Activity Report 2014

Project-Team STARS

Spatio-Temporal Activity Recognition Systems

Table of contents

Keywords: Perception, Semantics, Machine Learning, Software Engineering, Cognition

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

1. Members

Research Scientists

François Brémond [Team leader, Inria, Senior Researcher, HdR]
Guillaume Charpiat [Inria, Researcher]
Daniel Gaffé [Univ. Nice, Associate Professor]
Sabine Moisan [Inria, Researcher, HdR]
Annie Ressouche [Inria, Researcher]
Monique Thonnat [Inria, Senior Researcher, HdR]
Jean-Paul Rigault [External collaborator,Professor Univ. Nice]

Engineers

Slawomir Bak [Inria]
Vasanth Bathrinarayanan [Inria, granted by FP7 DEM@CARE project]
Carlos-Fernando Crispim Junior [Inria, granted by FP7 DEM@CARE project]
Giuseppe Donatiello [Inria, granted by FP7 DEM@CARE project]
Anaïs Ducoffe [NeoSensys start-up,from Oct 2014]
Baptiste Fosty [Inria, granted by Caisse des Dépôts et Consignations]
Rachid Guerchouche [NeoSensys start-up, from Oct 2014]
Julien Gueytat [Inria, until Oct 2014]
Anh-Tuan Nghiem [Inria, until Aug 2014, granted by Caisse des Dépôts et Consignations]
Jacques Serlan [Inria]
Malik Souded [Inria, granted by Min. de l'Economie]
Sofia Zaidenberg [NeoSensys start-up, until Aug 2014]
Bernard Boulay [NeoSensys start-up]
Yves Pichon [NeoSensys start-up]
Annunziato Polimeni [NeoSensys start-up, from Apr 2014]

PhD Students

Julien Badie [Inria]
Piotr Tadeusz Bilinski [Inria]
Carolina Garate Oporto [Inria]
Michal Koperski [Inria]
Thi Lan Anh Nguyen [Inria, from Dec 2014]
Minh Khue Phan Tran [Cifre grant]
Fnu Ratnesh Kumar [Inria]
Auriane Gros [Nice Hospital University, since Sep 2014]

Post-Doctoral Fellows

Duc Phu Chau [Inria, granted by Min. de l'Economie]
Serhan Cosar [Inria]
Antitza Dantcheva [Inria, from Mar 2014]
Salma Zouaoui-Elloumi [Inria, until Aug 2014]

Visiting Scientists

Vania Bogorny [Guest Professor, from Apr 2014]
Luis Campos Alvares [Guest Professor, from May 2014]
Jesse Hoey [Guest Professor, from Sep 2014]

Adlen Kerboua [PhD and associate professor (Algeria), from Oct 2014 until Nov 2014]
Pavel Vacha [Research engineer (Honeywell Praha), from May 2014]

Administrative Assistants

Jane Desplanques [Inria]
Nadežda Lacroix-Coste [Inria, from Oct 2014]

Others

Mohamed Bouatira [Inria, Internship, from Mar 2014 until Sep 2014]
Javier Ortiz [Inria, pre PhD, from Sep 2014]
Ines Sarray [Inria, from Apr 2014 until Oct 2014]
Omar Abdalla [Inria, Internship, from Apr 2014 until Sep 2014]
Jean Barboni [Inria, Internship, from May 2014 until Aug 2014]
Agustín Caverzasi [Inria, Internship, until Feb 2014]
Marine Chabran [Inria, Internship, from Sep 2014]
Sara Elkerdawy [Inria, Internship, Mar 2014]
Alvaro Gomez Uria Covella [Inria, Internship, from Mar 2014 until Dec 2014]
Filipe Martins de Melo [Inria, Internship, from Apr 2014 until Sep 2014]
Pablo Daniel Pusiol [Inria, Internship, from Apr 2014 until Sep 2014]
Carola Strumia [Inria, Internship, from Oct 2014]
Kouhua Zhou [Inria, Internship, from Jun 2014 until Sep 2014]
Etienne Corvée [External Collaborator, Linkcare Services]
Carolina Da Silva Gomes Crispim [Internship, from Jul 2014 until Sep 2014]
Alexandre Derreumaux [CHU Nice]
Auriane Gros [CHU Nice, from Oct 2014]
Vaibhav Katiyar [Internship, until Feb 2014]
Farhood Negin [Inria, Internship, from Apr 2014 until Nov 2014]
Thi Lan Anh Nguyen [Inria, Internship, from Mar 2014 until Oct 2014]
Ngoc Hai Pham [Inria, Internship, from May 2014 until Nov 2014]
Silviu-Tudor Serban [Internship, until Jan 2014]
Kartick Subramanian [Internship, until Aug 2014]
Jean-Yves Tigli [External collaborator, Associate professor Univ. Nice]
Philippe Robert [External Collaborator, Professor CHU Nice]
Alexandra Konig [External Collaborator, PhD Maastrich University]

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.
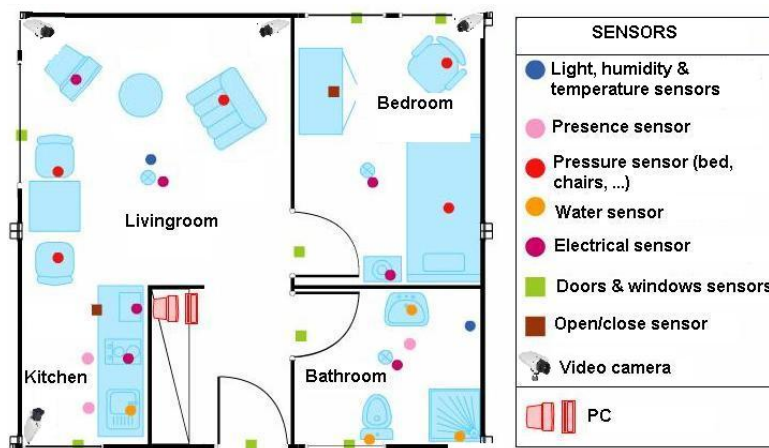
Figure 1. Homecare monitoring: the set of sensors embedded in an apartment

Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

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

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 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).

3. Research Program

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 engineering* reserach 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 RGB-D sensors 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 framework 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 analyzing 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 following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models), learning (how to learn the models needed for activity recognition) and activity recognition and discrete event systems.

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. For the second direction, we 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*

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 [14], 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.
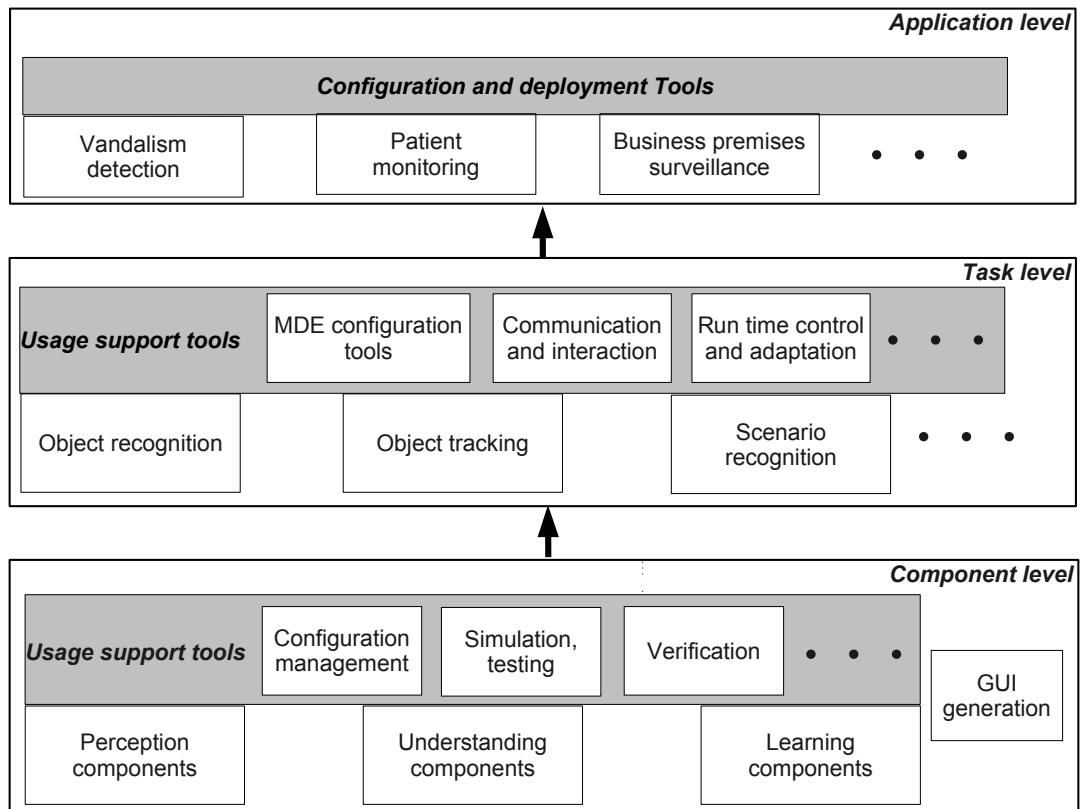
Figure 4. Global Architecture of an Activity Recognition The grey areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

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 parameterized 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 for verification, 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 [56]. 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 [75]. 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 [55], [84], [78].

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 [6], [1], [8] (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 SUP 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 [67], [85] (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).

4.3.1. Topics

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longitudinal experimentation. 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 behavioral 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.

4.3.2. Ethical and Acceptability Issues

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. Now, Stars can benefit from the support of the COERLE (Comité Opérationnel d'Evaluation des Risques Légaux et Ethiques) to help it to respect ethical policies in its applications.

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. New Software and Platforms

5.1. SUP
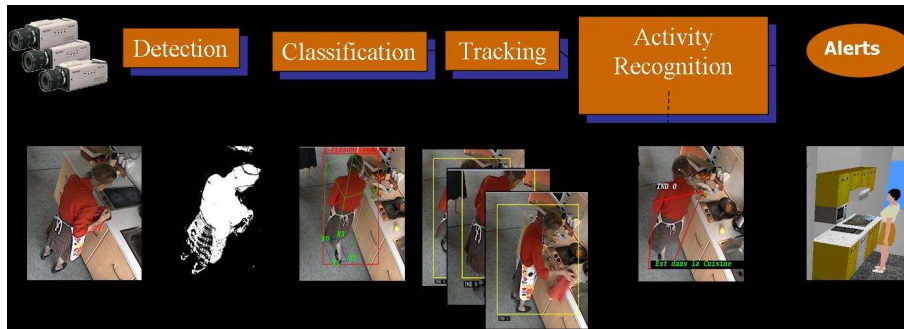
Figure 5. SUP workflow

5.1.1. Presentation

SUP is a Scene Understanding Software Platform (see Figure 5) written in C++ designed for analyzing video content . SUP is able to recognize events such as 'falling', 'walking' of a person. SUP divides the workflow of a video processing into several separated modules, such as acquisition, segmentation, up to activity recognition. Each module has a specific interface, and different plugins (corresponding to algorithms) can be implemented for a same module. We can easily build new analyzing systems thanks to this set of plugins. The order we can use those plugins and their parameters can be changed at run time and the result visualized on a dedicated GUI. This platform has many more advantages such as easy serialization to save and replay a scene, portability to Mac, Windows or Linux, and easy deployment to quickly setup an experimentation anywhere. SUP takes different kinds of input: RGB camera, depth sensor for online processing; or image/video files for offline processing.

This generic architecture is designed to facilitate:

1. integration of new algorithms into SUP;
2. iharing of the algorithms among the Stars team. Currently, 15 plugins are available, covering the whole processing chain. Some plugins use the OpenCV library.

Goals of SUP are twofold:

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

The plugins cover the following research topics:

- algorithms : 2D/3D mobile object detection, camera calibration, reference image updating, 2D/3D mobile object classification, sensor fusion, 3D mobile object classification into physical objects (individual, group of individuals, crowd), posture detection, frame to frame tracking, long-term tracking of individuals, groups of people or crowd, global tacking, basic event detection (for example entering a zone, falling...), human behaviour recognition (for example vandalism, fighting,...) and event fusion; 2D & 3D visualisation of simulated temporal scenes and of real scene interpretation results; evaluation of object detection, tracking and event recognition; image acquisition (RGB and RGBD cameras) and storage; video processing supervision; data mining and knowledge discovery; image/video indexation and retrieval.

- languages : scenario description, empty 3D scene model description, video processing and understanding operator description;

- knowledge bases : scenario models and empty 3D scene models;
- learning techniques for event detection and human behaviour recognition;

5.1.2. Improvements

Currently, the OpenCV library is fully integrated with SUP. OpenCV provides standardized data types, a lot of video analysis algorithms and an easy access to OpenNI sensors such as the Kinect or the ASUS Xtion PRO LIVE.

In order to supervise the GIT update progress of SUP, an evaluation script is launched automatically everyday. This script updates the latest version of SUP then compiles SUP core and SUP plugins. It executes the full processing chain (from image acquisition to activity recognition) on selected data-set samples. The obtained performance is compared with the one corresponding to the last version (i.e. day before). This script has the following objectives:

- Check daily the status of SUP and detect the compilation bugs if any.
- Supervise daily the SUP performance to detect any bugs leading to the decrease of SUP performance and efficiency.

The software is already widely disseminated among researchers, universities, and companies:

- PAL Inria partners using ROS PAL Gate as middleware
- Nice University (Informatique Signaux et Systèmes de Sophia), University of Paris Est Créteil (UPEC - LISSI-EA 3956)
- EHPAD Valrose, Institut Claude Pompidou
- European partners: Lulea University of Technology, Dublin City University,...
- Industrial partners: Toyota, LinkCareServices, Digital Barriers

Updates and presentations of our framework can be found on our team website https://team.inria.fr/stars/ software . Detailed tips for users are given on our Wiki website http://wiki.inria.fr/stars and sources are hosted thanks to the Inria software developer team SED.

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.
- To allow 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 amounts 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 user 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.
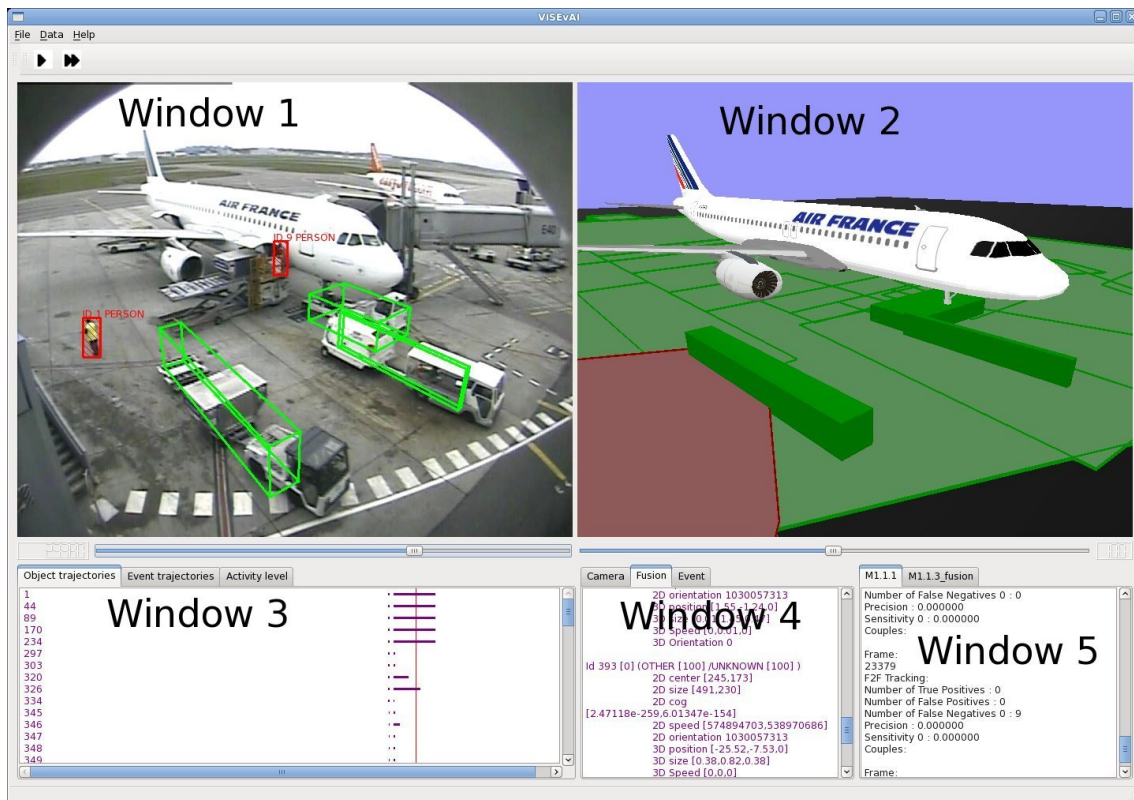
Figure 6. GUI of the ViSEvAl software

The GUI is composed of 3 different parts:

1. The visualization of results windows 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,...).

   – 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,

Figure 7. The object window enables users to choose the object to display
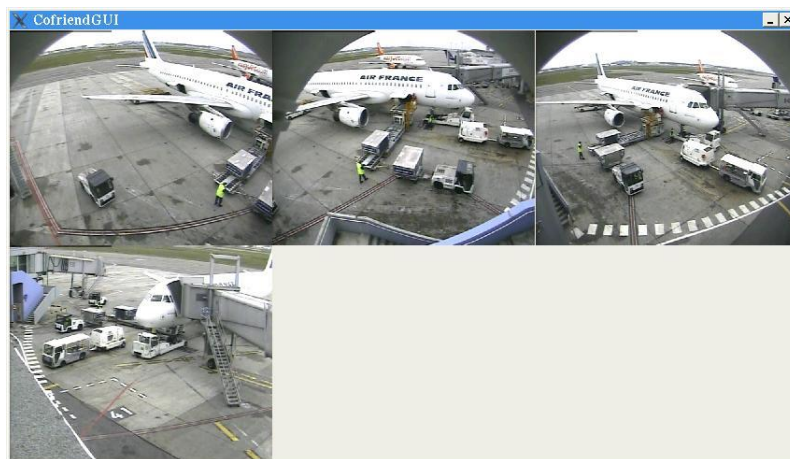
Figure 8. The multi-view window

– 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* [68](see Figure 9) is a set of tools devoted to design, simulate, verify and generate code for LE [18] [81] programs. LE is a synchronous language supporting a modular compilation. It also supports automata possibly designed with a dedicated graphical editor and implicit Mealy machine definition.

