTest	30	TP	28	93,3	90,3	91,8
		FP	3			
		FN	2			
TOTAL	**150**	**TP**	**145**	**96,6**	**94,2**	**95,4**
		FP	**9**			
		FN	**5**			

Figure 30. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

	Nb GT	Detected		Sensitivity (%)	Precision (%)	Fscore (%)
BalanceTest	37235	TP	37216	99,9	70,8	82,9
		FP	15322			
		FN	19			
WalkingTest_firstAttempt	1537	TP	849	55,2	94,1	69,6
		FP	53			
		FN	688			
WalkingTest_secondAttempt	1425	TP	857	60,1	62,6	61,3
		FP	513			
		FN	568			
RepeatedTransfersTest	5190	TP	4496	86,6	94,6	90,4
		FP	257			
		FN	694			
UpAndGoTest	3345	TP	2670	79,8	86,2	82,9
		FP	428			
		FN	675			
TOTAL	**48732**	**TP**	**46088**	**94,5**	**73,6**	**82,8**
		FP	**16573**			
		FN	**2644**			

Figure 31. Results by events (GT = ground truth, TP = true positive, FP = false positive, FN = false negative)

During his internship, Bertrand Simon proposed other kinds of descriptors, based on the articulations of the human skeleton given by OpenNI. These articulations are however very noisy too, so that a pre-filtering step of the data in time had to be performed. Various coordinate systems were studied, to reach the highest robustness. The work focused not only on descriptors but also on metrics suitable to compare gestures (in the phase space as well as in the space of trajectories). See figure 32 for an example.

These descriptors are designed to be robust to camera noise and to extract the relevant information from the videos; however their statistical analysis still remains to be done, to recognize Alzheimer symptoms during the different exercises.
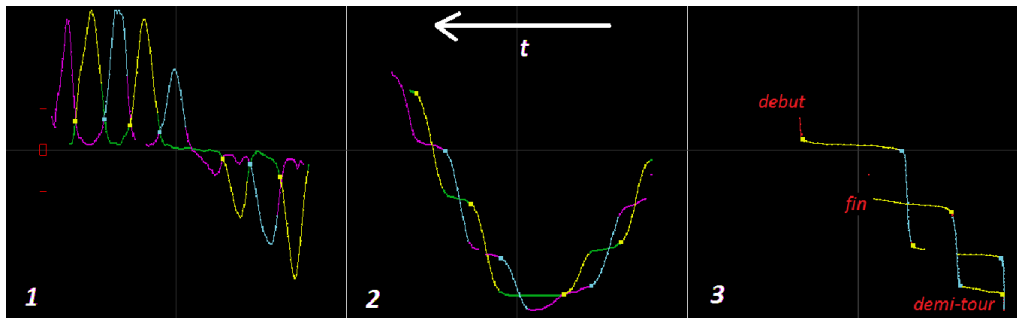

Figure 32. Curves obtained during a person's walk in backward then forward directions. Purple and blue curves stand for the right foot, while green and yellow ones stand for the left one. Graph 1 shows the speeds as a function of time; Graph 2 shows the locations as a function of time; Graph 3 shows the location of the right foot as a function of the location of the left foot.

6.18. Online Activity Learning from Subway Surveillance Videos

Participants: Jose-Luis Patino Vilchis, Abhineshwar Tomar, François Brémond, Monique Thonnat.

Keywords: Activity learning, clustering, trajectory analysis, subway surveillance

This work provides a new method for activity learning from subway surveillance videos. This is achieved by learning the main activity zones in the observed scene by taking as input the trajectories of detected mobile objects. This provides us the information on the occupancy of the different areas of the scene. In a second step, these learned zones are employed to extract people activities by relating mobile trajectories to the learned zones, in this way, the activity of a person can be summarised as the series of zones that the person has visited. If the person resides in the single zone this activity is also classified as a standing. For the analysis of the trajectory, a multiresolution analysis is set such that a trajectory is segmented into a series of tracklets based on changing speed points thus extracting the information when people stop to interact with elements of the scene or other people. Starting and ending tracklet points are fed to an advantageous incremental clustering algorithm to create an initial partition of the scene. Similarity relations between resulting clusters are modelled employing fuzzy relations. A clustering algorithm based on the transitive closure calculation of the fuzzy relations easily builds the final structure of the scene. To allow for incremental learning and update of activity zones (and thus people activities), fuzzy relations are defined with online learning terms. The approach is tested on the extraction of activities from the video recorded at one entrance hall in the Torino (Italy) underground system. Figure 33 presents the learned zones corresponding to the analyzed video. To test the validity of the activity extraction a one hour video was annotated with activities (corresponding to each trajectory) according to user defined ground-truth zones. After the comparison, following results were obtained: TP:26, FP:3, FN:1, Precision:0.89, Sensitivity:0.96. This work is published in [43].
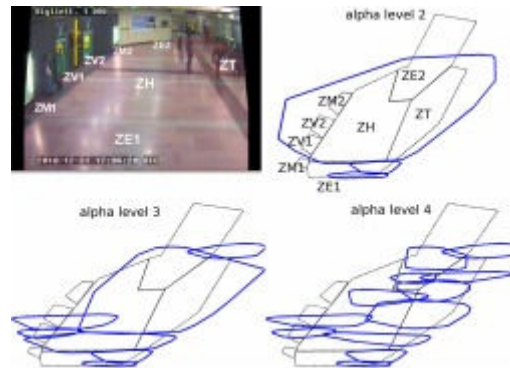
Figure 33. Left top panel: Original underground scene observed by the camera with user-defined areas delimiting the scene. Remaining panels: Learned zones in a 3D top view. They correspond to activity areas as discovered with our algorithm. Different granularity levels allow understanding the activity with different resolutions

6.19. Automatic Activity Detection Modeling and Recognition: ADMR

Participants: Guido-Tomas Pusiol, François Brémond.

This year a new Ph.D. thesis has been defended [30]. The main objective of the thesis is to propose a complete framework for the automatic activity discovery, modeling and recognition using video information. The framework uses perceptual information (e.g. trajectories) as input and goes up to activities (semantics). The framework is divided into five main parts:

1. We break the video into chunks to characterize activities. We propose different techniques to extract perceptual features from the chunks. This way, we build packages of perceptual features capable of describing activity occurring in small periods of time.

2. We propose to learn the video contextual information. We build scene models by learning salient perceptual features. The models end up containing interesting scene regions capable of describing basic semantics (i.e. region where interactions occur).