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 [65] in order to perform validation of program behaviors. In 2014, LE supports data value for automata and CLEM is used in 2 research axes of the team (SAM and SynComp). CLEM is registered at the APP since May 2014.

The work on CLEM was published in [68], [69], [18], [19].
Web page: http://www-sop.inria.fr/teams/pulsar/projects/Clem/

6. New Results

6.1. Highlights of the Year

NeoSensys, a spin off of the Stars team which aims at commercializing video surveillance solutions for the retail domain, has been created in September 2014.

6.2. 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.

6.2.1. *Perception for Activity Recognition*

Participants: Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Bernard Boulay, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Carolina Garate, Michal Koperski, Ratnesh Kumar, Filipe Martins, Malik Souded, Anh Tuan Nghiem, Sofia Zaidenberg, Monique Thonnat.

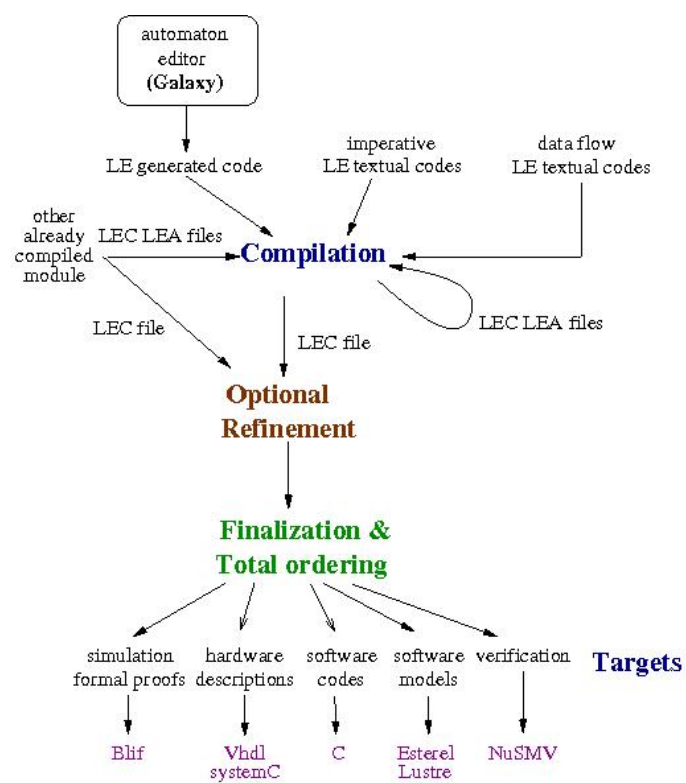*Figure 9. The Clem Toolkit*

For perception, the main achievements are:

- Our new covariance descriptor has led to many publications and applications already. The work on this topic is now more about the precise use of the descriptor in varied applications than the design of new descriptors.
- The new action descriptors have led to finer gesture classification. As our target application is the detection of the Alzheimer syndrome from gesture analysis, which requires still finer descriptors, we will continue the work on this topic.
- The different shape priors developed (for shape growth enforcement, shape matching, articulated motion) have been formulated and designed so that efficient optimization tools could be used, leading to global optimality guarantees. These particular problems can thus be considered as solved, but there is still much work to be done on shape and related optimization, in particular to obtain shape statistics for human action recognition.
- The success obtained in the control of trackers is a proof of concept, but this work still needs to be pursued to get more practical and to be applied on more real world videos.

More precisely, the new results for perception for activity recognition are:

- People Detection for Crowded Scenes (6.3),
- Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results (6.4),
- Head detection using RGB-D camera (6.5),
- Video Segmentation and Multiple Object Tracking (6.6),
- Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation (6.7),
- Multi-label Image Segmentation with Partition Trees and Shape Prior (6.8),
- Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking (6.9),
- An Approach to Improve Multi-object Tracker Quality using Discriminative Appearances and Motion Model Descriptor (6.10),
- Person re-identification by pose priors(6.11),
- Global tracker : an online evaluation framework to improve tracking quality (6.12),
- Human action recognition in videos (6.13),
- Action Recognition using 3D Trajectories with Hierarchical Classifier (6.14),
- Action Recognition using Video Brownian Covariance Descriptor for Human (6.15),
- Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance (6.16).

6.2.2. *Semantic Activity Recognition*

Participants: Vania Bogorny, Luis Campos Alvares, Vasanth Bathrinarayanan, Guillaume Charpiat, Duc Phu Chau, Serhan Cosar, Carlos F. Crispim Junior, Giuseppe Donatielo, Baptiste Fosty, Carolina Garate, Alvaro Gomez Uria Covella, Alexandra Konig, Farhood Negin, Anh-Tuan Nghiem, Philippe Robert, Carola Strumia.

For activity recognition, the main advances on challenging topics are:

- The utilization by clinicians for their everyday work of a first monitoring system able to recognize complex activities, to evaluate in real-time older people performance in an ecological room at Nice Hospital.
- The successful processing of over 80 older people videos and matching their performance for autonomy at home (e.g. walking efficiency) and cognitive disorders (e.g. realisations of executive tasks) with gold standard scales (e.g. NPI, MMSE). This research work contributes to the early detection of deteriorated health status and the early diagnosis of illness.
- The fusion of events coming from camera networks and heterogeneous sensors (e.g. RGB videos, Depth maps, audio, accelerometers).
- The management of the uncertainty of primitive events.
- The generation of event models in an unsupervised manner.

For this research axis, the contributions are :

- Autonomous Monitoring for Securing European Ports (6.17),
- Video Understanding for Group Behavior Analysis (6.18),
- Evaluation of an event detection framework for older people monitoring: from minute to hour-scale monitoring and Patients autonomy and dementia assessment (6.19),
- Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems (6.20),
- Assisted Serious Game for older people (6.21),
- Enhancing Pre-defined Event Models using Unsupervised Learning (6.22),
- Using Dense Trajectories to Enhance Unsupervised Action Discovery (6.23),
- Abnormal Event Detection in Videos and Group Behavior Analysis (6.24).

6.2.3. *Software Engineering for Activity Recognition*

Participants: François Brémond, Daniel Gaffé, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Omar Abdalla, Mohamed Bouatira, Ines Sarray, Luis-Emiliano Sanchez.

For the software engineering part, the main achievements are the Software Engineering methods and tools applied to video analysis. We have demonstrated that these approaches are appropriate and useful for video analysis systems:

- Run time adaptation using MDE is a promising approach. Our current prototype resorts to tools and technologies which were readily available. This made possible a proof of concepts.
- Introducing metrics in feature models was valuable to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration.
- The synchronous approach is well suited to describe reactive systems in a generic way, it has a well-established formal foundation allowing for automatic proofs, and it interfaces nicely with most model-checkers.

The contributions for this research axis are:

- Model-Driven Engineering for Activity Recognition Systems(6.25),
- Scenario Analysis Module (6.26),
- The Clem Workflow (6.27),
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.28).

6.3. People Detection for Crowded Scenes

Participants: Malik Souded, François Brémond.

keywords: people detection, crowded scenes, features, boosting.

This works aims at proposing an efficient people detection algorithm which can deal with crowded scenes.

6.3.1. *Early Work*

We have previously proposed an approach which optimizes state-of-the-art methods [Tuzel 2007, Yao 2008], based on training cascade of classifiers using LogitBoost algorithm on region covariance descriptors. This approach performs in real time and provides good detection performances in low to medium density scenes (see some examples in figure 10). However, this approach shows its limits on crowded scenes. Both detection accuracy and detection time are highly impacted in this case. The detection time increases dramatically due to the number of people in images, which forces the evaluation of many cascade levels, while the numerous partial occlusions highly decrease the detection rate (the considered detector is a full-body detector). To deal with these issues, we are working on a new approach.

6.3.2. Current Work

Our approach is based on training a cascade of classifiers using Boosting algorithms too, but on large sets of various features with several parameters for each of them (LBP, Haar-Like, HOG, Region Covariance Descriptor, etc.). The variety of features is motivated by three main reasons:

- Using fast features like LBP and Haar-like in the first levels of the cascade allows a fast rejection of a high part of negatives. The remaining ones will be rejected by a more sophisticated feature like Covariance Descriptor. This will highly decrease the detection time.

- Covariance Descriptor are not discriminative enough for very small regions. Our aim is to train the new detector on specific body parts, especially the upper one (shoulders and heads) to increase detection rate in highly crowded scenes (with a high rate of partial occlusions). Using a large set of various features allows the training system to select the ones which provide the best discriminative power for these regions.

- The possibility to combine several features to describe the same region, even by a simple concatenation, providing more discriminative power than using single features.

Another part of this approach consists in the optimization of the detector at two levels:

- Optimizing the training process by first clustering both positive and negative training samples. This clustering allows to focus on the hard samples which are too close to the other class from a classification point of view, providing more accurate detectors.

- Iterative training of several detectors on randomly selected samples, and weighting of the training samples according to their classification confidence, which allows to improve the clustering process.

The evaluation of this approach is still in progress.

6.4. Walking Speed Detection on a Treadmill using an RGB-D camera : experimentations and results

Participants: Baptiste Fosty, François Brémond.

keywords: RGB-D camera analysis, walking speed, serious games

Within the context of the developement of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed an algorithm to compute the walking speed of a person on a treadmill. The goal is to use this speed inside the game to control the displacement of the avatar, and then for the patient to perform some physical as well as cognitive tasks. For the evaluation of the accuracy of the algorithm, we collected a video data set of healthy people walking on a motorized treadmill.

Protocol. With the help of a specialist in the domain of physical activities, a protocol has been set up to cover the spectrum of the possible walking speeds and to prove the reproducibility of the results. This protocol consists in performing three times ten minutes of walking on the motorized treadmill, each attempt being itself divided in five times two minutes at the following speeds : 1.5 km/h, 2.5 km/h, 3.5 km/h, 4.5 km/h and 5.5 km/h. Participants, mostly people from the age of 18 to 60 without any physical disorder that could influence the gait, were asked to keep a natural gait and to follow the rotation of the treadmill.

Ground truth. The quantitative performances of the walking speed computation are evaluated by comparison with the speed of the walking person. The speed references are twofold :

- a theoretical value : the speed displayed by the treadmill, set up by the participant but imposed by the protocol (see Figure 12, red graph),

- a practical value : white marks have been painted on the treadmill to recompute the real speed of the rotation and so the walking speed (see Figure 12, green and blue graphs).

Results. The results presented herein are based on the videos of 36 participants who performed the protocol described above, with 17 males and 19 females, with an average age of 32.1±7.7 years, an average height of 171.1±9.1 cm and an average weight of 67.4±13.6 kg.

Figure 10. Some examples of detection using the previously proposed approach (see section Early Work).

Speed (km/h)	Number of observations	Average (km/h)	Standard deviation (km/h)	Mean absolute error (km/h)
1.5 (WM)	11823 (3170)	1.59 (1.508)	0.20 (0.014)	0.15 (0.011)
2.5 (WM)	14549 (5113)	2.56 (2.478)	0.20 (0.026)	0.16 (0.027)
3.5 (WM)	17399 (7133)	3.53 (3.453)	0.20 (0.040)	0.16 (0.051)
4.5 (WM)	19800 (9129)	4.47 (4.427)	0.21 (0.059)	0.17 (0.079)
5.5 (WM)	22163 (11180)	5.36 (5.397)	0.26 (0.116)	0.21 (0.116)
Total (WM)	**85734 (35725)**	**3.5 (3.453)**	**0.22 (0.074)**	**0.17 (0.071)**

Figure 11. Walking speed results. The number of observations corresponds to the number of time the speed has been computed. The WM lines refer to the results of the detection of the white mark.

The table in figure 11 shows the statistical evaluation of the performances of the system. The average column shows that the accuracy of the system is better for the median speeds (around 4.5km/h). When the person is walking slower, the system overestimate the speed due to the wrongly detected steps whereas when faster, there is an underestimation because of missing the exact time when the distance between feet is maximum (framerate too low).

A paper reporting this work is actually under writing process.

6.5. Head Detection Using RGB-D Camera

Participants: Marine Chabran, François Brémond.

keywords: RGB-D camera analysis, head detection, serious games

The goal of this work is to improve a head detection algorithm using RGB-D sensor (like a Kinect camera) for action recognition as part of a study of autism. The psychologists want to compare the learning process of children with autism syndrome depending on games (digital or physical toys).

The algorithm described in [79] represents a head by its center position. It takes three steps to determine this point :

- Determine possible head center positions using a head model : inner circle radius=6 cm, outer circle radius=20 cm (Figure 13).
  A good inner point is a point on the inner circle verifying :

$$depthHeadCenter + 30cm > depthInnerPoint > depthHeadCenter - 30cm.$$
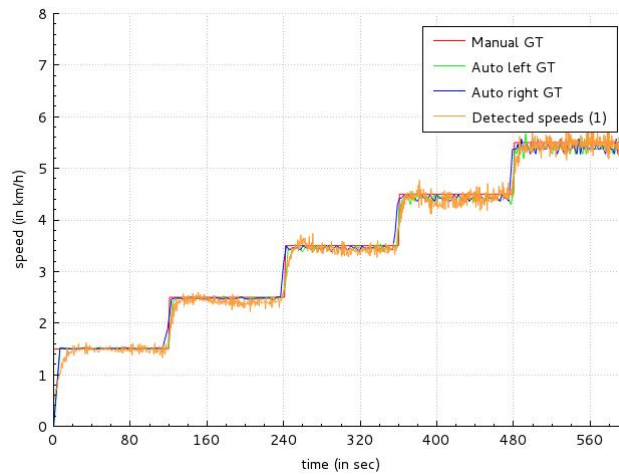
Figure 12. Graphical results of one attempt of protocal issue. The red graph corresponds to the theoretical speed, green and blue graphs correspond respectively to the detected speed of the left and right white marks on the treadmill, and the orange graph is the detected walking speed of the person.

A good outer point is a point on the outer circle verifying :

$$depthHeadCenter < depthOuterPoint + 15cm.$$

- Merge close head centers separated by less than 4 pixels.
- Select final head center according to its score (calculated according to the number of good inner and outer points).
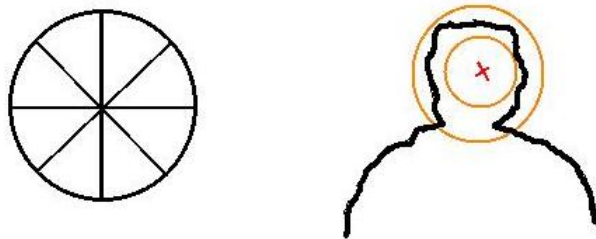
Figure 13. Each circle is divided in n parts (n=8). The points on the inner circle must have a similar depth with the center point, the points on the outer circle must be further than the center point compared to the camera

For now, it works well within video where people are close to the camera (about 1 meter) and without any background just behind them (Figure 14).

The problem is when the person is sitting and the head is ahead of the body (Figure 15) or close to a wall, the difference between head depth and outer circle depth becomes not sufficient (about 10 cm).

We have evaluated the performance of this algorithm with two data sets (Table 1). For Lenval Hospital data set, we have evaluated 2 series of 200 frames, for the Smart Home data set, we have evaluated 3 series of 300 frames (a total of 1300 heads).

Table 1. Performance of head detection and people detection on two different data sets.