3. We propose to reduce the gap between low-level vision information and semantic interpretation, by building an intermediate layer composed of Primitive Events. The proposed representation for primitive events aims at describing the meaningful motions over the scene. This is achieved by abstracting perceptual features using contextual information in an unsupervised manner.

4. We propose a pattern-based method to discover activities at multiple resolutions (i.e. activities and sub-activities). Also, we propose a generative method to model multi-resolution activities. The models are built as a flexible probabilistic framework easy to update.

5. We propose an activity recognition method that finds in a deterministic manner the occurrences of modelled activities in unseen datasets. Semantics are provided by the method under user interaction. All this research work has been evaluated using real datasets of people living in an apartment (home-care application) and elder patients in a hospital.

The work has also been evaluated for other types of applications such as sleeping monitoring. For example, Figure 34 display the results of the activity discovery method during 6 hours (left to right) applied to the center of mass (3D) of a tracked sleeping person. The colored segments represent hierarchical (bottom-up is finer-coarse) discovered activity which matches with sleeping postural movements. The segments have similar color when postural movements are similar. For example, the segment (j) is the only time the person sleeps upside down. Also, health professionals analysed the results claiming that the segments corresponds to normal

sleeping cycle, where low motion is noticed at the beginning of the sleep and more motion is shown when the person have a lighter sleep when starts waking up.
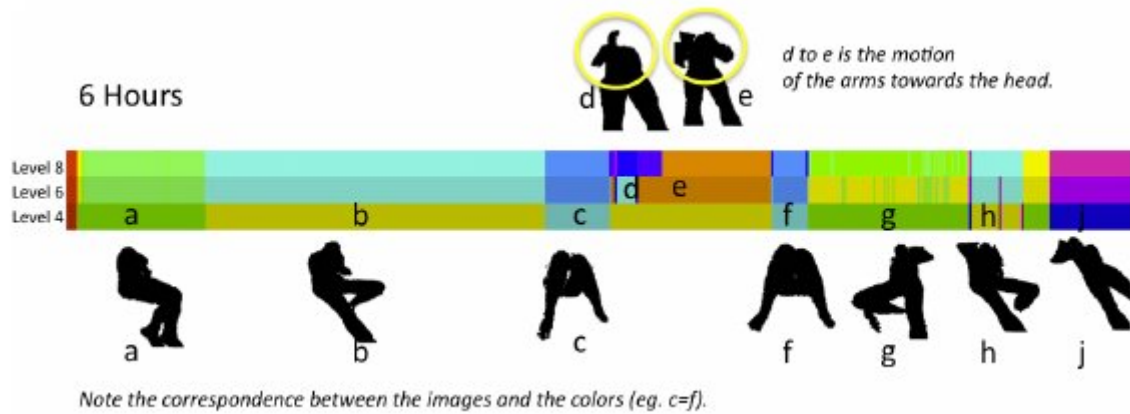

Figure 34. Results of the activity discovery method during 6 hours

6.20. SUP Software Platform

Participants: Julien Gueytat, Baptiste Fosty, Anh tuan Nghiem, Leonardo Rocha, François Brémond.

Our team focuses on developing Scene Understanding Platform (SUP) (see section 5.1). This platform has been designed for analyzing a video content. SUP is able to recognize simple events such as 'falling', 'walking' of a person. We can easily build new analyzing system thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, ... All those advantages are available since we are working together with the software developers team DREAM. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.
Currently, we have fully integrated OpenCV library with SUP and the next step is to integrate OpenNI to get depth map processing algorithms from PrimeSense running in SUP. Updates and presentations of our framework can be found on our team website http://team.inria.fr/stars. Detailed tips for users are given on our Wiki website http://wiki.inria.fr/stars and sources are hosted thanks to the new Source Control Management tool.

6.21. Qualitative Evaluation of Detection and Tracking Performance

Participants: Swaminathan Sankaranarayanan, François Brémond.

We study an evaluation approach for detection and tracking systems. Given an algorithm that detects people and simultaneously tracks them, we evaluate its output by considering the complexity of the input scene. Some videos used for the evaluation are recorded using the Kinect sensor which provides for an automated ground truth acquisition system. To analyse the algorithm performance, a number of reasons due to which an algorithm might fail is investigated and quantified over the entire video sequence. A set of features called Scene Complexity measures are obtained for each input frame. The variability in the algorithm performance is modelled by these complexity measures using a polynomial regression model. From the regression statistics, we show that we can compare the performance of two different algorithms and also quantify the relative influence of the scene complexity measures on a given algorithm. This work has been published in [44].

6.22. Model-Driven Engineering and Video-surveillance

Participants: Sabine Moisan, Jean-Paul Rigault, Luis-Emiliano Sanchez.

keywords: Feature Model Optimization, Software Metrics, Requirement specification, Component-based system, Dynamic Adaptive Systems, Model-Driven Engineering, Heuristic Search, Constraint Satisfaction Problems

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks and the video analysis algorithms [41]. The various VS tasks (counting, intrusion detection, tracking, scenario recognition) have different requirements. Observation conditions, objects of interest, device configuration... may vary from one application to another. On the implementation side, selecting the components themselves, assembling them, and tuning their parameters to comply with context may lead to a multitude of variants. Moreover, the context is not fixed, it evolves dynamically and requires run time adaptation of the component assembly.

Our work relies on Feature Models, a well-known formalism to represent variability in software systems. This year we have focused on an architecture for run time adaptation and on metrics to drive dynamic architecture changes.

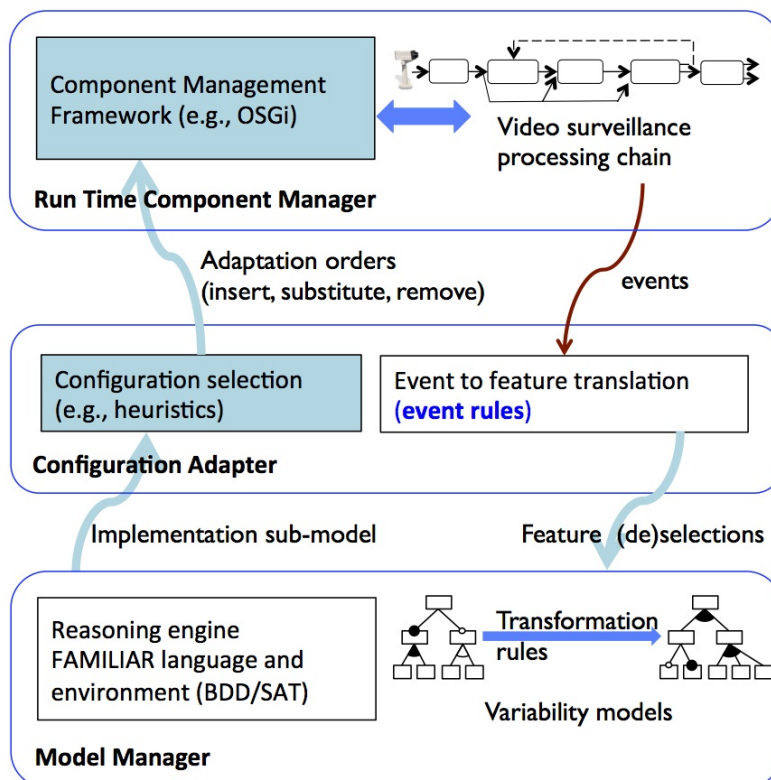### 6.22.1. Run Time Adaptation Architecture

Figure 35. Architecture of an adaptive system using models at run time. (In light blue the elements studied this year.)

The architecture of the run time system (also used for initialization at deployment time) is based on three collaborating modules as shown in Figure 35. A *Run Time Component Manager* (RTCM) cooperates with the low levels (to manage the software components and capture events) and applies configuration changes. A *Configuration Adapter* (CA) receives events from the RTCM, and propagates them as features into the models to obtain a new configuration. The *Model Manager* (MM) embeds a specialized scripting language for Feature Models (FAMILIAR [52], [53][1]) to manage the representation of the two specialized feature models and applies constraints and model transformations on them. The *Model Manager* produces new component configurations (a model specialization) that it sends to the CA. At its turn, the CA selects one single configuration (possibly using heuristics) and converts it into component operations to be applied by the RTCM.

This year we first finalized the interface between the Model Manager and the Configuration Adapter. On one hand, we transform the feature models obtained from FAMILIAR into C++ representations enriched with software component information. On the other hand, we dynamically transform context change events into requests to FAMILIAR.

Second, we searched for a suitable technology for handling components in the Run Time Component Manager. OSGi is an adequate de facto standard but it is mainly available in the Java world. However we could find a C++ implementation, complete enough for our needs (SOF, Service Oriented Framework [65]). However, SOF has to be completed to adjust to the needs of our end users who are the video system developers. Thus, we are currently building a multi-threaded service layer on top of SOF, easy to use and hiding most of the nitty-gritty technical details of thread programming and SOF component manipulation. This layer provides end users with a set of simple patterns and allow them to concentrate only on the code of video services (such as acquisition, segmentation, tracking...).

As a matter of feasability study we are building an experimental video self adaptive system based on the afore mentionned architecture. Software components are implemented with the OpenCV library. In the final system, feature models and software components continuously interact in real time, modifying the whole system in response to changes in its environment.

6.22.2. *Metrics on Feature Models to Optimize Configuration Adaptation at Run Time*

As shown on figure 35, the Configuration Adapter has to set up a suitable component configuration of the run time system. For this, each time the context changes, it receives a set of valid configurations (a feature *sub*-model) from the Model Manager. In most cases, this set contains more than one configuration. Of course, only one configuration can be applied at a given time and the problem is to select the "best" one. Here, "best" is a trade-off between several non-functional aspects: performance, quality of service, time cost for replacing the current configuration, etc.

It is thus necessary to rank the configurations. Our approach is to define metrics suitable for comparing configurations. Then the problem comes down to the widely studied problem of *Feature model optimization* [55]. This problem is known to be an intractable combinatorial optimization problem in general.

We started with a study of the state of the art: metrics for general graphs as well specific to feature models, optimization and requirement specification on feature models... We obtained a structured catalog of quality and feature model metrics. Then we selected solutions based on heuristic search algorithms using quality and feature model metrics. We thus propose several strategies and heuristics offering different properties regarding optimality of results and execution efficiency.

These strategies and heuristics have been implemented, tested, and analyzed using random generated feature models. We got empirical measures about their properties, such as completeness, optimality, time and memory efficiency, scalability... This allows us to compare the performance of the different algorithms and heuristics, and to combine them in order to achieve a good trade-off between optimality and efficiency. Finally, the proposed algorithms have been introduced as part of the Configuration Adapter module.

---

[1]FAMILIAR has been developed at the I3S laboratory by the Modalis team.

This work is quite original from several aspects. First, we did not find any study using heuristic search algorithms for solving the feature optimization problem. Most studies apply Artificial Intelligence techniques such as CSP solvers, planning agents, genetic algorithms... Second, we do not restrict to the optimization of linear objective functions, but we also address non-linear ones allowing us to take into account a broader set of criteria. Among the possible criteria we consider quality of service of components, their performance, their set up delay, the cost of their replacement, etc. Finally, we apply our metrics at run time whereas most studies consider metrics only for static analysis of feature models.

Currently, we are still working on new variants of the search algorithms and new heuristics relying on techniques proposed in the domains of heuristic search and constraint satisfaction problems.

6.23. Synchronous Modelling and Activity Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Daniel Gaffé.

6.23.1. Scenario Analysis Module (SAM)

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 other 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 [76] and we adapt usual techniques of synchronous modelling approach to express scenario behaviours. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model.

In addition, 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. To face this challenge, we introduced in the generation of 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. Indeed, we relied on Lustre [66] synchronous language to express the automata semantics of scenario models as Boolean equation systems. This approach was successful and shows that we can consider a synchronous framework to generate validated scenario recognition engines. This year, in order to improve efficiency (and to tackle the real time recognition problem), we begin to rely on CLEM (see section 6.23.2) toolkit to generate such recognition engines. The reason is threefold: (1) CLEM is becoming a mature synchronous programming environment; (2) we can use the CLEM compiler to build our own compiler; (3) CLEM supplies the possibility of using NuSMV [61] model checker, which is more powerful than the Lustre model-checker. Moreover, thanks to CLEM compiler into Boolean equation systems, we can compute the expected events of the next instant on the fly, by propagation of information related to the current instant.

6.23.2. The clem Workflow

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 9) 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. This year, we mainly work on theoretical aspects of CLEM.