Videos	Head Detection (%)	People detection (%)
Lenval Hospital dataset (Figure 14)	89.7	96.9
Rest home dataset (Figure 15)	62.8	85.3
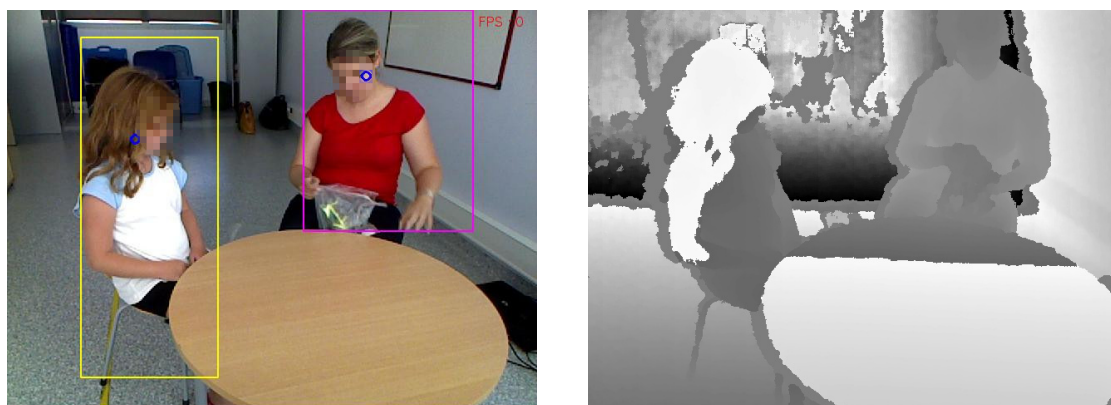
Figure 14. Result of head detection - good detection. The bounding box represents the person, the small blue circle represents the head center.
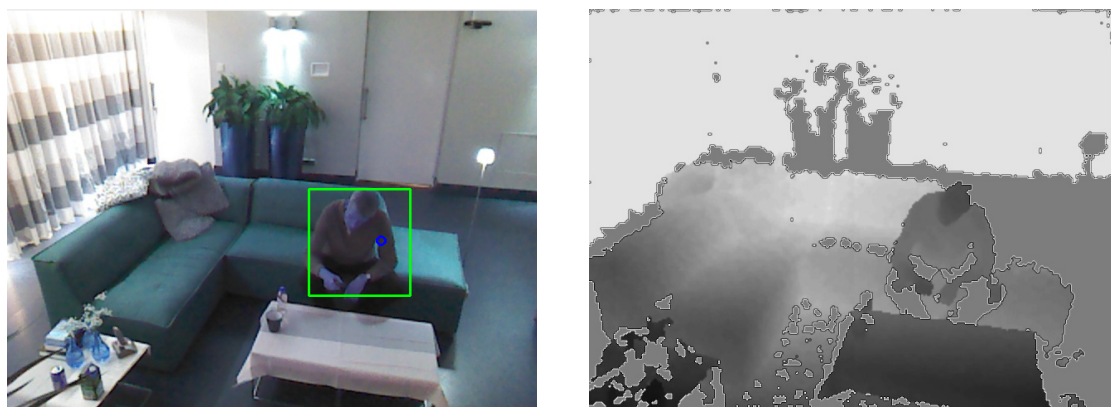
Figure 15. Result of head detection - wrong detection. The green bounding box represents the person, the small blue circle represents the head center.

6.6. Video Segmentation and Multiple Object Tracking

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

keywords: Fibers, Graph Partitioning, Message Passing, Iterative Conditional Modes, Video Segmentation, Video Inpainting

This year we focussed on multiple object tracking, and writing of the thesis manuscript of Ratnesh (defense on December 2014).

The first contribution of this thesis is in the domain of video segmentation wherein the objective is to obtain a dense and coherent spatio-temporal segmentation. We propose joining both spatial and temporal aspects of a video into a single notion *Fiber*. A Fiber is a set of trajectories which are spatially connected by a mesh. Fibers are built by jointly assessing spatial and temporal aspects of the video. Compared to the state-of-the-art, a fiber based video segmentation presents advantages such as a natural spatio-temporal neighborhood accessor by a mesh, and temporal correspondences for most pixels in the video. Furthermore, this fiber-based segmentation is of quasi-linear complexity w.r.t. the number of pixels. The second contribution is in the realm of multiple object tracking. We proposed a tracking approach which utilizes cues from point tracks, kinematics of moving objects and global appearance of detections. Unification of all these cues is performed on a Conditional Random Field. Subsequently this model is optimized by a combination of message passing and an Iterated Conditional Modes (ICM) variant to infer object-trajectories. A third, minor, contribution relates to the development of suitable feature descriptor for appearance matching of persons. All of our proposed approaches achieve competitive and better results (both qualitatively and quantitatively) than state-of-the-art open source datasets.

Figure 16. **Consistent people crossing in dense scenarios.** *The two images are 121 frames apart.*

This first part of the thesis was published at IEEE WACV at the beginning of this year [43], and the work on *multiple object tracking* was recently presented at Asian Conference on Computer Vision [44]

Sample visual results from our recent publication [44] can be seen in Figure 16.

6.7. Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation

Participant: Guillaume Charpiat [contact].

This work has been done in collaboration with Yuliya Tarabalka (Ayin team, Inria-SAM), Bjoern Menze (Technische Universität München, Germany), and Ludovic Brucker (NASA GSFC, USA) [http://www.nasa.gov].

keywords: Video segmentation, graph cut, shape analysis, shape growth

The automatic segmentation of objects from video data is a difficult task, especially when image sequences are subject to low signal-to-noise ratio or low contrast between the intensities of neighboring structures. Such challenging data are acquired routinely, for example, in medical imaging or satellite remote sensing. While individual frames can be analyzed independently, temporal coherence in image sequences provides a lot of information not available for a single image. In this work, we focused on segmenting shapes that grow or shrink monotonically in time, from sequences of extremely noisy images.

We proposed a new method for the joint segmentation of monotonically growing or shrinking shapes in a time sequence of images with low signal-to-noise ratio [32]. The task of segmenting the image time series is expressed as an optimization problem using the spatio-temporal graph of pixels, in which we are able to impose the constraint of shape growth or shrinkage by introducing unidirectional infinite-weight links connecting pixels at the same spatial locations in successive image frames. The globally-optimal solution is computed with graph-cuts. The performance of the proposed method was validated on three applications: segmentation of melting sea ice floes; of growing burned areas from time series of 2D satellite images; and of a growing brain tumor from sequences of 3D medical scans. In the latter application, we imposed an additional inter-sequences inclusion constraint by adding directed infinite-weight links between pixels of dependent image structures. Figure 17 shows a multi-year sea ice floe segmentation result. The proposed method proved to be robust to high noise and low contrast, and to cope well with missing data. Moreover, in practice, its complexity was linear in the number of images.

6.8. Multi-label Image Segmentation with Partition Trees and Shape Prior

Participant: Guillaume Charpiat [contact].

This work has been done in collaboration with Emmanuel Maggiori and Yuliya Tarabalka (Ayin team, Inria-SAM).

keywords: partition trees, multi-class segmentation, shape priors, graph cut

The multi-label segmentation of images is one of the great challenges in computer vision. It consists in the simultaneous partitioning of an image into regions and the assignment of labels to each of the segments. The problem can be posed as the minimization of an energy with respect to a set of variables which can take one of multiple labels. Throughout the years, several efforts have been done in the design of algorithms that minimize such energies.
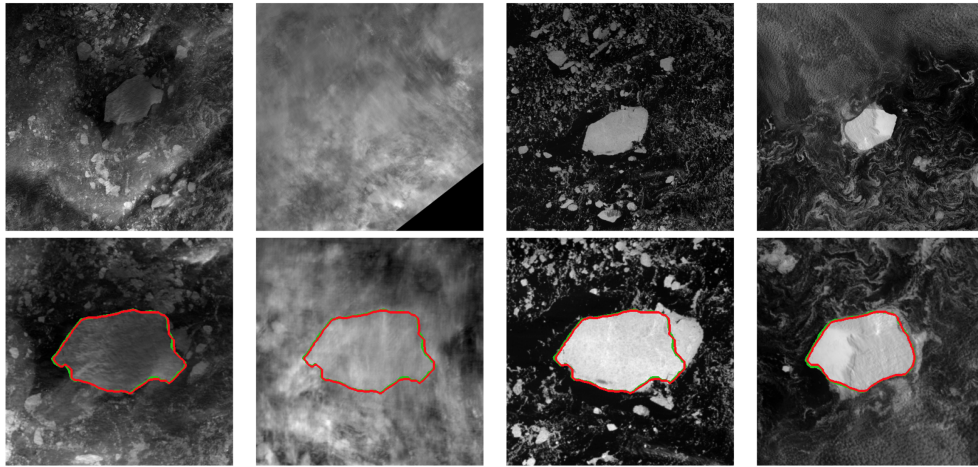
Figure 17. Top: MODIS images at four time moments (days 230, 233, 235 and 267 of 2008, respectively). Bottom: corresponding aligned images with segmentation contours (in red). Manual segmentation (ground truth) is shown in green.

We propose a new framework for multi-label image segmentation with shape priors using a binary partition tree [50]. In the literature, such trees are used to represent hierarchical partitions of images, and are usually computed in a bottom-up manner based on color similarities, then processed to detect objects with a known shape prior. However, not considering shape priors during the construction phase induces mistakes in the later segmentation. This study proposes a method which uses both color distribution and shape priors to optimize the trees for image segmentation. The method consists in pruning and regrafting tree branches in order to minimize the energy of the best segmentation that can be extracted from the tree. Theoretical guarantees help reducing the search space and make the optimization efficient. Our experiments (see Figure 18) show that the optimization approach succeeds in incorporating shape information into multi-label segmentation, outperforming the state-of-the-art.

6.9. Automatic Tracker Selection and Parameter Tuning for Multi-object Tracking

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

Keywords: object tracking, machine learning, tracker selection, parameter tuning

Many approaches have been proposed to track mobile objects in a scene [87], [45]. However the quality of tracking algorithms always depends on video content such as the crowded level or lighting condition. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has already been determined, there are still some issues (e.g. the determination of the best parameter values or the online estimation of the tracking reliability) for adapting online this tracker to the video content variation. In order to overcome these limitations, we propose the two following approaches.

The main idea of the first approach is to learn offline how to tune the tracker parameters to cope with the tracking context variations. The tracking 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. In an offline phase, training video sequences are classified by clustering their contextual features. Each context cluster is then associated to satisfactory tracking parameters using tracking annotation associated to training videos. In the online control phase, once a context change is

	$\mathcal{A}$	$\mathcal{D}$
SVM	0.764	0.50
GC	0.799	0.49
TC	0.804	0.47
Opt.	0.813	0.69

(d) Manual. (e) SVM. (f) GC. (g) BPT Opt.

(h) Performance.

Time [s]	Time [s] (Props. 1-2)
214	84

(i) Execution time.

(a) Color image. (b) Manual segment. (c) BPT optimization.
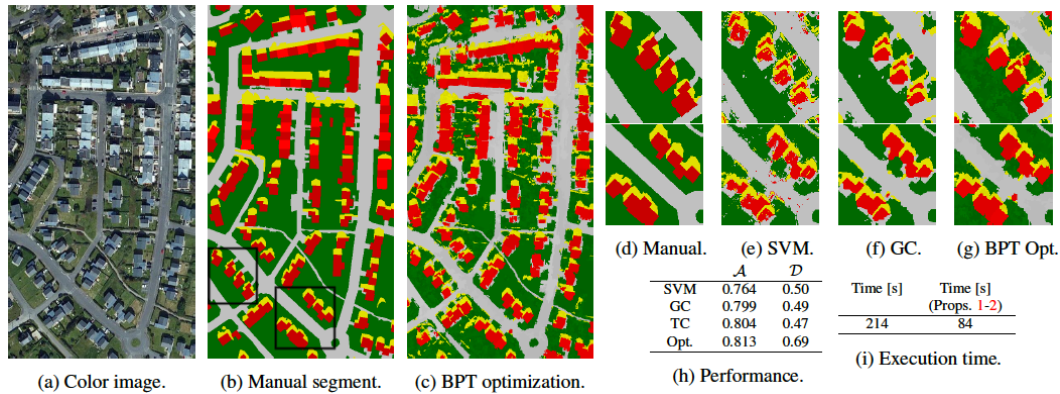
Figure 18. Classification results for the satellite image over Brest. $\mathcal{A}$ denotes overall classification accuracy, and $\mathcal{D}$ denotes average building's overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with α-expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).

detected, the tracking parameters are tuned using the learned parameter values. This work has been published in [30].

A limitation of the first approach is the need of annotated data for training. Therefore we have proposed a second approach without training data. In this approach, the proposed strategy combines an appearance tracker and a KLT tracker for each mobile object to obtain the best tracking performance (see figure 19). This helps to better adapt the tracking process to the spatial distribution of objects. Also, while the appearance-based tracker considers the object appearance, the KLT tracker takes into account the optical flow of pixels and their spatial neighbours. Therefore these two trackers can improve alternately the tracking performance.
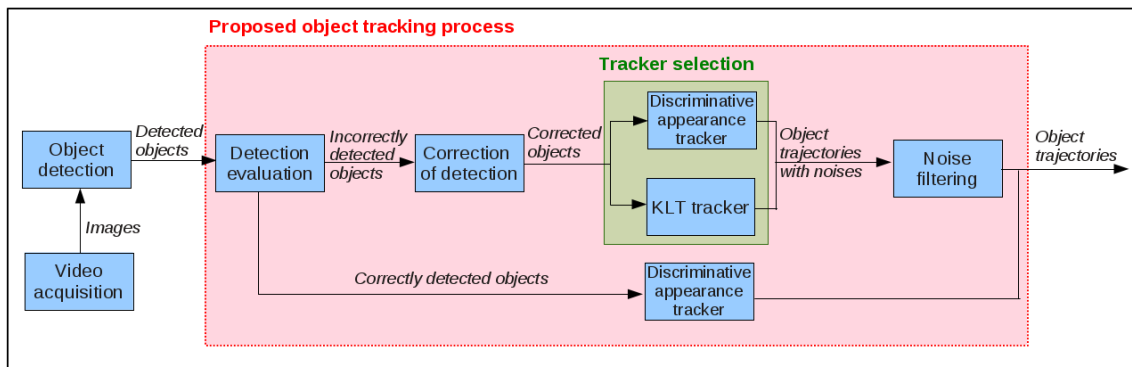
Figure 19. The scheme of the second approach

The second approach has been experimented on three public video datasets. Figure 20 presents correct tracking results of this approach even with strong object occlusion in PETS 2009 dataset. Table 2 presents the evaluation results of the proposed approach, the KLT tracker, the appearance tracker and different trackers from the

state of the art. While using separately the KLT tracker or the appearance tracker, the performance is lower than other approaches from the state of the art. The proposed approach by combining these two trackers improves significantly the tracking performance and obtains the best values for both metrics. This work has been published in [39].
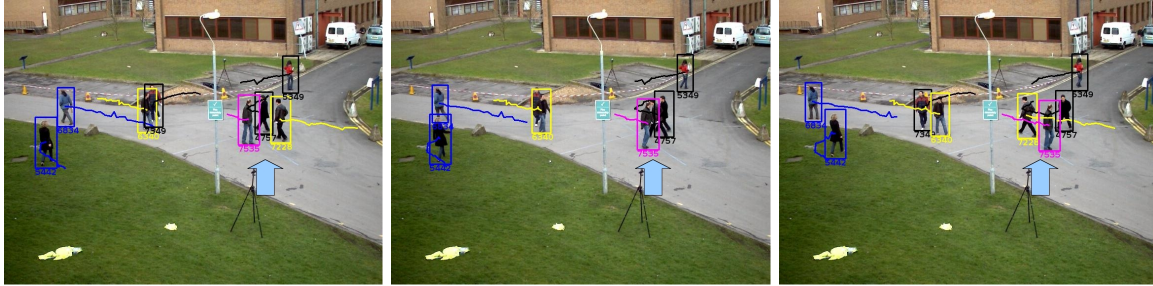
Figure 20. Tracker result: Three persons of Ids 7535, 7228 and 4757 (marked by the cyan arrow) are occluded each other but their identities are kept correctly after occlusion.

Table 2. Tracking results on the PETS sequence S2.L1, camera view 1, sequence time 12.34. The best values are printed in bold.

Method	MOTA	MOTP
Berclaz et al. [60]	0.80	0.58
Shitrit et al. [86]	0.81	0.58
KLT tracker	0.41	**0.76**
Appearance tracker	0.62	0.63
Proposed approach	**0.86**	0.72

6.10. An Approach to Improve Multi-object Tracker Quality Using Discriminative Appearances and Motion Model Descriptor

Participants: Thi Lan Anh Nguyen, Duc Phu Chau, François Brémond.

Keywords: Tracklet fusion, Multi-object tracking

Many recent approaches have been proposed to track multi-objects in a video. However, the quality of trackers is remarkably effected by video content. In the state of the art, several algorithms are proposed to handle this issue. The approaches in [39] and [64] propose methods which compute online or learn descriptor weights during tracking process. These algorithms adapt the tracking to the scene variations but are less effective when mis-detection occurs in a long period of time. Inversely, the algorithms in [59] and [58] can recover a long-term mis-detection by fusing tracklets. However, the descriptor weights in these tracklet fusion algorithms are fixed in the whole video. Furthermore, above algorithms track objects based on object appearance which is not reliable enough when objects look similar to each other.