First, synchronous language semantics usually characterizes each output and local signal status (as present or absent) according to input signal status. To reach our goal, we defined a semantics that translates LE programs into equation systems. This semantics bears and grows richer the knowledge about signals and is never in contradiction with previous deduction (this property is called constructiveness). In such an approach, causality turns out to be a scheduling evaluation problem. We need to determine all the partial orders of equation systems and to compute them, we consider a 4-valued algebra to characterize the knowledge of signal status (unknown, present, absent, overknown). Previously, we relied on 4-valued Boolean algebra [19], [20] which defines the negation of unknown as overknown. The advantage of this way is to benefit from Boolean algebras laws to compute equation system solutions. The drawback concerns signal status evaluation which does not correspond to usual interpretation (not unknown = unknown and not overknown = overknown). To avoid this drawback, we study other kinds of algebras well suited to define synchronous languages semantics. In [49], we choose an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [64]. But, the algebra we defined is no more a Boolean algebra, but we prove (always in [49]), that the main laws of Boolean algebras hold as distributivity laws, associativity laws, idempotence laws, etc. After compilation, signals have to be projected into Boolean values. Bilattice theory offers an isomorphism between 4-valued status and pair of Boolean.

Second, the algorithm which computes partial orders relies 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. Inputs (resp. outputs) have date 0 and the algorithm recursively increases the dates of nodes in the upstream (resp downstream) dependencies graph. Hence, the algorithm determines an *earliest* date and a *latest* date for equation system variables. Moreover, we can compute the dates of variables of a global equation system starting from dates already computed for variables which were inputs and outputs in a sub equation system corresponding to a sub program of the global program[2]. This way of compiling is the corner stone of our approach [20]. We defined two approaches to compute all the valid partial orders of equation systems, either applying critical path scheduling technique (CPM) [3] or applying fix point theory: the vector of earliest (resp. latest) dates can be computed as the least fix point of a monotonic increasing function. This year we have proved that we can compute dates either starting from a global equation system or considering equation system where some variables are abstracted (i.e they have no definition) and whose dates have been already computed. To achieve the demonstration, we rely on an algebraic characterization of dates and thanks to uniqueness property of least fix points, we can deduce that the result is the same for a global equation systems as for its abstraction. We are in the process of publishing this result. From an implementation point of view, we use CPM approach to implement our scheduling algorithm since it is more efficient than fix point consideration. Of course both ways yield the same result. Indeed, fix point approach is useful for a theoretical concern.

6.23.3. Multiple Services for Device Adaptive Platform for Scenario Recognition

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 5.1) 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 [85], [84], [68]) 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 behaviour models associated with a well sound modelling 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 behavioural 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.

---

[2]these variables are local in the global equation system
[3] http://pmbook.ce.cmu.edu/10_Fundamental_Scheduling_Procedures.html

To define such synchronous component, user can specify a synchronous component per sub assembly corresponding to a concern and compose the synchronous components connected with the same critical component in order to get an only synchronous component. Thus, we supply a *composition under constraints* of synchronous components and we proved that this operation preserves already separately verified properties of synchronous components [79], [78].

The main challenge of this approach 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 other, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To tackle this problem, we replace the effective description of constraints by a generic specification of them in the critical component. But, we must offer a way to express these generic constraints. Then, each synchronous component has a synchronous controller, which is the projection of the generic constraints on its output set. The global synchronous component is the synchronous parallel composition of all basic components and their synchronous controllers. Moreover, according to synchronous parallel composition features, the property preservation result we have still hold.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. Collaborations

- Stars has a strong collaboration with the CobTek team (CHU Nice).
- G. Charpiat works with Yuliya Tarabalka (AYIN team) and with Bjoern Menze (Computer Vision Laboratory at ETH Zurich, Medical Vision group of CSAIL at MIT, and collaborator of Asclepios team) on the topic of shape growth/shrinkage enforcement for the segmentation of time series.
- G. Charpiat worked with former members from the ARIANA team: Ahmed Gamal Eldin (now LEAR team), Xavier Descombes (MORPHEME team) and Josiane Zerubia (AYIN team) on the topic of multiple object detection.

7.2. National Initiatives

7.2.1. ANR

7.2.1.1. 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.

7.2.1.2. SWEET-HOME

Program: ANR Tecsan

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.

7.2.2. FUI

7.2.2.1. QUASPER

Program: FUI

Project acronym: QUASPER

Project title: QUAlification et certification des Systèmes de PERception

Duration: June 2010 - May 2012

Coordinator: THALES ThereSIS

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.

7.2.3. Investment of future

7.2.3.1. Az@GAME

Program: DGCIS

Project acronym: Az@GAME

Project title: un outil d'aide au diagnostic médical sur l'évolution de la maladie d'Alzheimer et les pathologies assimilées.

Duration: January 2012- December 2015

Coordinator: Groupe Genious

Other partners: IDATE, Inria(Stars), CMRR (CHU Nice) and CobTek team.

See also: http://www.azagame.fr/

Abstract: This French project aims at providing evidence concerning the interest of serious games to design non pharmacological approaches to prevent dementia patients from behavioural disturbances, most particularly for the stimulation of apathy.

7.2.4. Large Scale Inria Initiative

7.2.4.1. PAL

Program: Inria

Project acronym: PAL

Project title: Personally Assisted Living

Duration: 2010 -2014

Coordinator: COPRIN team

Other 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.

7.2.5. Collaborations

- G. Charpiat works with Gabriel Peyré, François-Xavier Vialard and Giacomo Nardi (CNRS, CERE-MADE, Université Paris-Dauphine) on the topic of piecewise rigid movements.

- G. Charpiat works with Yann Ollivier (Computer Science department in Paris-Sud University (Orsay)), on the topic of image compression.

7.3. European Initiatives

7.3.1. FP7 Projects

7.3.1.1. PANORAMA

Title: PANORAMA

Duration: April 2012 - March 2015

Coordinator: Philips Healthcare (Netherlands)

Other partners :Medisys (France), Grass Valley (Netherlands), Bosch Security Systems (Netherlands), STMicroelectronics (France), Thales Angenieux (France), CapnaDST (UK), CMOSIS (Belgium), CycloMedia (Netherlands), Q-Free (Netherlands), TU Eindhoven (Netherlands) , University of Leeds (UK), University of Catania (Italy), Inria(France), ARMINES (France), IBBT (Belgium).