In order to overcome mentioned issues, the proposed approach brings three contributions: (1) appearance descriptors and motion model combination, (2) online discriminative descriptor weight computation and (3) discriminative descriptors based tracklet fusion. In particular, the appearance of one object can be discriminative with other objects in this scene but can be similar with other objects in another scene. Therefore, tracking objects based on only object appearance is less effective. In order to improve tracker quality, assuming that objects move with constant velocity, this approach firstly combines a constant velocity model from [70]

and other appearance descriptors. Continuously, discriminative descriptor weights are computed online to adapt the tracking to each video scene. The more a descriptor discriminates one tracklet over other tracklets, the higher its weight value is. Next, based on these descriptor weights, the similarity score between the target tracklet with its candidate is computed. In the last step, tracklets are fused to a long trajectory by Hungarian algorithm with the optimization of global similarity scores.

The proposed approach gets results of tracker in [63] as input and is tested on challenge datasets. This approach achieves comparable results with other trackers from the state of the art. Figure 1 shows that the tracklet keeps its ID even when occlusion occurs. Table 1 shows the better performance of this approach compared to other trackers from the state of the art.
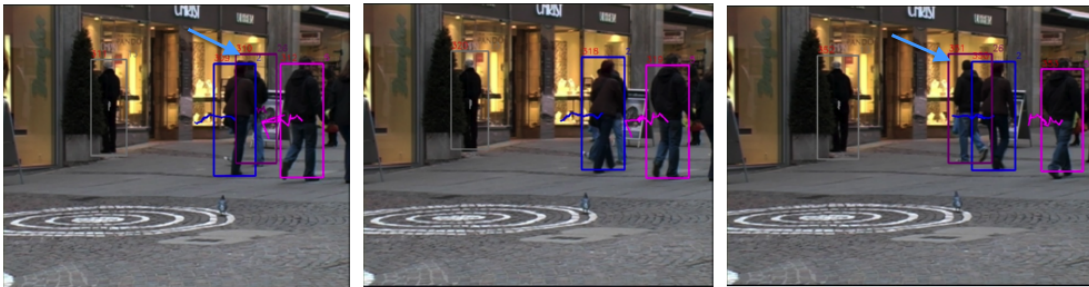
Figure 21. The proposed approach with TUD dataset: Object ID_{26} (presented by pink bounding box) keeps its ID correctly after 11 frames of mis-detection.

Table 3. Tracking results on datasets: TUD-Stadtmitte and TUD-crossing. The best values are printed in bold

Dataset	Method	MT(%)	PT(%)	ML(%)
TUD-Stadtmitte	[57]	60.0	30.0	10.0
TUD-Stadtmitte	[30]	**70.0**	10.0	20.0
TUD-Stadtmitte	[71]	**70.0**	30.0	**0.0**
TUD-Stadtmitte	[95]	**70.0**	30.0	**0.0**
TUD-Stadtmitte	Ours	**70.0**	30.0	**0.0**
TUD-Crossing	[89]	**53.8**	38.4	7.8
TUD-Crossing	Ours	**53.8**	46.2	**0.0**

6.11. Person Re-identification by Pose Priors

Participants: Slawomir Bak, Sofia Zaidenberg, Bernard Boulay, Filipe Martins, Francois Brémond.

keywords: re-identification, pose estimation, metric learning
Human appearance registration, alignment and pose estimation

Re-identifying people in a network of cameras requires an invariant human representation. State of the art algorithms are likely to fail in real-world scenarios due to serious perspective changes. Most of existing approaches focus on invariant and discriminative features, while ignoring the body alignment issue. In this work we proposed 3 methods for improving the performance of person re-identification. We focus on eliminating perspective distortions by using 3D scene information. Perspective changes are minimized by affine transformations of cropped images containing the target (1). Further we estimate the human pose for (2) clustering data from a video stream and (3) weighting image features. The pose is estimated using 3D scene

(a) left   (b) right
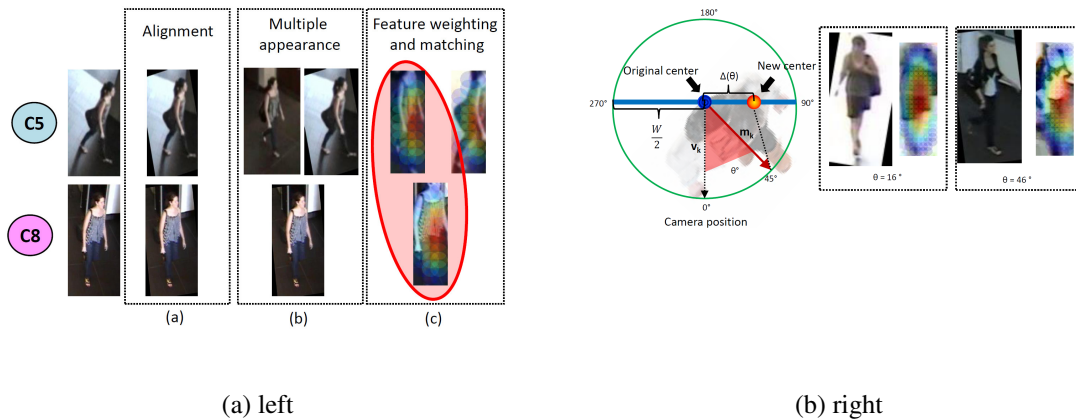
Figure 22. Improvements on re-identification using viewpoint cues: (a) target alignment; (b) multiple target appearance based on clustering; (c) pose orientation-driven weighting. The left illustration shows an example of the same person viewed from two different cameras. The right image presents pose estimation algorithm.

information and motion of the target. Pose orientation is computed by dot product between viewpoint vector and motion of the target (see figure 22). We validated our approach on a publicly available dataset with a network of 8 cameras. The results demonstrated significant increase in the re-identification performance over the state of the art [36].

Matching employing pose priors

Currently we are working on learning the matching strategy of appearance extracted from different poses. We employ well known metric learning tools for matching given poses. Let us assume that pose can be described by the angle between the motion vector of the target and the viewpoint vector of the camera (see figure 22). Thus for each target appearance we can express the pose as the angle in the range of [0,360). We decide to divide this range into n bins. Given n bins of estimated poses, we learn how to match different poses corresponding to different bins. In the result, we learn $n * (n + 1)/2$ metrics. While learning metrics, we follow a well known scheme based on image pairs, containing two different poses of the same target as positives and pairs of different poses containing different targets as negatives. The learned metrics stand for the metric pool. This metric pool is learned offline and does not depend on camera pair. In the result, once metric pool is learned, it can be used for any camera pair.

Given two images from different (or the same) camera, we first estimate the poses for each image. Having two poses, we select a corresponding metric from the metric pool. The selected metric provides the strategy to compute similarity between two images (see figure 23).

6.12. Global Tracker : An Online Evaluation Framework to Improve Tracking Quality

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

keywords: online quality estimation, re-identification, tracking results improvements

This work addresses the problem of estimating the reliability of a tracking algorithm during runtime and correcting the anomalies found. Evaluating and tuning a tracking algorithm generally requires multiple runs and ground truth. The proposed framework called global tracker overcomes these limitations by combining an online evaluation algorithm and a recovering post-process.
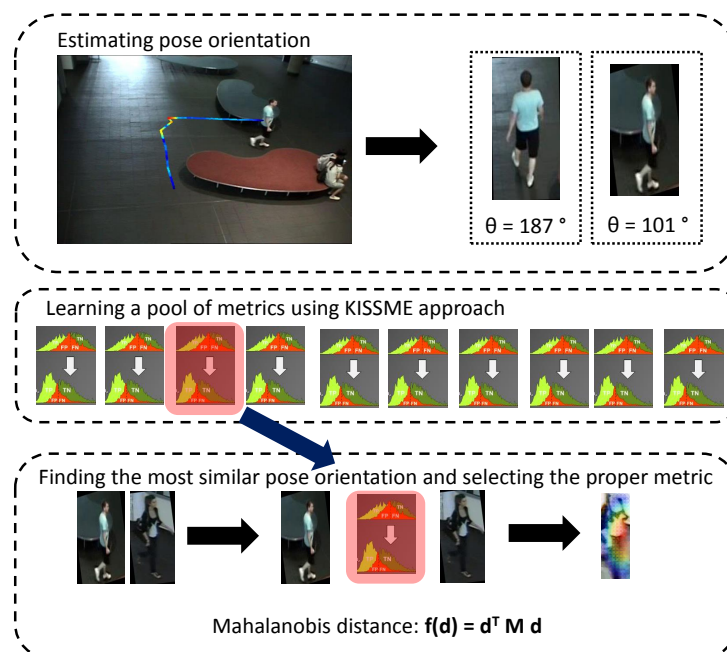
Figure 23. The proper metric is selected from the pool of previously learned metric for matching different poses.

Designing an evaluation framework that does not require ground truth has many different applications. One of them is to provide feedback to the tracking algorithm that can tune its own parameters to improve the results on the next frame. Another convenient application is to filter the reliable information from the tracking algorithm that can be used by the next processing step such as event recognition or re-identification.
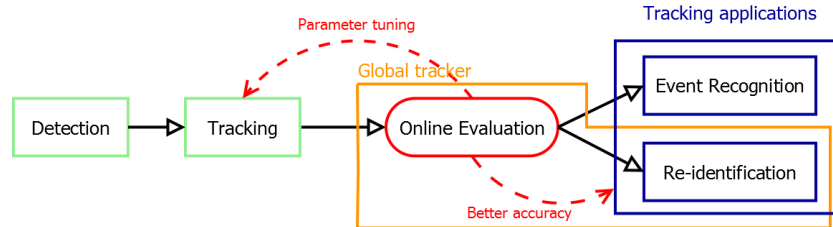
Figure 24. Integration of the global tracker, combining online evaluation and re-identification

The proposed online evaluation framework is based on control features. It means that several representative values or characteristics (the features) are chosen to monitor what is happening. Control features are the features that the online evaluation framework uses to monitor the status of the tracked objects. The framework is divided into two steps :

- computing the control features related to each tracked object of the scene on the current frame
- detecting the possible anomalies and then classifing them into two categories : anomalies due to tracking errors or benign anomalies (when a person leaves the scene or cross an obstacle for example).

This approach has been tested on two datasets (PETS 2009 and Caviar) with two different tracking algorithms (a multi-feature tracker and a tracker based on graph partitioning). The results show that the global tracker, even associated with a tracking algorithm that does not have good results, can perform as well as the state of the art.

Table 4. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset using CLEAR metrics

Methods	MOTA	MOTP
Heili et al. [72]	0.89	0.66
Zamir et al. [96]	**0.90**	0.69
Milan et al. [77]	**0.90**	**0.74**
Tracker 1	0.62	0.63
Tracker 1 + global tracker	0.85	0.71
Tracker 2	0.85	**0.74**
Tracker 2 + global tracker	**0.90**	**0.74**

Table 5. Tracking results on the Caviar dataset using Mostly Tracked (MT), Partially Tracked (PT) and Mostly Lost (ML) metrics

Method	MT (%)	PT (%)	ML (%)
Li et al. [76]	84.6	14.0	1.4
Kuo et al. [74]	84.6	14.7	**0.7**
Tracker 1	78.3	16.0	5.7
Tracker 1 + global tracker	**86.4**	**8.3**	5.3

This approach has been published in AVSS 2014 [33] which details the differences between real errors and benign anomalies.

6.13. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

keywords: Action Recognition; Human Action Recognition

This Ph.D. thesis targets the automatic recognition of human actions in videos. Human action recognition is defined as a requirement to determine what human actions occur in videos. This problem is particularly hard due to enormous variations in visual and motion appearance of people and actions, camera viewpoint changes, moving background, occlusions, noise, and enormous amount of video data.

Firstly, we review, evaluate, and compare the most popular and the most prominent state-of-the-art techniques, and we propose our action recognition framework based on local features, which we use throughout this thesis work embedding the novel algorithms. Moreover, we introduce a new dataset (CHU Nice Hospital) with daily self care actions of elder patients in a hospital.

Then, we propose two local spatio-temporal descriptors for action recognition in videos. The first descriptor is based on a covariance matrix representation, and it models linear relations between low-level features. The second descriptor is based on a Brownian covariance, and it models all kinds of possible relations between low-level features.

Then, we propose three higher-level feature representations to go beyond the limitations of the local feature encoding techniques.

The first representation is based on the idea of relative dense trajectories. We propose an object-centric local feature representation of motion trajectories, which allows to use the spatial information by a local feature encoding technique.

The second representation encodes relations among local features as pairwise features. The main idea is to capture the appearance relations among features (both visual and motion), and use geometric information to describe how these appearance relations are mutually arranged in the spatio-temporal space.

The third representation captures statistics of pairwise co-occurring visual words within multi-scale feature-centric neighbourhoods. The proposed contextual features based representation encodes information about local density of features, local pairwise relations among the features, and spatio-temporal order among features.

Finally, we show that the proposed techniques obtain better or similar performance in comparison to the state-of-the-art on various, real, and challenging human action recognition datasets (Weizmann, KTH, URADL, MSR Daily Activity 3D, HMDB51, and CHU Nice Hospital).
The Ph.D. thesis was defended on December 5, 2014.

6.14. Action Recognition Using 3D Trajectories with Hierarchical Classifier

Participants: Michal Koperski, Piotr Bilinski, François Brémond.

keywords: action recognition, computer vision, machine learning, 3D sensors

The goal of our work is to extend recently published approaches ([61], [93]) for Human Action Recognition to take advantage of the depth information from 3D sensors.

We propose to add depth information to trajectory based algorithms ([61], [93]). Currently mentioned algorithms compute trajectories by sampling video frames and then tracking points of interest - creating the trajectory. Our contribution is to create even more discriminative features by adding depth information to previously detected trajectories. In our work we propose methods to deal with noise and missing measurements in depth map.

The second contribution is a technique to deal with actions which do not contain enough motion to compute discriminative trajectory descriptors. Actions like sitting, standing, laptop use do not contain large amount of motion, or motion is occluded by the object. For such cases we proposed LDP (Local Depth Pattern) descriptor which does not require motion to be computed.

Proposed descriptors are further processed using a Bag of Words method and SVM classifier. We use hierarchical approach where at first level we train classifier to recognize if given example contains high or low amount of motion. Then at second layer we train SVM classifier to recognize action labels.

Figure 25. Visualization of MSR Dailiy Activty 3D data set (left) - video input frame, (center) - frame with detected trajectories (red - static points, green detected trajectories, (right) - corresponding depth map

The evaluation of our method was conducted on "Microsoft Daily Activity3D" data set [94] which consists of 16 actions (drink, eat, read book, call cellphone, write on a paper, use laptop etc.) performed by 10 subjects. We achieve superior performance among techniques which do not require skeleton detection. This work was published in proceedings of the 21st IEEE International Conference on Image Processing, ICIP 2014 [42]

6.15. Action Recognition using Video Brownian Covariance Descriptor for Human

Participants: Piotr Bilinski, Michal Koperski, Slawomir Bak, François Brémond.

keywords: action recognition, computer vision, machine learning

This work addresses a problem of recognizing human actions in video sequences. Recent studies have shown that methods which use bag-of-features and space-time features achieve high recognition accuracy [61], [93], [42]. Such methods extract both appearance-based and motion-based features. In image processing, a novel trend has emerged that ignores explicit values of given features, focusing instead on their pairwise relations. The most known example of such an approach is covariance descriptor [92]. Inspired by Brownian motion statistics [88] and application in people Re-identification [35]; we propose to model relationships between different pixel-level appearance features such as intensity and gradient using Brownian covariance, which is a natural extension of classical covariance measure. While classical covariance can model only linear relationships, Brownian covariance models all kinds of possible relationships. We propose a method to compute Brownian covariance on space-time volume of a video sequence. We show that proposed Video Brownian Covariance (VBC) descriptor carries complementary information to the Histogram of Oriented Gradients (HOG) descriptor. The fusion of these two descriptors gives a significant improvement in performance on three challenging action recognition datasets. The result of this work was published in proceedings of the 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014 [38].
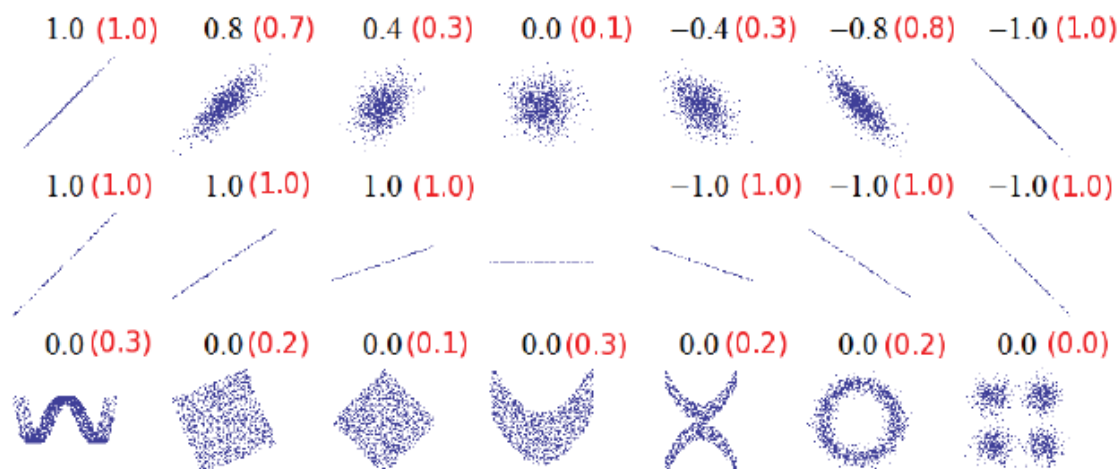
Figure 26. Comparison between Covariance and Brownian distance correlation. Covariance values in black, Brownian values in red.

6.16. Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance

Participants: Sofia Zaidenberg, Piotr Bilinski, François Brémond.

keywords: Event detection; Motion estimation; Anomaly estimation; Situation awareness; Scene Understanding; Group Activity Recognition; Stream Selection

We present a novel and unsupervised approach for discovering "sudden" movements in surveillance videos. The proposed approach automatically detects quick motions in a video, corresponding to any action. A set of possible actions is not required and the proposed method successfully detects potentially alarm-raising actions without training or camera calibration. Moreover, the system uses a group detection and event recognition framework to relate detected sudden movements and groups of people, and to provide a semantical interpretation of the scene. We have tested our approach on a dataset of nearly 8 hours of videos recorded from two cameras in the Parisian subway for a European Project. For evaluation, we annotated 1 hour of sequences containing 50 sudden movements.
This work has been published in [47].

6.17. Autonomous Monitoring for Securing European Ports

Participants: Vasanth Bathrinarayanan, François Brémond.

Keywords: Event Recognition, Port Surveillance

This work is done for the European research project SUPPORT (Security UPgrade for PORTs). This project addresses 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 is securing and efficient operation of European ports enabling uninterrupted flows of cargo and passengers while suppressing attacks on high value port facilities, illegal immigration and trafficking of drugs, weapons and illicit substances.

Scene understanding platform was tested on this new dataset, which has archived footage from past incidents and some acted scenarios. The processing pipeline of algorithms contains camera calibration, background subtraction using GMM (Gaussian Mixture Model), people detection using DPM (Deformable Parts Model), Tracking (Frame to Frame), Event recognition.

We collected several hours of videos which contained security related events like Intrusion to port by different methods (sea, gates, fences), Spying activities from outside the port, robbery or theft, ticketless travelling, restricted zone access, abondon luggage and some abnormal behaviors. The system was modelled and validated for all the above events to be detected and also a live real time demo was done for the completion of the project. All the events from our systems are later sent to project partners for fusion of the data with other sensors data and information from police, internal and external reports, etc., to detect complex security threats (see figure 27).
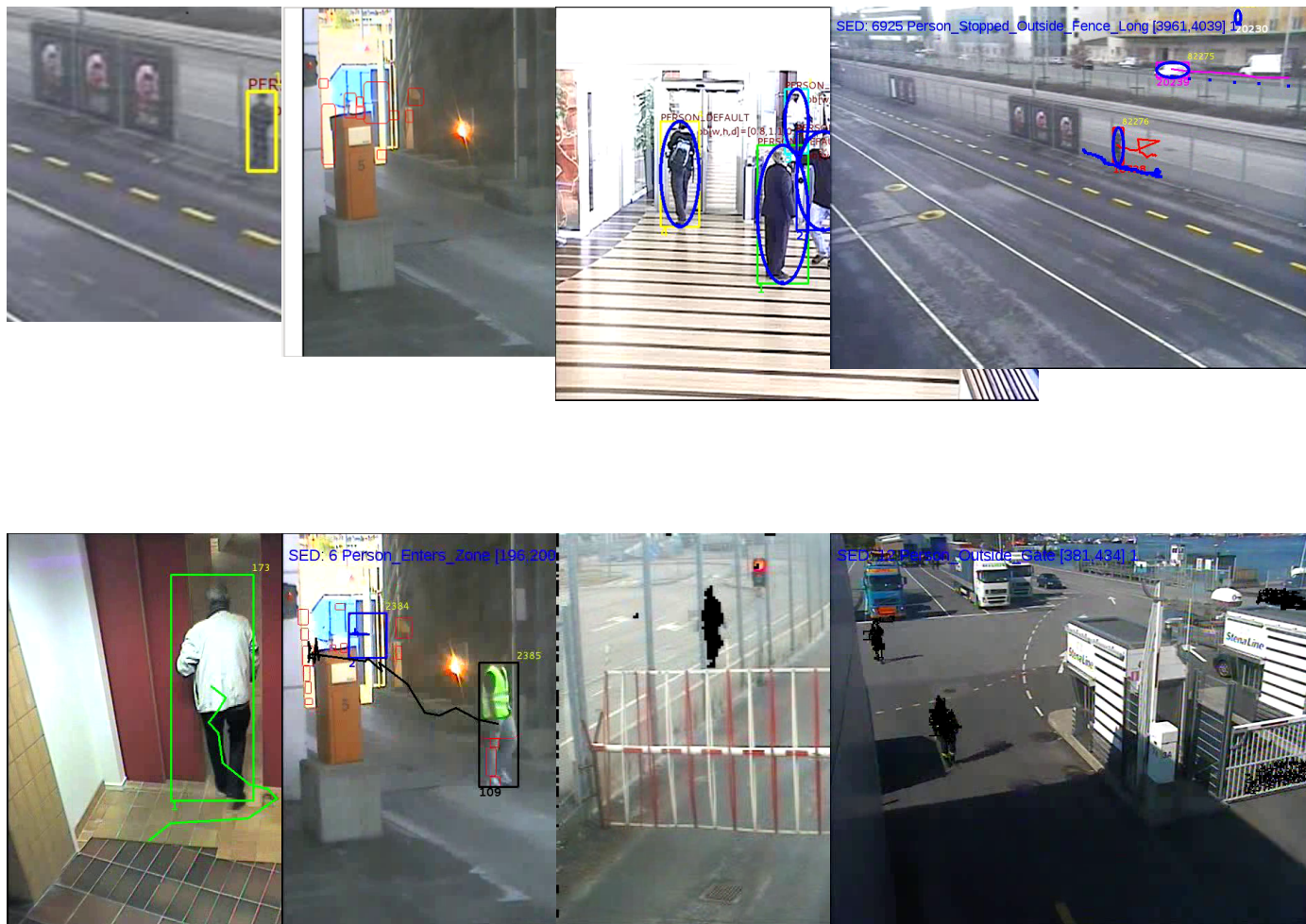

Figure 27. Detection of complex security threats from the tested dataset for Port surveillance

6.18. Video Understanding for Group Behavior Analysis

Participants: Carolina Garate, François Brémond.

keywords: Computer vision, group tracking, scene understanding, group behavior recognition, video surveillance, event detection.

The main work in this PhD thesis concerns the recognition of the behaviors of a group of people (2-5 persons) involved in a scene depicted by a video sequence.

Our goal focuses on the automatic recognition of behavior patterns in video sequence for groups of people (2-5 persons). We want to build a real time system able to recognize various group scenarios.

The approach includes different tasks to achieve the final recognition. The first one consists in tracking groups of moving regions detected in the video sequence acquired by the cameras. The second task attempts to classify these moving regions into people classes. Finally, the last task recognizes group scenarios using a priori knowledge containing scenario models predefined by experts and also 3D geometric and semantic information of the observed environment.

Our approach considers a chain process consisting of 5 consecutive steps for video processing. The steps are : 1) segmentation, 2) physical object detection, 3) physical objects tracking, 4) group tracking and 5) group behavior recognition. Our research focuses on the last two phases.

First, group scenarios have been defined (and then recognized) using the general scenario description language. Second, the likelihood of the group scenario recognition has been quantified. Third, machine learning techniques have been investigated to learn and recognize these scenarios.

We have processed the data set from 1 month video surveillance camera in the Torino subway and the Minds eye data set. Recognizing several and different events such as: walking groups, standing still groups, running groups, calm groups (i.e. having a bounding box with stable size), active groups (i.e. with bounding box's size variations, meaning that group members move a lot).

6.19. Evaluation of an Event Detection Framework for Older People Monitoring: from Minute to Hour-scale Monitoring and Patients Autonomy and Dementia Assessment

Participants: Carlos F. Crispim-Junior, Alvaro Gomez Uria Covella, Carola Strumia, Baptiste Fosty, Duc Phu Chau, Anh-Tuan Nghiem, Alexandra Konig, Auriane Gros, Philippe Robert, François Brémond.

keywords: RGBD cameras, description-based activity recognition, older people,

Two main works are reported here: the continuous evaluation and extention of our event detection framework for older people monitoring, and the proposal of a behavioral classification model for the assessment of autonomy and cognitive health level of older people using automatically detected events. The evaluation of our event monitoring framework was extended from 29 to 49 recordings of senior participants undertaking physical tasks (7 min per participant, total : 5.71 hours) and instrumental activities of daily living (IADL, 15 minutes per participant, total: 12.25 hours). The recordings have taken place in a ecological observation room set in the Memory Center of Nice hospital. In the extended evaluation we employed a RGBD sensor as input instead of a standard RGB camera due to its advantages like invariance to illumination changes and real-time measurements of 3D information which foster better performance of the underlying algorithms for people detection and tracking. Table 6 presents the event monitoring performance of the present system for 49 participants. Event detection performance on physical task generalized to the larger dataset with a small performance increase of 1.4% (average F-Score). Concerning IADL detection although the global performance value (F-Score, 80.7 %) are the same, the new approach have made a trade-off between recall and precision to obtain more reliable detection of activities and their parameter estimations. Low precision values on preparing drink (e.g., making coffee) and watering plant are due to these activities being performed in very close - if not overlapping - locations (contextual zones). Low precision values in reading are due to the preferred reading location be close to image edges where most parts of person body are frequently outside the camera field of view.

Using the event monitoring system as input we have devised a behavioral classification model for the automatic assessment of participant cognitive health and autonomy level. Besides to event data the model also uses fine-grained data about person gait attributes (*e.g.*, stride-length, cadence, etc), obtained by a RGBD-based algorithm for gait analysis also developed in STARS team. Briefly, the event monitoring system supports the doctor by automatically annotating the patient daily living activities and assessing his/her gait parameters in a quantitative way, and the behavioral model performs the classification of participant's dementia and autonomy levels as a complement for standard psychometric scales for autonomy. We achieved an average accuracy of 83.67 % at the prediction of patient autonomy (poor, mediocre, good), and of 73.46 % for cognitive level class (healhy, memory cognitive impairment - MCI, alzheimer's disease), all models using a Naíve Bayes classifier. The results suggest that the behavioral classification model using automatically detected events outperforms the same model using events manually annotated by domain experts (81 %). On the contrary, the model using annotated data still outperforms the automated detection at dementia classification (79.46 %). Results indicate it is easier to predict the autonomy level than the Dementia, since the latter may be seen as the cause /source and the first its consequences. Deciding whether a decay on cognitive abilities relates to normal aging or early MCI or a given mild cognitive decay is an early symptom of Alzheimer's diasease or a severe case of MCI is also a open-problem for medical community. Future work will focus on investigating whether the remaining performance to achieve is related to the performance failures of the underling event monitoring system, to important behavioral aspects still not covered by the behavioral model, or even to the inherently ambiguous nature of the dementia classes.

Table 6. Event Monitoring Performance

Physical Tasks	Recall	Precision
Single Task	100%	88%
Dual Task	100%	98%
IADLs	Recall	Precision
Preparing drug box	87%	93%
Watering plant	80%	63%
Reading	60%	88%
Prepare drink	90%	68%
Talk on phone	89%	89%

We have also started the evaluation of the event monitoring system in Nursing home scenario passing from a minute time-scale to hours. A first participant was monitored with two RGBD sensors, one for bed-related events (sleep, bed exits) and one for living room and daily living activity events for 14 days. Preliminary results are 80 % for entering in bed and 100 % for bed exit in set of 6 events of each class in 13 hours monitoring (6 pm - 7am). Figure 28 illustrates the detection of restroom usage during the night. The automatic monitoring of participant activities during night is an important contribution to medical/nursing staff as wandering behavior at night is a common cause of accident in older people population. For instance, detecting whether a bed-exit during the night will be followed by a restroom visit or a bedroom exit plays a significant role at predicting a possibly dangerous situation.

Two papers are envisaged to report the results of this year research to scientific community, one describing the new version of the event monitoring system, and a second one for the developed behavioral classification model. As a publication of this year we hightlight the paper in partnership with Alexandra Konig and Philippe Robert - entitled Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients - in the Journal of Alzheimer disease where we summarize the results of the validation of our event monitoring system for the recognition of activities of daily living of participants of Alzheimer's disease study.

6.20. Uncertainty Modeling Framework for Constraint-based Event Detection in Vision Systems
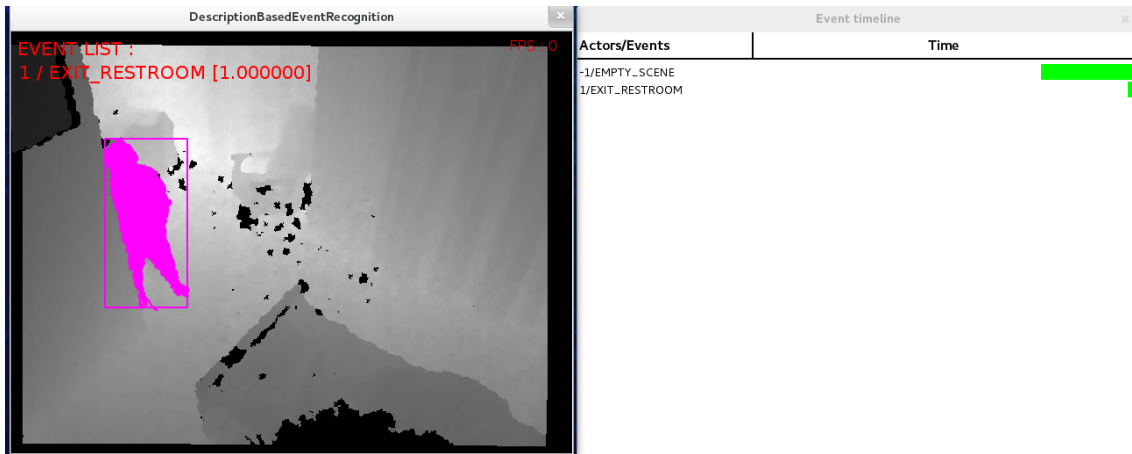
Figure 28. Long-term monitoring of Older People: Sleep-related events

Participants: Carlos F. Crispim-Junior, François Brémond.

keywords: description-based activity recognition, uncertainty modeling, vision system, older people

Event detection has advanced significantly in the past decades relying on pixel- and feature-level representations of video-clips. Although effective, those representations have difficulty on incorporating scene semantics. Alternatively, ontology and description-based approaches for event modeling can explicitly embed scene semantics, but the deterministic nature of such languages is susceptible to noise from underlying components of vision systems. We have developed a probabilistic framework to handle uncertainty on our constraint-based ontology framework for event detection. This task spans from elementary scenarios uncertainty handling (from low-level data and event intra-class variance) to complex scenario semantic modeling, where time ordering in between event sub-components and the effect of missing components (for instance, due to miss-detection) plays a significant role.

Preliminary results of this work have been published in [40], where the presented formalism for elementary event (scenario) uncertainty handling is evaluated on the detection of activities of daily living of participants of the Alzheimer's disease study of Nice hospital using the newest version of our vision system using a RGB-D sensor (Kinect®, Microsoft©) as input. Two evaluations have been carried out: the first one, (a 3-fold cross-validation) focuses on elementary scenario constraint modeling and recognition, and the second one was devoted to complex scenario recognition following a semi-probabilistic approach (n:45).

Table 7 presents the performance of the uncertainty modeling framework on elementary scenario (primitive state) detection for $N : 10\ participants$; $15\ min.\ each$; $Total : 150\ min$. The 3-fold cross-validation scheme (n:10 participants) is employed for constraint probabilistic distribution learning and event detection evaluation on 10 RGB-D recordings of participants of the Nice hospital clinical protocol for Alzheimer's disease study. "Crisp" term stands for our deterministic constraint-based ontology language for event modeling. Results are reported as the average performance on the crisp and uncertainty frameworks on the validation sets. Results confirm that the uncertainty modeling improves the detection of elementary scenarios in recall (*e.g.,* In zone phone: 84 to 100 %) and precision indices (*e.g.,* In zone Reading: 54.5 to 85.7%).