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

Abstract: PANORAMA aims to research, develop and demonstrate generic breakthrough technologies and hardware architectures for a broad range of imaging applications. For example, object segmentation is a basic building block of many intermediate and low level image analysis methods. In broadcast applications, segmentation can find people's faces and optimize exposure, noise reduction and color processing for those faces; even more importantly, in a multi-camera set-up these imaging parameters can then be optimized to provide a consistent display of faces (e.g., matching colors) or other regions of interest. PANORAMA will deliver solutions for applications in medical imaging, broadcasting systems and security & surveillance, all of which face similar challenging issues in the real time handling and processing of large volumes of image data. These solutions require the development of imaging sensors with higher resolutions and new pixel architectures. Furthermore, integrated high performance computing hardware will be needed to allow for the real time image processing and system control. The related ENIAC work program domains and Grand Challenges are Health and Ageing Society - Hospital Healthcare, Communication & Digital Lifestyles - Evolution to a digital lifestyle and Safety & Security - GC Consumers and Citizens security.

7.3.1.2. 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)

Other 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.

7.3.1.3. SUPPORT

Title: Security UPgrade for PORTs

Type: COOPERATION (SECURITE)

Instrument: IP

Duration: July 2010 - June 2014

Coordinator: BMT Group (UK)

Other 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.

7.3.1.4. 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)

Other 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.

7.3.2. Collaborations in European Programs, except FP7

7.3.2.1. 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.

7.4. International Initiatives

7.4.1. Inria International Partners

7.4.1.1. Collaborations with Asia

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

7.4.1.2. Collaboration with U.S.

Stars collaborates with the University of Southern California.

7.4.1.3. Collaboration with Europe

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

7.4.2. Participation In International Programs

7.4.2.1. 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.

Stars is involved in the EIT ICT Labs - Health and Wellbeing .

7.5. International Research Visitors

7.5.1. Visits of International Scientists

7.5.1.1. Internships

This year Stars has hosted 12 internships:

- Pierre Aittahar, Nice University.
- Guillaume Barbe, Nice University.
- Sorana Capalnean, Cluj-Napoca University.
- Cintia Corti, FCEIA Facultad de Ciencias Exactas Ingenieria y Agrimensura, National University of Rosario.
- Eben Freeman, MIT USA.
- Vaibhav Katiyar, Asian Institute of Technology Khlong Luang Pathumtani, Thailand.
- Vannara Loch, Nice University.
- Qioa Ma, Ecole centrale de Pékin, University of Beihang (China).
- Firat Ozemir, Sabancı Universitesi Orta Mahalle, University Caddesi Istanbul.
- Luis Sanchez, Buenos Aires University.
- Abhineshwar Tomar , Ku Leuven University, Belgium.
- Swaminathan Sankaranarayanan, Delft University of Technology.

8. Dissemination

8.1. Scientific Animation

8.1.1. Conference Organization

- In the framework of VANAHEIM project, Stars has organized a summer school which was held at Inria in October 2012, entitled "Human Activity and Vision Summer School"[4]. This summer school addressed the human activity or behaviour recognition, focusing dominantly on the video and audio modalities. In this context, the topics addressed ranged from low-level feature extraction (background subtraction, space-time interest points, tracklets) to active learning, as well as object detection (human, body), tracking (multi-object, multi-camera, audio-visual), behaviour cue extraction (body or head pose), crowd monitoring and supervised behaviour recognition (statistical and symbolic approaches). The summer school counted 26 outside participants, 19 Inria participants and 21 invited speakers. Most of the participants were PhD students, but master students and postdoctoral researchers were also registered.

8.1.2. Journals

- G. Charpiat reviewed for the journals TIP (Transactions on Image Processing), SIIMS (SIAM Journal on Imaging Sciences), RIA (Revue d'Intelligence Artificielle)
- Jean-Paul Rigault reviewed for the SoSym (Software and System Modeling) journal (Springer)
- M. Thonnat reviewed for Image and Vision Computing Journal
- F. Bremond was reviewer for the International Journal of Neural Systems and for the IEEE Pervasive Computing

8.1.3. Conferences

---

[4]http://www.multitel.be/events/human-activity-and-vision-summer-school/home.php

- G. Charpiat reviewed for the conferences CVPR (Computer Vision and Pattern Recognition) and SIGPRO (Signal Processing).

- Jean-Paul Rigault is a member of AITO (Association Internationale pour les Technologies Objets), the steering committee of several international conferences including ECOOP.

- M. Thonnat reviewed for ICPR Conference

- The conference paper entitled "Alzheimer's patient activity assessment using different sensors" by C. Crispim et all has received the Best Paper Award in ISG*ISARC 2012 conference in Eindhoven, Netherlands in June 2012.

- F. Bremond was Session chair and chairman of discussion panels at AVSS'12 conference

- F. Bremond was reviewer for the conferences CVPR'12-13, ICPR'12, ECCV'12, AVSS'12, PETS'12-13, Visage'12, IROS'12 Workshop, ICRA'13

- F. Bremond was program committee member of the CVPR 3nd Intl. Workshop on Socially Intelligent Surveillance and Monitoring (SISM 2012)

- F. Bremond was program committee member of the 2 ECCV 2012 workshops, International Workshop on Re-Identification (Re-Id 2012) and ARTEMIS 2012

- F. Bremond was program committee member of the IEEE Workshop on Applications of Computer Vision (WACV 2013)

- F. Bremond was program committee member of the Workshop Interdisciplinaire sur la Sécurité Globale (WISG 2013)

8.1.4. Invited Talk

- F. Bremond was invited at the Gerontological Workshop organized by Universitas Tarumanagara in Jakarta on the 28-29 September 2012.

- F. Bremond was invited at the French American Biotech Symposium FABS 2012, invited by EUROBIOMED and French Embassy in US, Nice on the 25-26 October 2012,

- F. Bremond was invited at the IA (Innovation Alzheimer) Workshop 2012: Intersection between ICT & Health – defining guidelines on October 30th in Monaco, France.

- F. Bremond was invited at the International Workshop on Human Behavior Understanding (HBU'2012) held in conjunction with IROS'2012 in Algarve, Portugal, October 2012.

8.1.5. Advisory Board

- 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 "Santé", Dependance et Autonomie since 2010.

- M. Thonnat is Scientific Board member of Ecole Nationale des Ponts since 2008

- F. Bremond is Scientific Board member of Fondation Médéric Alzheimer: European Dementia Biomedical Outlook, April and October 2012.