Table 8 presents the performance of the proposed framework on Composite Event Detection for $N : 45\ participants$; $15\ min.\ each$; $Total : 675min$. Here a hybrid strategy is adopted where the uncertainty modeling is used for elementary scenarios and the crisp constraint-based framework is used for composite event modeling. Results show improvement on recall index of event detection performance, but

Table 7. Framework Performance on Elementary Scenario Detection on a 3-fold-cross-validation scheme

	Crisp		Uncertainty	
IADL	Rec.	Prec.	Rec.	Prec.
In zone Pharmacy	100.0	71.4	100	**83.3**
In zone Phone	84.0	95.45	**100.0**	**100.0**
In zone Plant	100.0	81.8	100.0	81.8
In zone Tea	93.3	**77.7**	93.3	73.7
In zone Read	75.0	54.5	75.0	**85.7**

the uncertainty framework performance on precision index is still worse than the crisp approach. The latter performance may be attributed to the crisp constraints that did not have their uncertainty addressed yet.

Table 8. Framework Performance on Composite Event Detection Level

	Crisp		Uncertainty	
IADL	Rec.	Prec.	Rec.	Prec.
Talk on Phone	88.76	89.77	**88.76**	85.86
Preparing Tea/Coffee	81.42	40.36	**92.85**	**55.08**
Using Pharmacy Basket	87.75	95.65	**89.79**	**97.77**
Watering plant	78.57	**84.61**	**100.0**	28.86

Future work will focus on modeling complex scenario constraints such as time ordering and missing components, and on extending the set of low-level uncertainties which are addressed. Moreover, we have been conducting a joint work with partners of Dem@care project to evaluate the uncertainty framework for multiple sensor fusion at decision level. Currently, processed data from different visual modalities (standard RGB, RGBD, and wearable cameras) have been gathered for 17 participants of Nice hospital pilot@lab, and preliminary results are expected for the first semester of 2015.

6.21. Assisted Serious Game for Older People

Participants: Minh Khue Phan Tran, François Brémond, Philippe Robert.

keywords: interactive system, elderly people, serious game

A system able to interact with older people has been recently devised. The system consists of two parts: Recognition and Interaction. Recognition part, requiring an Asus Xtion Pro Live Camera, consists in observing the scene to decide when is the best moment to interact with users. Afterwards, the Interactive system tries to engage the patient via an interface and through Microsoft Kinect Camera, the patient can interact with the interface using voice or gesture. The interface is designed with Unity 3D game engine (see figure 29).

An experiment was conducted in a memory center for older people, Institut Claude Pompidou in Nice, in order to test different functionalities of the system. Here, participants can experiment the system in a private room (see figure 30) equipped with a large screen and can start the game without having to use devices (mouse, keyboard). The "best moments" to interact with participants are defined when they stay more than 5 seconds in front of the screen. Once these moments are recognized, the interface of Interactive part is called. The avatar indicates the place to be for playing and starts the game.

19 older people have participated to the experiment. 16 succeeded to follow the indications of avatar up to the start of the game. Most of them have appreciated the interaction with the avatar. Even 13 of them have continued to play the second game after suggestion of the avatar.

Future work aims at looking at other indicators (behavior, gaze) that the system can rely on to improve user's interaction.
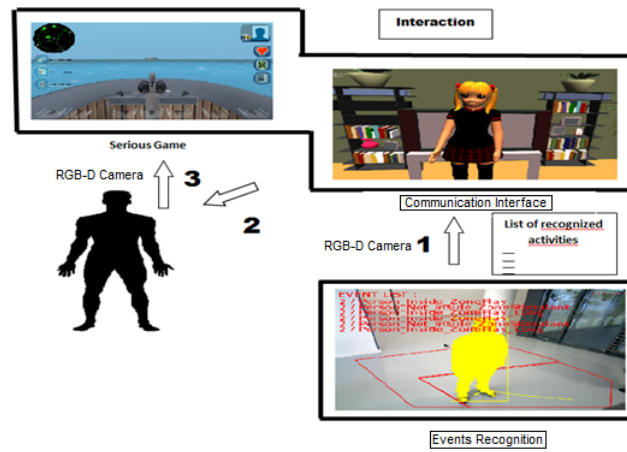
Figure 29. Functional diagram of our interactive system

Figure 30. Experimental private room

6.22. Enhancing Pre-defined Event Models Using Unsupervised Learning

Participants: Serhan Coşar, François Brémond.

keywords: Pre-defined activity models, unsupervised learning, tailoring activity models

In this work, we have developed a new approach to recognize human activities from videos, given models that are learned in an unsupervised way and that can take advantage of *a priori* knowledge provided by an expert of the application domain. The description-based methods use pre-defined models and rules to recognize concrete events. But, if the data has unstructured nature, such as daily activities of people, the models cannot handle the variability in data (e.g., the way of preparing meal is person dependent).

In order to overcome this drawback, we have combined the description-based method in [66] with an unsupervised activity learning framework, as presented in Figure 31. We have created a mutual knowledge loop system, in which both frameworks are combined in a way to compensate their individual limitations. In [66], scene regions are pre-defined and the activity models are created via defining an expected duration value (e.g., 2 seconds) and a posture type (e.g., standing) by hand. Thus, these hand-crafted models fail to cover the variability in data and require an update by experts whenever the scene or person changes. To automatically define these parameters, we utilize the unsupervised activity recognition framework. The unsupervised approach first learns scene regions (zones) in the scene using trajectory information and then, it learns the duration and posture distribution for each zone. By matching the pre-defined zones with learned zones, we connect the learned parameter distributions with hand-crafted models.

The knowledge is passed in a loopy way from one framework to another one. By knowledge we mean: (i) the geometric information and scene semantics of the description-based system are used to label the zones that are learned in an unsupervised way, (ii) the activity models that are learned in an unsupervised way are used to tune the parameters (i.e. tailoring) in the activity models of the description-based framework. It is assumed that the person detection and tracking are already performed and we have the trajectory information of people in the scene beforehand.
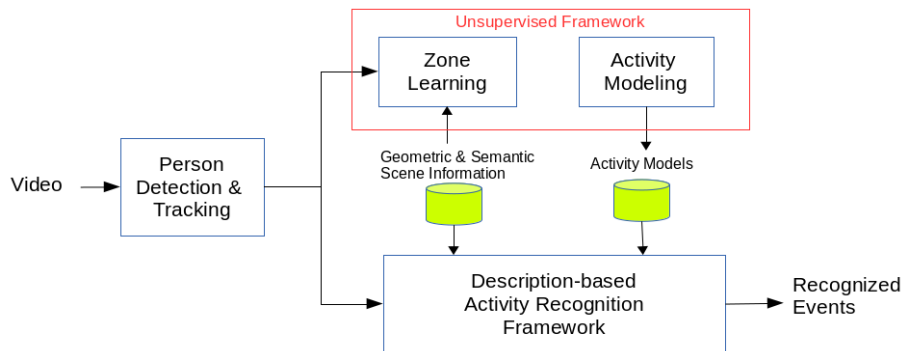
Figure 31. The flow diagram of the framework that combines pre-defined activity models with unsupervised models.

We have tested the performance of the knowledge-loop based framework on two datasets: i) Hospital-RGB, ii) Hospital-RGBD. Each dataset contains one person performing everyday activities in a hospital room. The activities considered in the datasets are "watching TV", "preparing tea", "answering phone", "reading newspaper/magazine", "watering plant", "organizing the prescribed drugs", "writing a check at the office desk" and "checking bus routes in a bus map". Each person is recorded using RGB and RGBD cameras of 640×480 pixels of resolution. RGB dataset consists of 41 videos and RGBD dataset contains 27 videos. For each person, video lasts approximately 15 minutes.

The performance of the approach in [66] with hand-crafted models and our approach with learned models for Hospital-RGB and Hospital-RGBD datasets are presented in Table 9 and in Table 10 . The results have been partially presented in Ellomiietcv2014 (**waiting hal acceptation**). It can be clearly seen that updating the constraints in activity models using data learned by the unsupervised approach enables us to detect activities missed by the pre-defined models. For "watching TV" and "using pharmacy basket" activities in RGB dataset and "answering phone" and "preparing tea" activities in RGBD dataset, there is increase in false positive rates. The reason is that, for some activities, the duration and posture distributions learned by the unsupervised approach can be inaccurate because of other actions occurring inside a zone (e.g., a person standing inside tea zone and reading). For this reason, the constraints updated in activity models are too wide and other activities that occur inside the zone are also detected. Despite the small increase of false positives in some activities, we have increased the true positive rates and obtained sensitivity rates around 90% and 87% on average in RGB and RGBD datasets, respectively, and precision rates around 81% on average in RGBD dataset. Thanks to the distributions learned for time duration and posture, we can enhance the activity models that are manually defined in the description-based, and thereby detect missed events.

Table 9. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGB. The bold values represent the best result for each activity class.

	Hand-crafted Models		Unsupervised Models	
ADLs	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
Answering Phone	70	82.35	**95**	**90.47**
Watching TV	84.61	78.57	**100**	54.16
Using Office Desk	91.67	47.82	91.67	**52.38**
Preparing Tea	**80.95**	70.83	76.19	**80**
Using Phar. Basket	100	**90.90**	100	76.92
Watering Plant	100	81.81	88.89	**88.89**
Reading	45.46	83.34	**81.82**	**90**
TOTAL	81.81	**76.52**	**90.5**	76.11

Table 10. The activity recognition results of the description-based approach and our knowledge-loop based approach for the Hospital-RGBD datasets. The bold values represent the best result for each activity class.

	Hand-crafted Models		Unsupervised Models	
ADLs	Sensitivity (%)	Precision (%)	Sensitivity (%)	Precision (%)
Answering Phone	80	**100**	**84.21**	88.89
Watching TV	55.56	45.46	**77.78**	**58.34**
Preparing Tea	100	**73.68**	92.85	65
Using Phar. Basket	100	90	100	**100**
Watering Plant	40	66.67	**83.34**	**71.42**
Reading	**100**	66.67	71.42	**83.34**
Using Bus Map	50	71.42	**100**	**100**
TOTAL	75.07	73.41	**87.08**	**80.99**

6.23. Using Dense Trajectories to Enhance Unsupervised Action Discovery

Participants: Farhood Negin, Serhan Coşar, François Brémond.

keywords: zone learning, action descriptors, dense trajectories, supervised action recognition, unsupervised activity recognition

The main purpose in this work is to monitor older people in an unstructured scene (e.g., home) and to recognize the types of activities they perform. We have extended the work in Ellomiietcv2014 that was basically an unsupervised method to learn behavioral patterns of individuals without restraining subjects to follow a predefined activity model. The main concern in previous work is to find different zones in the scene where activities take place (scene topology) by employing trajectory information provided by tracking algorithm. The previous work in Ellomiietcv2014 (**waiting hal acceptation**) proposes a Hierarchical Activity learning Model (HAM) to learn activities based on previously identified topologies. The current work examines the same potential while first, incorporating image descriptors [93] in a bag-of-word representation to differentiate actions in a supervised manner and second, combining the two approaches (supervised and unsupervised) to provide clues about actions inside each zone by classifying retrieved descriptors using a classifier.

Recently, dense trajectories are widely used for action recognition and have been shown state-of-the-art performance [93]. For the purpose of the current work, we use HOG and HOF descriptors for supervised action recognition. Figure 32 shows a general description of the supervised framework. For the learning phase, the dense trajectories are extracted from input images coming from RGBD camera. Following Ellomiietcv2014, three-level topology of the scene is constructed by trajectory information coming from tracking algorithm [62]. The topology is used to split input video stream into chunks by checking where the person is with respect to the learned zones. Then, for every video chunk, dense descriptors are extracted and stored. A codebook representation is obtained by applying a k-means clustering algorithm on the whole set of extracted features. Next, the action histograms are calculated by employing the codebook. A SVM classifier is trained and stored to use in test phase via calculated histograms.

Figure 32. Flow diagram for supervised action recognition.

In recognition phase, we similarly split the test videos by comparing each trajectory point with learned topologies, extract the descriptor for each split, and the histograms are calculated via k-NN using the codebook generated in learning phase. Then, the histograms are classified using the trained SVM classifier and resulting labels are evaluated by comparing with the ground truth.

We have assessed the performance of the supervised activity recognition framework using 183 video splits of 26 subjects. We divided the video dataset to training and testing groups. Training set includes 93 videos of 15 subjects and the test set includes 90 videos of 11 subjects. Notice that the number of videos is counted after splitting process has been done on input data. We used the videos recorded from CHU Nice hospital while real patients are visiting their doctors and are asked to perform several activities in specified locations of the room. The activities we considered in our tests include: "preparing tea", "watching TV", "using phone", "reading on chair", "using pharmacy", and "using bus map". For RGB-D camera, we have used the person detection algorithm in [79] and tracking algorithm in [62]. The classification results for using HOG and HOF descriptors and corresponding confusion matrices are depicted in Table 11 and in Table 12. For SVM classifier, we used RBF kernel.

As a future work, we are going to benefit from the action descriptors to discriminate different activities occurring in the same zone.

6.24. Abnormal Event Detection in Videos and Group Behavior Analysis

Participants: Giuseppe Donatiello, Vania Bogorny, Serhan Cosar, Luis Campos Alvares, Carolina Garate, François Brémond.

Table 11. Confusion matrix for recognition results for HOG descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	11	0	0	0	0	0
2 Preparing Tea	0	18	0	0	0	0
3 Reading in Chair	1	0	10	0	0	0
4 Using Bus Map	0	0	0	14	0	0
5 Using Pharmacy Basket	0	0	0	0	10	0
6 Using Phone	0	0	0	0	0	25
Total	98.89%					

Table 12. Confusion matrix for recognition results for HOF descriptor

Activity Names	1	2	3	4	5	6
1 Watching TV	4	1	1	4	0	1
2 Preparing Tea	0	5	0	5	0	8
3 Reading in Chair	1	0	2	4	0	4
4 Using Bus Map	0	0	0	13	0	1
5 Using Pharmacy Basket	1	0	0	0	9	0
6 Using Phone	0	1	1	5	0	19
Total	57.78%					

keywords: activity recognition, abnormal events, group behavior analysis, trajectory clustering

This work addresses two different issues: (i) abnormal event detection and (ii) group behavior analysis in videos.

6.24.1. Abnormal Event Detection

For abnormal event detection we are proposing a fused approach that combines trajectory-based and pixel-based analysis. In this work we first discover the activity zones based on object trajectories, and we investigate abnormal events considering objects that move in wrong direction and/or with abnormal speed. Second, inside each zone we extract dense tracklets and using the clustering technique we discover different types of actions, and are able to distinguish between normal and abnormal actions inside each zone.

While existing approaches for abnormal behavior detection do either use trajectory based or pixel based methods, we propose a fused solution which can detect simple abnormal behavior based on speed and direction, as well as more complex behavior as abnormal activities. In a first step we automatically learn the zones of the scene where most activities occur, by taking as input the trajectories of detected mobiles, analyzing then statistical information of each mobile in each zone (speed and direction), through the use of a scale-resolution analysis. This approach implies a considerable complexity decrease of having huge data set and then an extensive impact of the algorithm speed, without losing useful information. Figure 33 shows an example of this first part.

The next step concerns a pixel based analysis inside each zone. This step takes as input each zone computed in the previous step and the bounding box of the object trajectories, and extracts action descriptors inside the bounding box of each object trajectory inside the zone. With this step we obtain the different body movements of each detected mobile inside a zone. By clustering the body motions and using Bags of Words, we detect different types of abnormal activities inside each zone. Figure 34 shows an example of what just mentioned.

Figure 33. Example of Trajectories (left), Trajectories over the grid (center), that represents a given scale resolution, Zones discovering (right)

Figure 34. Tracked object

The last step of our approach is a clustering operation of all information gathered in the previous two steps, that is for each mobile, speed, direction, and body movements-actions in each zone are applied to discriminate between different types of abnormal behavior in the scene. A flow diagram of our approach is presented in Figure 35.
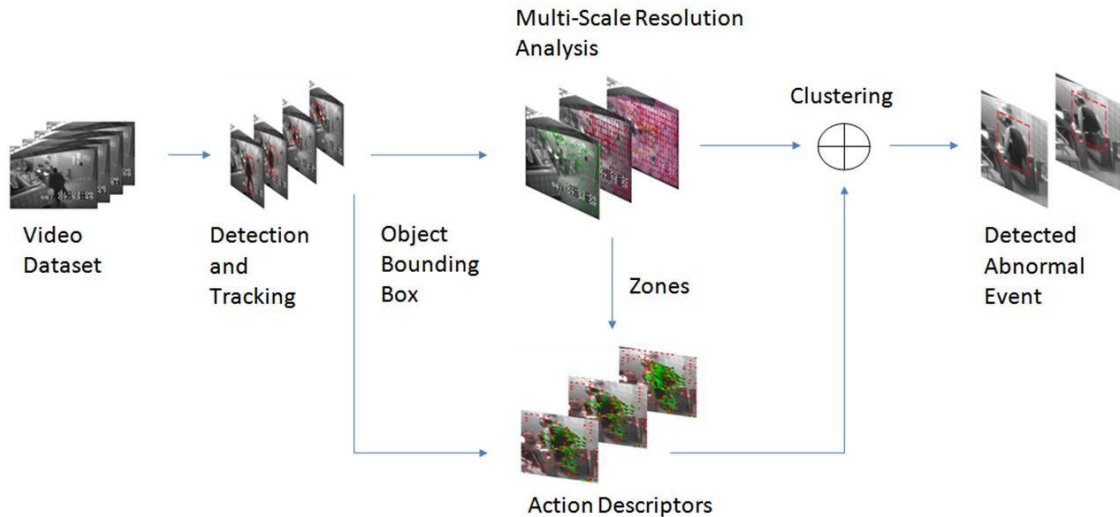
Figure 35. Flow Diagram of our approach

We have tested our approach on several real videos recorded. We show with experiments on two open datasets that our approach is able to detect several types of abnormal behavior.

6.24.2. Group Behavior Analysis

Group behavior analysis is focused on the extraction of groups based on object trajectories and the analysis is performed over the dense tracklets, computed for the groups bounding boxes. From the analysis of the dense tracklets we detect different levels of agitation. These works are ongoing and have not yet been published.

6.25. Model-Driven Engineering for Activity Recognition Systems

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

We continue to explore the applicability of model driven engineering (MDE) to activity recognition systems. Of course, setting up a complete methodology is a long term objective.

6.25.1. Feature Models

Features models are convenient representations of system variability but the drawback is a risk of combinatorial explosion of the number of possible configurations. Hence we have extended feature models with quality attributes and associated metrics to facilitate the choice of an optimal configuration, at deployment as well as at run time. We have proposed several strategies and heuristics offering different properties regarding optimality of results and execution efficiency [41].

This year we have conducted further experiments to evaluate the optimization algorithm and the metrics. In particular, we studied the prediction accuracy of the additive metrics for estimating two properties of interest: frame processing time and reconfiguration time. The goal was to compare predicted against measured properties of a running system. We used a simple video chain implemented with OpenCV components (acquisition, filtering, various detections, and visualisation) and we tested it on a video sample of 48s (i.e., about 1350 frames). We defined a feature model for this chain, which exhibits 14 valid configurations.

We first computed the properties of each component in isolation (based on a set of repetitive measurements), then we measured the actual frame processing time and reconfiguration time, and finally we compared the estimated and actual values.
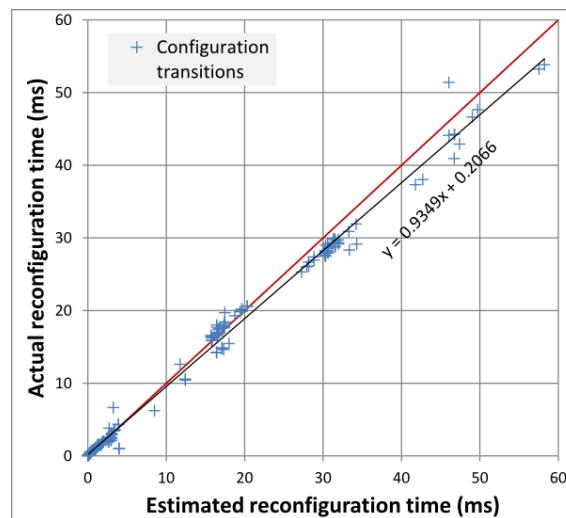

Figure 36. Measured and estimated reconfiguration time. Black line represents regression curve and red line the ideal trend

Figure 36 displays the results for reconfiguration time. In our experiment, we have a total of 210 transitions over a set of 15 valid configurations (14 system configuration + one for system shutdown). For frame processing time, we achieved an accuracy of 96.7% on average, and for reconfiguration time the accuracy was between 90.5% and 87.6%.

Introducing metrics in feature models is a precious help to reduce the huge set of valid configurations after a dynamic context change and to provide a real time selection of an appropriate running configuration. However, more evaluation remains to be done on other runtime properties and for other video chains.

6.25.2. Configuration Adaptation at Run Time

To react to environment changes we favor the "model at run-time" approach. Our current prototype ressorts to tools and technologies which were readily available. This made possible a proof of concepts. However, this induces several redundant representations of the same data, consistency problems, coexistence of several formalisms or programming languages, and superfluous back and forth inter-module communications. This year we started to design a more homogeneous and better integrated prototype. The two key points are, first, a component management framework, second, a feature model managemen tool.

This year, we have continued to develop our OSGi-like component framework, but more adapted to real time and compatible with our extended feature models. Concerning feature model management, we started to

study a FAMILIAR replacement that integrates smoothly into the C++ video analysis chain instead of being a separate (Java) tool. Moreover, the new tool should cope with our feature extensions (e.g., quality attributes).

6.26. Scenario Analysis Module

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

Keywords: Synchronous Modelling, Model checking, Mealy machine, Cognitive systems.

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. 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 [80] 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.

Setting up our tools on top of an existing language such as LUSTRE was convenient for rapid prototyping. However, it appeared delicate for efficiency reasons on the one hand, but also because it is a closed environment, difficult to customize. Hence we developed our own language LE and its environment CLEM (see section 6.27). This year, we focus on the expression of scenario models in CLEM through the internship of Omar Adballa [51] and we define in CLEM a specific back end to generate recognition engines (see figure 37). However, mastering all aspects of this environment will allow the user scenario description language to rely directly on the semantics of LE and not on its syntax. This reduces the number of necessary translations.

Currently, SAM implements an "exact" algorithm in the sense that it generates, at each instant, *all* possible scenario instances although many of them will freeze, still holding system resources. We have started scalability studies to evaluate the risk of combinatorial explosion. In parallel we enriched the synchronous scenario descriptions to reduce the number of generated scenario instances as well as the number of instances to awake at each instant. We are currently modifying our recognition engine generator to take advantage of this supplementaty information.

6.27. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé, Mohamed Bouatira, Ines Sarray.

Keywords: Synchronous languages, Synchronous Modelling, Model checking, Mealy machine.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see Figure 37) 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 focus on the improvement of both LE language and compiler concerning data handling and in the generation of back-ends required by other research axis of the team (see 6.26 and 6.28). We also improve the design of a new simulator for LE programs which integrates our new approach.

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). Last year, we chose an algebra which is a bilattice and we showed that it is well suited to solve our problem. To compute the partial orders of equation systems, we introduced two ways : a CPM (Critical Path Method) like algorithm, efficient and a "fix point" approach that allows us to show that we can compute partial orders locally and in an incremental way, or globally (thanks to the uniqueness if fix points). We introduced "abstract" equation systems and our method allows us to compute their partial orders.

We defined a new intermediate format LEA (see figure 37) to record these sorted abstract eaquation systems and they will be expanded latter in concrete equation systems with a refinement operation. We apply this technique to the compilation of Grafcet language. Our work in under publication in two journal papers.

In CLEM, we added types and data a few years ago and this year we complete this addition, we know can express automata where control and output signals are valued. From last year, we rely on CLEM both to design SAM (see section 6.26) and to perform validation in a component-based middleware (see section 6.28). To this aim, we generate now two specific output formats dedicated to these applications [54]

Finally, in CLEM, we generate an independent intermediate code (LEC) before specific target generations. This code represents the semantics of programs with 4-valued equation systems. In our design flow, we need to simulate programs at this level. This year, we complete a simulator begun last year but which did not integrate the data part of the language. The simulator GUI has been designed again in Qt and the simulator takes into account the values carried by signals. This work has been done by Mohamed Bouatira during his internship.
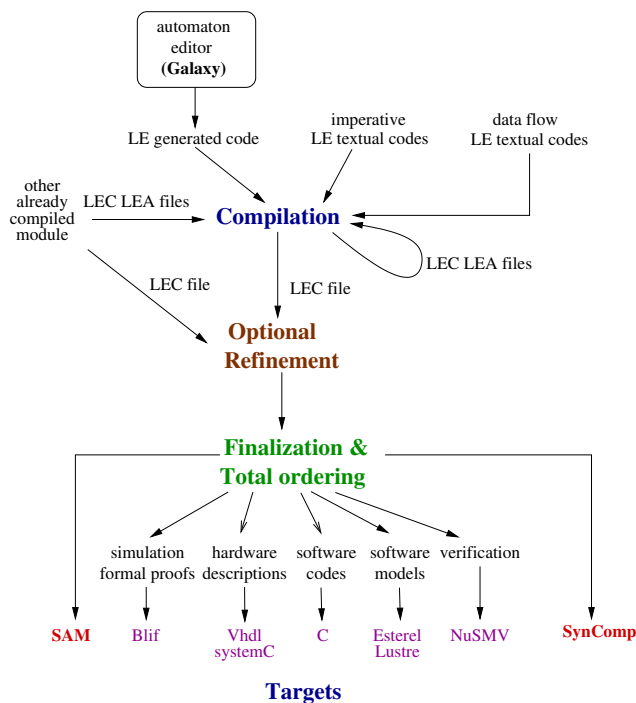
Figure 37. The CLEM *toolkit*

6.28. Multiple Services for Device Adaptive Platform for Scenario Recognition

Participants: Annie Ressouche, Daniel Gaffé, Ines Sarray, Jean-Yves Tigli.

Keywords: Synchronous Modelling, Model checking, Mealy machine, Ubiquitous Computing.

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 [91], [90], [73]) 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 behavioral model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

In [83], [82], we showed that an efficient means to define the synchronous components which allow to validate critical component behaviours, is to specify them with Mealy machines. Previously, we used a classical synchronous language (Lustre) to specify synchronous components, but the integration of the synchronous component code into WComp was not straightforward because Lustre compiler is not opened and cannot integrate new target code needed by WComp. This year, we rely on CLEM to design synchronous monitor as CLEM automata and we extend CLEM to generate automatically the internal code of WComp(C#).

On another hand, for some critical components, we can be led to introduce several synchronous monitors, each of them being related to a sub assembly. Then, some outputs of these synchronous monitors can be linked to the same input of a critical component. To face this problem, we introduced in [83], [82] a composition under constraints operation, which composed all the synchronous monitors linked to a critical component according to a set of constraints telling us how the respective outputs of monitors which access the same input are combined. We proved that this operation preserved safety properties, however it cannot ensure adaptivity and incrementality. This year, we have introduced a new way to compose synchronous monitors. We introduce an additional monitor (called constrain monitor) which express as a CLEM Mealy machine (set of equations) the constraints and we perform a usual parallel composition of all the monitors. Moreover, we define a syntactic mean (language DCL) to describe constraints in a generic way and we can derive automatically the constraint monitor for each specific application. In such a setting, we are able to manage the appearance and the desappearance of synchronous monitors.

Moreover, in middleware as WComp, communication is asynchronous while synchronous monitors respect the synchrony paradigm. Thus, we need to introduce in the middleware a means to transform asynchronous events into synchronous entries for synchronous monitors and the opposite to connect again the synchronous events into asynchronous design. To this aim, we introduced in WComp, specific components that receive asynchronous events and generate synchronous ones. Indeed, the part of this component is to decide how asynchronous events will be gather into a synchronous instant (logical time). To this aim, the user can choose between different politics. Then, all the events composing an instant are serialized and deserialized by the synchronous monitor. The desynchronisation operation performs the opposite way[54].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- **Toyota europ**: this project with Toyota runs from the 1st of August 2013 up to 2017 (4 years). It aims at detecting critical situations in the daily life of older adults living home alone. We believe that a system that is able to detect potentially dangerous situations will give peace of mind to frail older people as well as to their caregivers. This will require not only recognition of ADLs but also an evaluation of the way and timing in which they are being carried out. The system we want to develop is intended to help them and their relatives to feel more comfortable because they know potentially dangerous situations will be detected and reported to caregivers if necessary. The system is intended to work with a Partner Robot (to send real-time information to the robot) to better interact with the older adult.

- **LinkCareServices**: this project with Link Care Services runs from 2010 upto 2014. It aims at designing a novel system for Fall Detection. This study consists in evaluating the performance of video-based systems for Fall Detection in a large variety of situations. Another goal is to design a novel approach based on RGBD sensors with very low rate of false alarms.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. MOVEMENT

Program: ANR CSOSG

Project acronym: MOVEMENT

Project title: AutoMatic BiOmetric Verification and PersonnEl Tracking for SeaMless Airport ArEas Security MaNagemenT

Duration: January 2014-June 2017

Coordinator: MORPHO (FR)

Other partners: SAGEM (FR), Inria Sophia-Antipolis (FR), EGIDIUM (FR), EVITECH (FR) and CERAPS (FR)

Abstract: MOVEMENT is focusing on the management of security zones in the non public airport areas. These areas, with a restricted access, are dedicated to service activities such as maintenance, aircraft ground handling, airfreight activities, etc. In these areas, personnel movements tracking and traceability have to be improved in order to facilitate their passage through the different areas, while insuring a high level of security to prevent any unauthorized access. MOVEMENT aims at proposing a new concept for the airport's non public security zones (e.g.customs control rooms or luggage loading/unloading areas) management along with the development of an innovative supervision system prototype.

8.1.1.2. SafEE

Program: ANR TESCAN

Project acronym: SafEE

Project title: Safe & Easy Environment for Alzheimer Disease and related disorders

Duration: December 2013-May 2017

Coordinator: CHU Nice

Other partners: Nice Hospital(FR), Nice University (CobTeck FR), Inria Sophia-Antipolis (FR), Aromatherapeutics (FR), SolarGames(FR), Taichung Veterans General Hospital TVGH (TW), NCKU Hospital(TW), SMILE Lab at National Cheng Kung University NCKU (TW), BDE (TW)

Abstract: SafEE project aims at investigating technologies for stimulation and intervention for Alzheimer patients. More precisely, the main goals are: (1) to focus on specific clinical targets in three domains behavior, motricity and cognition (2) to merge assessment and non pharmacological help/intervention and (3) to propose easy ICT device solutions for the end users. In this project, experimental studies will be conducted both in France (at Hospital and Nursery Home) and in Taiwan.

8.1.2. Investment of Future

8.1.2.1. 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 (FR), Inria(Stars), CMRR (CHU Nice) and CobTek(Nice University).

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.

8.1.3. Large Scale Inria Initiative

8.1.3.1. PAL

Program: Inria

Project acronym: PAL

Project title: Personally Assisted Living

Duration: 2010 -2014

Coordinator: COPRIN team

Other partners: AROBAS, DEMAR, E-MOTION, STARS, PRIMA, MAIA, TRIO, and LAGADIC Inria teams

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

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

8.1.4. Other Collaborations

- G. Charpiat works with Yann Ollivier and Jamal Atif (TAO team) as well as Rémi Peyre (École des Mines de Nancy / Institut Élie Cartan) on the topic of image compression.
- G. Charpiat works with Giacomo Nardi, Gabriel Peyré and François-Xavier Vialard (Ceremade, Paris-Dauphine University) on the generalization of gradient flows to non-standard metrics.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. CENTAUR

Type: FP7

Defi: Computer vision

Instrument: Industry-Academia Partnerships and Pathway

Objective: a network of scientific excellence addressing research topics in computer vision

Duration: January 2013 - December 2016

Coordinator: Honeywell (CZE)

Partner: Neovison (CZE), Inria Sophia-Antipolis (CZE), Queen Mary University of London (UK) and EPFL in Lausanne (CH).

Inria contact: François Brémond

Abstract: CENTAUR aims at developing a network of scientific excellence addressing research topics in computer vision and advancing the state of the art in video surveillance. The cross fertilization of ideas and technology between academia, research institutions and industry will lay the foundations to new methodologies and commercial solutions for monitoring crowded scenes.Three thrusts identified will enable the monitoring of crowded scenes: (a) multi camera, multicoverage tracking of objects of interest, (b) Anomaly detection and fusion of multimodal sensors, c) activity recognition and behavior analysis in crowded environments.

8.2.1.2. PANORAMA

Type: FP7

Defi: Computer vision

Instrument: Industry-Academia Partnerships and Pathway

Objective: techniques and architectures for imaging applications

Duration: April 2012 - March 2015

Coordinator: Philips Healthcare (NL)

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

Inria contact: François Brémond

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 (see also: http://www.panorama-project.eu/).

8.2.1.3. SUPPORT

Title: Security UPgrade for PORTs

Type: FP7

Defi: Port Security

Instrument: Industry-Academia Partnerships and Pathway

Objective: secure European ports

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).

Inria contact: François Brémond

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

8.2.1.4. Dem@Care

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

Type: FP7

Defi: Cognitive Systems and Robotics

Instrument: Industry-Academia Partnerships and Pathway

Objective: development of a complete system providing personal health services to persons with dementia

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).

Inria contact: François Brémond

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

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

8.3.1.1.1. 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.

8.3.1.1.2. Collaboration with U.S.A.:

Stars collaborates with the University of Southern California.

8.3.1.1.3. Collaboration with Europe:

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

8.3.2. Participation in Other International Programs

- The ANR SafEE (see section 8.1.1.2 collaborates with international partners as Taichung Veterans General Hospital TVGH (TW), NCKU Hospital(TW), SMILE Lab at National Cheng Kung University NCKU (TW) and BDE (TW).