- F. Bremond is Scientific Board member of Workshop Interdisciplinaire sur la Sécurité Globale (WISG)

- F. Bremond is Member of the advisory Boards of the "Éthique, technologie et maladie d'Alzheimer ou apparentée" EREMA

8.1.6. Expertise

- M. Thonnat participated to the evaluation of ANR Tecsan proposals, award committee member for the best scientific project at Ecole Polytechnique Paris

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Master : François Brémond, Video Understanding Techniques at the Human Activity and Vision Summer school, Sophia-Antipolis, 3h., Oct 2012, FR;

Master : Annie Ressouche, Critical Systems and Vérification. Application to WComp Platform, 10h, M2, Polytechnic School of Nice Sophia Antipolis University, FR;

Jean-Paul Rigault is Full Professor of Computer Science at Polytech'Nice (University of Nice): courses on C++ (beginners and advanced), C, System Programming, Software Modeling.

8.2.2. Supervision

PhD & HdR

PhD : Slawomir Bak, People Detection in Temporal Video Sequences by Defining a Generic Visual Signature of Individuals, Nice Sophia Antipolis University, 5th July 2012, François Brémond [28]

PhD : Duc Phu Chau, Object Tracking for Activity Recognition, Nice Sophia Antipolis University, 30th March 2012 , François Brémond and Monique Thonnat [29];

PhD : Guido-Tomas Pusiol, Learning Techniques for Video Understanding, Nice Sophia Antipolis University, 31st May 2012, François Brémond [30];

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

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

PhD in progress : Carolina Garate, Video Understanding for Group Behaviour 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 : 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 ;

8.2.3. Juries

- G. Charpiat reviewed applications to a Cordi S PhD grant as well as a proposal to the project call Digiteo.

- M.Thonnat was jury member for the PhD defence of Edouard Auvinet, Montreal University, the 14th June 2012.

- A. Ressouche was jury member for the PhD defence of Salma Zouaoui-Elloumi, Mines Paris Tech, in July 2012.

- F. Bremond was jury member for the PhD defence of Cyrille Migniot, GIPSA Lab, Univ. P. Mendes-France, Grenoble, 17 January 2012.

- F. Bremond was jury member for the PhD defence of Carmelo Velardo, EURECOM, 23rd April 2012.

- F. Bremond was jury member for the Mid-term PhD defence of Usman Farrokh Niaz, EURECOM, 2nd May 2012.

- F. Bremond was jury member for the Mid-term PhD defence of Claudiu Tanase, Mid-term defense, EURECOM, 2nd May 2012.

- F. Bremond was jury member for the Mid-term PhD defence of Alban Meffre, Mid-term defense, Télécom Physique Strasbourg (ex ENSPS) - LSIIT CNRS, June 5 2012.

- F. Bremond was jury member for the Mid-term PhD defence of Hajer Fradi, Mid-term defense, EURECOM, 29th October 2012.
- F. Bremond was jury member for the PhD defence of Sun Lin, Université Pierre & Marie Curie (UPMC) - TELECOM SudParis, 12th Dec 2012.
- F. Bremond was jury member for the HDR defence of Christian Wolf, Université de Lyon, INSA-Lyon, LIRIS CNRS, Team Imagine and M2Disco, 10th December 2012.

8.3. Popularization

- 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) and attended a media training.

8.3.1. Press Release

- F. Bremond has a Press interview on Assisted Living in February 2012 with Le Monde, Le Figaro, Les Echos, Notre Temps, Le Quotidien du Médecin, RFI, France Culture, La Croix, France Inter.
- Stars article in Le Monde on the 10th of March on "Vidéosurveillance : Trop de caméras pas assez d'yeux ?",
- F. Bremond has an Interview on Assisted Living on the 21st of March 2012 with France Inter,
- F. Bremond has an Interview on CobTek for the ITV Inria Web Site (by TEchnoscope company) in April 2012.
- Dem@Care has been mentioned in the EU eHealth Newsletters, firstly in March 2012 to announce the launch of the project, secondly in August 2012 to announce the IA (Innovation Alzheimer) Workshop 2012.

9. Bibliography

Major publications by the team in recent years

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

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

[3] B. BOULAY, F. BREMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System.*, in "Pattern Recognition Letter.", 2006, vol. 27, n⁰ 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. *Learning Shape Metrics based on Deformations and Transport*, in "Proceedings of ICCV 2009 and its Workshops, Second Workshop on Non-Rigid Sh ape Analysis and Deformable Image Alignment (NORDIA)", Kyoto, Japan, September 2009.

[6] G. CHARPIAT, P. MAUREL, J.-P. PONS, R. KERIVEN, O. FAUGERAS. *Generalized Gradients: Priors on Minimization Flows*, in "International Journal of Computer Vision", 2007.

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

[8] 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.

[9] F. FUSIER, V. VALENTIN, F. BREMOND, M. THONNAT, M. BORG, D. THIRDE, J. FERRYMAN. *Video Understanding for Complex Activity Recognition*, in "Machine Vision and Applications Journal", 2007, vol. 18, p. 167-188.

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

[11] 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.

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

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

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

[15] 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.

[16] 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° 4, p. 501-507.

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

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

[19] 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.

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

[21] 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° 5.

[22] 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.

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

[24] 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.

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

[26] V. 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 (IJ-CAI'03)", 9-15 September 2003.

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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[28] S. BAK. *Ré-identification de personne dans un réseau de cameras vidéo*, Université de Nice Sophia-Antipolis, July 2012, http://tel.archives-ouvertes.fr/tel-00763443.

[29] D. P. CHAU. *Suivi dynamique et robuste d'objets pour la reconnaissance d'activités*, Institut National de Recherche en Informatique et en Automatique (Inria), March 2012, http://hal.inria.fr/tel-00695567.

[30] G.-T. PUSIOL. *Discovery of human activities in video.*, Institut National de Recherche en Informatique et en Automatique (Inria), May 2012.

Articles in International Peer-Reviewed Journals

[31] C. F. CRISPIM-JUNIOR, V. JOUMIER, Y.-L. HSU, M.-C. PAI, P.-C. CHUNG, A. DECHAMPS, P. ROBERT, F. BREMOND. *Alzheimer's patient activity assessment using different sensors*, in "Gerontechnology", 2012, vol. 11, n° 2, p. 266-267 [*DOI :* 10.4017/GT.2012.11.02.597..678], http://hal.inria.fr/hal-00721549.

[32] M.-B. KAÂNICHE, F. BREMOND. *Recognizing Gestures by Learning Local Motion Signatures of HOG Descriptors*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2012, http://hal.inria.fr/hal-00696371.

International Conferences with Proceedings

[33] J. BADIE, S. BAK, S.-T. SERBAN, F. BREMOND. *Recovering people tracking errors using enhanced covariance-based signatures*, in "Fourteenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance - 2012", Beijing, Chine, July 2012, p. 487-493 [*DOI :* 10.1109/AVSS.2012.90], http://hal.inria.fr/hal-00761322.

[34] S. BAK, G. CHARPIAT, E. CORVEE, F. BREMOND, M. THONNAT. *Learning to Match Appearances by Correlations in a Covariance Metric Space*, in "12th European Conference on Computer Vision", Florence, Italy, A. FITZGIBBON, S. LAZEBNIK, P. PERONA, Y. SATO, C. SCHMID (editors), Lecture Notes in Computer Science - LNCS, Springer, October 2012, vol. 7574, p. 806-820 [*DOI :* 10.1007/978-3-642-33712-3_58], http://hal.inria.fr/hal-00731792.

[35] S. BAK, D. P. CHAU, J. BADIE, E. CORVEE, F. BREMOND, M. THONNAT. *Multi-target Tracking by Discriminative Analysis on Riemann Manifold*, in "ICIP - International Conference on Image Processing - 2012", Orlando, United States, IEEE Computer Society, June 2012, vol. 1, p. 1-4, http://hal.inria.fr/hal-00703633.

[36] P. BILINSKI, F. BREMOND. *Contextual Statistics of Space-Time Ordered Features for Human Action Recognition*, in "9th IEEE International Conference on Advanced Video and Signal-Based Surveillance", Beijing, China, September 2012, http://hal.inria.fr/hal-00718293.

[37] P. BILINSKI, F. BREMOND. *Statistics of Pairwise Co-occurring Local Spatio-Temporal Features for Human Action Recognition*, in "4th International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR), in conjunction with 12th European Conference on Computer Vision (ECCV)", Florence, Italy, October 2012, http://hal.inria.fr/hal-00760963.

[38] P. BILINSKI, E. CORVEE, S. BAK, F. BREMOND. *Relative Dense Tracklets for Human Action Recognition*, in "10th IEEE International Conference on Automatic Face and Gesture Recognition (FG)", Shanghai, China, 2012, To Appear in April 2013.

[39] C. F. CRISPIM-JUNIOR, F. BREMOND, V. JOUMIER. *A Multi-Sensor Approach for Activity Recognition in Older Patients*, in "The Second International Conference on Ambient Computing, Applications, Services and Technologies - AMBIENT 2012", Barcelona, Espagne, XPS/ThinkMindTM Digital Library, September 2012, in press, http://hal.inria.fr/hal-00726184.

[40] C. F. CRISPIM-JUNIOR, V. JOUMIER, Y.-L. HSU, P.-C. CHUNG, A. DECHAMPS, M.-C. PAI, P. ROBERT, F. BREMOND. *Alzheimer's patient activity assessment using different sensors*, in "ISG*ISARC 2012: 8th World Conference of the International Society for Gerontechnology in cooperation with the ISARC, International Symposium of Automation and Robotics in Construction", Eindhoven, Netherlands, J. VAN BRONSWIJK (editor), Gerontechnology 2012:11(2):63, ISG, IAARC and TU/e (Eindhoven University of Technology), 2012, p. 266-267, Best Paper Award of ISG*ISARC2012 [*DOI :* 10.4017/GT.2012.11.02.597.00], http://hal.inria.fr/hal-00721575.

[41] S. MOISAN, M. ACHER, J.-P. RIGAULT. *A Feature-based Approach to System Deployment and Adaptation*, in "ICSE, MISE Workshop - 34th International Conference on Software Engineering, Workshop on Modelling in Software Engineering", Zurich, Switzerland, June 2012, http://hal.inria.fr/hal-00708745.

[42] S. MOISAN. *Intelligent Monitoring of Software Components*, in "ICSE RAISE Workshop - 34th International Conference on Software Engineering, Workshop on Realizing AI Synergies in Software Engineering - 2012", Zurich, Switzerland, June 2012, http://hal.inria.fr/hal-00708737.

[43] L. PATINO, F. BREMOND, M. THONNAT. *On-line learning of activities from video*, in "AVSS - IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance 2012", Beijing, Chine, 2012, p. 234-239, http://hal.inria.fr/hal-00761461.

[44] S. SANKARANARAYANAN, F. BREMOND, D. TAX. *Qualitative Evaluation of Detection and Tracking Performance*, in "9th IEEE International Conference On Advanced Video and Signal Based Surveillance (AVSS 12)", Beijing, Chine, Sep 2012, http://hal.inria.fr/hal-00763587.

[45] S. ZAIDENBERG, B. BOULAY, F. BREMOND. *A generic framework for video understanding applied to group behavior recognition*, in "Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on", IEEE, September 2012, p. 136 -142 [*DOI :* 10.1109/AVSS.2012.1], http://hal.inria.fr/hal-00702179.

Scientific Books (or Scientific Book chapters)

[46] F. BRÉMOND, G. SACCO. *Technologies de l'information, limiter les effets de la maladie d'Alzheimer*, in "Alzheimer, éthique et société", F. Gzil and E. Hirsh, Sep 2012, p. 518-526.

[47] A. MARA, L. S. MASTELLA, M. PERRIN, M. THONNAT. *Ontologies and their use in geological knowledge formalization.*, in "Shared Earth Modeling: Knowledge based solutions for building and managing subsurface structural models", M. PERRIN, J. RAINAUD (editors), Technip, Paris, 2012, http://hal.inria.fr/hal-00761496.

[48] P. VERNEY, M. THONNAT, J.-F. RAINAUD. *Knowledge based approach of a data intensive problem: seismic interpretation*, in "Shared Earth Modeling: Knowledge based solutions for building and managing subsurface structural models", M. PERRIN, J. RAINAUD (editors), Technip, 2012, http://hal.inria.fr/hal-00761476.

Research Reports

[49] D. GAFFÉ, A. RESSOUCHE. *Algebras and Synchronous Language Semantics*, Inria, November 2012, n$^{\text{o}}$ RR-8138, 107, http://hal.inria.fr/hal-00752976.

Other Publications

[50] M. SOUDED, F. BRÉMOND. *Optimized Cascade of Classifiers for People Detection Using Covariance Features*, 2012, To appear in 2013 in International Conference on Computer Vision Theory and Applications Proceedings.

References in notes

[51] M. ACHER, P. COLLET, F. FLEUREY, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Modeling Context and Dynamic Adaptations with Feature Models*, in "Models@run.time Workshop", Denver, CO, USA, October 2009, http://hal.inria.fr/hal-00419990/en.

[52] M. ACHER, P. COLLET, P. LAHIRE, R. FRANCE. *Managing Feature Models with FAMILIAR: a Demonstration of the Language and its Tool Support*, in "Fifth International Workshop on Variability Modelling of Software-intensive Systems(VaMoS'11)", Namur, Belgium, VaMoS, ACM, January 2011.

[53] M. ACHER, P. COLLET, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Modeling Variability from Requirements to Runtime*, in "16th International Conference on Engineering of Complex Computer Systems (ICECCS'11)", Las Vegas, IEEE, April 2011.

[54] M. ACHER, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Tackling High Variability in Video Surveillance Systems through a Model Transformation Approach*, in "ICSE'2009 - MISE Workshop", Vancouver, Canada, May 2009, http://hal.inria.fr/hal-00415770/en.

[55] D. BENAVIDES, S. SEGURA, A. RUIZ-CORTES. *Automated Analysis of Feature Models 20 Years Later: A Literature Review*, in "Information Systems", September 2010, vol. 35, p. 615–636.

[56] J. BERCLAZ, F. FLEURET, E. TURETKEN, P. FUA. *Multiple object tracking using k-shortest paths optimization*, in "PAMI", 2011, vol. 33, n⁰ 9, p. 1806–1819.

[57] F. BREMOND, N. MAILLOT, M. THONNAT, V. VU. *Ontologies For Video Events*, in "Inria Research Report RR-5189", 2004.

[58] F. BREMOND, M. THONNAT. *Tracking multiple non-rigid objects in video sequences*, in "in proceedings of the IEEE Transactions On Automatic Control", 1998, vol. 8,5.

[59] D. P. CHAU, F. BREMOND, 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/.

[60] D. P. CHAU, F. BREMOND, 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, This work is supported by the PACA region, The General Council of Alpes Maritimes province, France as well as The ViCoMo, Vanaheim, Video-Id, Cofriend and Support projects., http://hal.inria.fr/inria-00599734/en/.

[61] A. CIMATTI, E. CLARKE, E. GIUNCHIGLIA, F. GIUNCHIGLIA, M. PISTORE, M. ROVERI, R. SEBASTIANI, A. TACCHELLA. *NuSMV 2: an OpenSource Tool for Symbolic Model Checking*, in "Proceeeding CAV", Copenhagen, Danmark, E. BRINKSMA, K. G. LARSEN (editors), LNCS, Springer-Verlag, July 2002, n⁰ 2404, p. 359-364, http://nusmv.fbk.eu/NuSMV/papers/cav02/ps/cav02.ps.

[62] R. DAVID, E. MULIN, P. MALLEA, P. ROBERT. *Measurement of Neuropsychiatric Symptoms in Clinical Trials Targeting Alzheimer's Disease and Related Disorders*, in "Pharmaceuticals", 2010, vol. 3, p. 2387-2397.

[63] 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.

[64] M. GINSBERG. *Multivalued Logics: A Uniform Approach to Inference in Artificial Intelligence*, in "Computational Intelligence", 1988, vol. 4, p. 265–316.

[65] M. GROSAM. *SOF: Service Oriented Framework*, Web site, http://sof.tiddlyspot.com/.

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

[67] J. F. HENRIQUES, R. CASEIRO, J. BATISTA. *Globally optimal solution to multi-object tracking with merged measurements*, 2011, In ICCV.

[68] 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.

[69] C. KUO, C. HUANG, R. NEVATIA. *Multi-target tracking by online learned discriminative appearance models*, 2010, In CVPR.

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

[71] Y. LI, C. HUANG, R. NEVATIA. *Learning to Associate: HybridBoosted Multi-Target Tracker for Crowded Scene*, 2009, The International Conference on Computer Vision and Pattern Recognition (CVPR).

[72] 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, September 2011, http://hal.inria.fr/inria-00617279/en.

[73] A.-T. NGHIEM, F. BRÉMOND, M. THONNAT. *Controlling Background Subtraction Algorithms for Robust Object Detection*, in "The 3rd International Conference on Imaging for Crime Detection and Prevention", London, United Kingdom, 3 December 2009.

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

[75] X. PENNEC, P. FILLARD, N. AYACHE. *A Riemannian Framework for Tensor Computing*, 2006, Int. Journal of Comp. Vision, 66(1):41–66..

[76] 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.

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

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

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

[80] 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.

[81] R. ROMDHANE, E. MULIN, A. DERREUMEAUX, N. ZOUBA, J. PIANO, L. LEE, I. LEROI, P. MALLEA, R. DAVID, M. THONNAT, F. BREMOND, P. ROBERT. *Automatic Video Monitoring system for assessment of Alzheimer's Disease symptoms*, in "The Journal of Nutrition, Health and Aging Ms(JNHA)", 2011, vol. JNHA-D-11-00004R1, http://hal.inria.fr/inria-00616747/en.

[82] H. B. SHITRIT, J. BERCLAZ, F. FLEURET, P. FUA. *Tracking multiple people under global appearance constraints*, 2011, In ICCV.

[83] B. C. STAUFFER, W. E. L. GRIMSON. *Adaptive background mixture models for real-time tracking*, in "in the IEEE Computer Vision and Pattern Recognition", 1999, vol. 2, p. 246-252.

[84] 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, n⁰ 3-4, ISSN 0003-4347 (Print) ISSN 1958-9395 (Online).

[85] 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, n⁰ 1, ISSN (Online): 1694-0784 ISSN (Print): 1694-0814.

[86] O. TUZEL, F. PORIKLI, P. MEER. *Human Detection via Classification on Riemannian Manifolds*, 2007, In IEEE Conf. Comp. Vision and Pattern Recognition (CVPR)..

[87] J. YAO, J. ODOBEZ. *Fast Human Detection from Videos Using Covariance Feature*, 2008, In: ECCV 2008 Visual Surveillance Workshop..

[88] S. ZAIDENBERG, B. BOULAY, C. GARATE, D. P. CHAU, E. CORVEE, F. BREMOND. *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, 10, http://hal.inria.fr/inria-00624356/en/.

[89] L. ZHANG, Y. LI, R. NETVATIA. *Global data association for multi-object tracking using network flows*, 2008, In CVPR.