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

ABDALLA OMAR

Date: from Apr 2014 until Sep 2014

Institution: Université Française du Caire (Egypt)

BOUATIRA Mohamed

Date: from Mar 2014 until Sep 2014

Institution: Ecole Mohammadia d'Ingénieurs (Marocco)

CAVERZASI Augustin

Date: until Feb 2014

Institution: Universidad Nacional de Córdoba, Facultad de Ciencias Exactas Físicas y Naturales, Argentina

GOMEZ URIA COVELLA Alvaro

Date: from Mar 2014 until Dec 2014

Institution: National University of Rosario, Argentina

MARTINS DE MELO Filipe

Date: from Apr 2014 until Sep 2014

Institution: Federal University of Penambucco, Brazil

NEGIN Farood

Date: from Apr 2014 until Nov 2014

Institution: Sabanci University, Turkey

NGUYEN Thi Lan Anh

Date: from Mar 2014 until Oct 2014

Institution: Dhai Nguyen Uiversity of Information and Communication Technology, Vietnam

PHAM Ngoc Hai

Date: from May 2014 until Nov 2014

Institution: Science and Technologu University of Hanoi, Vietnam

PUSIOL Pablo Daniel

Date: from Apr 2014 until Sep 2014

Institution: National University of Cordoba, Argentina

SARRAY Ines

Date: Apr 2014 - Oct 2014

Institution: ESPRIT (Ecole d'ingénieurs Tunis) (Tunisia)

STRUMIA Carola

Date: from Oct 2014

Institution: University of Genova, Italy

SUBRAMANIAN Kartick

Date: until August 2014

Institution: Nanyang Technological University, Singapore

ZHOU Kouhua

Date: from Jul 2014 until Sep 2014

Institution: Polytech University of Dalan, China

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General chair, scientific chair

François Brémond was organizer of a PANORAMA (see section 8.2.1.2) special session, part of VISAPP Lisbon, Portugal, 5-8 January 2014.

9.1.1.2. Member of the organizing committee

François Brémond was a member of the Management Committee and COST Action IC1307 in 2014.

Guillaume Charpiat organized a weekly computer vision student seminar [1] in Sophia-Antipolis (about 40 participants).

9.1.2. Scientific Events Selection

9.1.2.1. Member of the conference program committee

Jean-Paul Rigault is a member of the *Association Internationale pour les Technologies à Objets* (AITO) which organizes international conferences such as ECOOP.

François Brémond was program committee member of the conferences and workshops: IEEE Workshop on Applications of Computer Vision (WACV 2014).

François Brémond was program committee member of VS-Re-ID-2014, CAVAU 2014, PETS2014, ATC 2014 and the first International Conference on Cognitive Computing and Information Processing (CCIP-15) at JSSATEN.

François Brémond was area chair of AVSS'14.

9.1.2.2. Reviewer

Guillaume Charpiat reviewed for the conference CVPR (Computer Vision and Pattern Recognition).

Monique Thonnat reviewed for ICPR Conference.

François Brémond was reviewer for the conferences : CVPR2014, ECCV2014, Intelligent Vehicles 2014, VOT2014, CONTACT2014, ChaLearn2014, ACCV2014, WACV 2015.

9.1.3. Journal

9.1.3.1. Member of the editorial board

François Brémond was handling editor of the international journal "Machine Vision and Application".

9.1.3.2. Reviewer

Guillaume Charpiat reviewed for the journals IJCV (International Journal of Computer Vision), JMIV (Journal of Mathematical Imaging and Vision), CVIU (Computer Vision and Image Understanding), and TIP (Transactions on Image Processing).

Jean-Paul Rigault reviewed for the SoSym (Software and System Modeling) journal (Springer).

Monique Thonnat reviewed for Image and Vision Computing Journal.

François Brémond was reviewer for the International Journal of Neural Systems and for the IEEE Pervasive Computing.

François Brémond was reviewer for the journal "IEEE Computing Now".

---

[1] http://www-sop.inria.fr/members/Guillaume.Charpiat/VisionSeminar/

François Brémond was reviewer for the Pervasive and Mobile Computing Journal, ACM Multimedia and Society Journal, Journal of Healthcare Engineering, Sensors Journal, and for Health and Wellbeing Journal.

Sabine Moisan was reviewer for the Journal of Software Engineering for Robotics (JOSER) and for the Computer Science and Information Systems (ComSIS) International Journal.

9.1.4. Invited Talks

François Brémond was invited at the Dagstuhl Seminar "Robots Learning from Experiences", Germany, 17-21 February 2014.

François Brémond was invited to give a talk in VideoSense Summer School on Privacy-respecting video analytics in Eurecom premises in Sophia Antipolis on 17 April 2014.

François Brémond gave an invited talk at SMILE Lab on Activity recognition, Tainan Taiwan, May 2014.

Baptiste Fosty has participated to the 13th édition des Trophées de la Communication, Nice, June 2014.

François Brémond gave an invited talk at PETS on Tracking and surveillance, Seoul Korea, August 2014.

François Brémond was invited to give a talk at SAME, "Enabling the Cloud of THINGS", Sophia Antipolis, 2 October 2014.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master : Annie Ressouche,Critical Software Verification and application to WComp Middleware, 20h, niveau (M2), Polytech Nice School of Nice University, FR and ESPRIT (Tunisia).

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.

9.2.2. Supervision

PhD: Piotr Bilinski, Gesture Recognition in Videos, 5th December 2014, François Brémond.

PhD: Ratnesh Kumar, Fiber-based segmentation of videos for activity recognition, 15th December 2014, Guillaume Charpiat and Monique Thonnat.

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

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

PhD in progress : Auriane Gros, Evaluation and Specific Management of Emotionnal Disturbances with Activity Recognition Systems for Alzheimer patient, François Brémond.

PhD in progress : Minh Khue Phan Tran, Man-machine interaction for older adults with dementia, May 2013, François Brémond.

PhD in progress : Michal Koperski, Detecting critical human activities using RGB/RGBD cameras in home environment, François Brémond.

PhD in progress : Thi Lan Anh Nguyen, Complex Activity Recognition from 3D sensors, François Brémond.

9.2.3. Juries

9.2.3.1. PhD, HDR

François Brémond was jury member of the following PhD and HDR theses:

PhD, Hajer Fradi, EURECOM, 28th January 2014.

PhD, Martin Hirzer, Graz University of Technology, Graz, Austria , March 2014.

Pre-PhD Effrosyni Doutsi, Nice University - I3S, 23 June 2014.

Pre-PhD Antoine Basset, Institut Curie, Paris, 8 July 2014.

PhD, Usman Niaz, EURECOM, Sophia Antipolis, 8 July 2014.

HDR, Thierry BOUWMANS, November 2014.

Jean-Paul Rigault was reviewer and jury president of the PhD of Nadezhda Baklanova, University of Toulouse, December 2014.

Guillaume Charpiat reviewed Eduardo Ferández-Moral's PhD thesis (Malaga University, Spain, September 2014).

9.2.3.2. Expertise

Jean-Paul Rigault was a member of the ISO Committee on the C++ language.

François Brémond was expert for the Proposal KIC Healthy Ageing (Innolife) Expert Group1: Independent living.

François Brémond was expert for EU European Reference Network for Critical Infrastructure Protection (ERNCIP) - Video Anlaytics and surveillance Group, at European Commission's Joint Research Centre in Brussels in August and October 2014.

François Brémond was expert for reviewing tutorial proposals for ICIP'14.

François Brémond participated to the expertise for NEM (New European Media) an EU ETP concerning media, creativity and content to provide future strategies for EU research funding under H2020.

François Brémond was expert for the committee selection for the tenure track, Mines School - ParisTech, June 2014.

9.3. Popularization

François Brémond was invited to give a talk at Conférence des métiers at International Lycée (CIV) in Sophia 27 January 2014.

François Brémond was interviewed by Nice Matin on CoBTeK and Alzheimer issues, Opening of ICP, 10 March 2014.

François Brémond was interviewed by Monaco TV on MONAA and Autism issues, 12 March 2014.

François Brémond has participated to the ERCIM News 98 -Special theme: Smart Cities, July, 2014.

Guillaume Charpiat was part of the popularization MASTIC committee.

10. Bibliography

Major publications by the team in recent years

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

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

[3] B. BOULAY, F. BREMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System*, in "Pattern Recognition Letter", 2006, vol. 27, n° 15, pp. 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, pp. 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 Shape Analysis and Deformable Image Alignment (NORDIA)", Kyoto, Japan, September 2009

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

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

[8] 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, pp. 167-188

[9] 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, pp. 189-205

[10] 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, pp. 91-103

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

[12] 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, pp. 1182-1191

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

[14] 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

[15] 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, pp. 501-507

[16] 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

[17] 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

[18] 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, pp. 157-171, selected as one of the 17 best papers of SERA'08 conference

[19] 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, pp. 441-471, http://hal.inria.fr/inria-00524499/en

[20] 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

[21] 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

[22] M. THONNAT. *Special issue on Intelligent Vision Systems*, in "Computer Vision and Image Understanding", May 2010, vol. 114, n° 5, pp. 501-502

[23] 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

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

[25] 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 (IJCAI'03)", 9-15 September 2003

[26] 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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[27] P. T. BILIŃSKI. *Human Action Recognition in Videos*, Universite de Nice-Sophia Antipolis, December 2014, https://tel.archives-ouvertes.fr/tel-01097111

[28] R. KUMAR. *Video Segmentation and Multiple Object Tracking*, Inria Sophia Antipolis, December 2014, https://tel.archives-ouvertes.fr/tel-01096638

Articles in International Peer-Reviewed Journals

[29] F. BREMOND, V. BOGORNY, L. PATINO, S. COSAR, G. PUSIOL, G. DONATIELLO. *Monitoring People's Behaviour using Video Analysis and Trajectory Clustering*, in "ERCIM News", July 2014, n° 98, https://hal.inria.fr/hal-01061036

[30] D. P. CHAU, M. THONNAT, F. BREMOND, E. CORVEE. *Online Parameter Tuning for Object Tracking Algorithms*, in "Image and Vision Computing", February 2014, vol. 32, n° 4, pp. 287-302, https://hal.inria.fr/hal-00976594

[31] P. ROBERT, A. KOENING, H. AMIEVA, S. ANDRIEU, F. BREMOND, R. BULLOCK, M. CECCALDI, B. DUBOIS, S. GAUTHIER, P.-A. KENIGSBERG, S. NAVE, J.-M. ORGOGOZO, J. PIANO, M. BENOIT, J. TOUCHON, B. VELLAS, J. YESAVAGE, V. MANERA. *Recommendations for the use of Serious Games in people with Alzheimer's Disease, related disorders and frailty*, in "frontiers in Aging Neuroscience", March 2014, vol. 6, n° 54 [*DOI :* 10.3389/FNAGI.2014.00054], https://hal.inria.fr/hal-01059594

[32] Y. TARABALKA, G. CHARPIAT, L. BRUCKER, B. MENZE. *Spatio-Temporal Video Segmentation with Shape Growth or Shrinkage Constraint*, in "IEEE Transactions on Image Processing", September 2014, vol. 23, n° 9, pp. 3829-3840, https://hal.inria.fr/hal-01052543

International Conferences with Proceedings

[33] J. BADIE, F. BREMOND. *Global tracker: an online evaluation framework to improve tracking quality*, in "AVSS - 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance", Seoul, South Korea, August 2014, https://hal.inria.fr/hal-01062766

[34] S. BAK, V. BATHRINARAYANAN, F. BREMOND, A. CAPRA, D. GIACALONE, G. MESSINA, A. BUEMI. *Retrieval tool for person re-identification*, in "PANORAMA Workshop in conjunction with VISIGRAPP", Lisbon, Portugal, January 2014, https://hal.inria.fr/hal-00907455

[35] S. BAK, R. KUMAR, F. BREMOND. *Brownian descriptor: a Rich Meta-Feature for Appearance Matching*, in "WACV: Winter Conference on Applications of Computer Vision", Steamboat Springs CO, United States, March 2014, https://hal.inria.fr/hal-00905588

[36] S. BAK, S. ZAIDENBERG, B. BOULAY, F. BREMOND. *Improving Person Re-identification by Viewpoint Cues*, in "Advanced Video and Signal-based Surveillance", Seoul, South Korea, August 2014, pp. 1-6, https://hal.archives-ouvertes.fr/hal-01004390

[37] P. BILINSKI, J. AHRENS, M. R. P. THOMAS, I. J. TASHEV, J. C. PLATT. *HRTF Magnitude Synthesis via Sparse Representation of Anthropometric Features*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", Florence, Italy, May 2014, pp. 4468 - 4472 [*DOI :* 10.1109/ICASSP.2014.6854447], https://hal.inria.fr/hal-01097303

[38] P. BILINSKI, M. KOPERSKI, S. BAK, F. BREMOND. *Representing Visual Appearance by Video Brownian Covariance Descriptor for Human Action Recognition*, in "AVSS - 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance", Seoul, South Korea, IEEE, August 2014, https://hal.inria.fr/hal-01054943

[39] D. P. CHAU, F. BREMOND, M. THONNAT, S. BAK. *Automatic Tracker Selection w.r.t Object Detection Performance*, in "IEEE Winter Conference on Applications of Computer Vision (WACV)", Steamboat Springs CO, United States, March 2014, https://hal.inria.fr/hal-00974693

[40] C. CRISPIM, F. BREMOND. *Uncertainty Modeling Framework for Constraint-based Elementary Scenario Detection in Vision System*, in "1st International Workshop on Computer vision + ONTology Applied Cross-disciplinary Technologies in Conjunction with ECCV 2014", Zurich, Switzerland, M. CRISTANI, R. FERRARIO, J. J. CORSO (editors), September 2014, https://hal.inria.fr/hal-01054769

[41] L. EMILIANO SÁNCHEZ, J. ANDRÉS DIAZ-PACE, A. ZUNINO, S. MOISAN, J.-P. RIGAULT. *An Approach for Managing Quality Attributes at Runtime using Feature Models*, in "8th Brazilian Symposium on Software Components, Architectures and Reuse (SBCARS)", Maceio, Brazil, September 2014, 10 p. , https://hal.inria.fr/hal-01093085

[42] M. KOPERSKI, P. BILINSKI, F. BREMOND. *3D Trajectories for Action Recognition*, in "ICIP - The 21st IEEE International Conference on Image Processing", Paris, France, IEEE, October 2014, https://hal.inria.fr/hal-01054949

[43] R. KUMAR, G. CHARPIAT, M. THONNAT. *Hierarchical Representation of Videos with Spatio-Temporal Fibers*, in "IEEE Winter Conference on Applications of Computer Vision", Colorado, United States, March 2014, https://hal.inria.fr/hal-00911012

[44] R. KUMAR, G. CHARPIAT, M. THONNAT. *Multiple Object Tracking by Efficient Graph Partitioning*, in "ACCV - 12th Asian Conference on Computer Vision", Singapore, Singapore, T. SIM, J. WU (editors), Lecture Notes on Computer Science, Springer, November 2014, https://hal.inria.fr/hal-01061450

[45] S. MUKANAHALLIPATNA SIMHA, D. P. CHAU, F. BREMOND. *Feature Matching using Co-inertia Analysis for People Tracking*, in "The 9th International Conference on Computer Vision Theory and Applications (VISAPP)", Lisbon, Portugal, January 2014, https://hal.inria.fr/hal-00909566

[46] S.-T. SERBAN, S. MUKANAHALLIPATNA SIMHA, V. BATHRINARAYANAN, E. CORVEE, F. BREMOND. *Towards Reliable Real-Time Person Detection*, in "VISAPP - The International Conference on Computer Vision Theory and Applications", Lisbon, Portugal, January 2014, https://hal.inria.fr/hal-00909124

[47] S. ZAIDENBERG, P. BILINSKI, F. BREMOND. *Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance*, in "VISAPP - 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications", Lisbon, Portugal, S. BATTIATO (editor), SCITEPRESS Digital Library, January 2014, https://hal.inria.fr/hal-00878580

National Conferences with Proceedings

[48] M. K. PHAN TRAN, F. BREMOND, P. ROBERT. *Comment intéresser les personnes âgées aux Serious Game ?*, in "JA-SFTAG 2014", Paris, France, November 2014, https://hal.archives-ouvertes.fr/hal-01092329

Research Reports

[49] G. CHARPIAT, G. NARDI, G. PEYRÉ, F.-X. VIALARD. *Piecewise rigid curve deformation via a Finsler steepest descent*, CNRS, December 2014, https://hal.archives-ouvertes.fr/hal-00849885

[50] E. MAGGIORI, Y. TARABALKA, G. CHARPIAT. *Multi-label segmentation of images with partition trees*, Inria Sophia Antipolis, November 2014, https://hal.inria.fr/hal-01084166

Other Publications

[51] O. ABDALLA. *Contribution au développement d'une plateforme de reconnaissance d'activités*, Polytech'Nice, September 2014, https://hal.inria.fr/hal-01095189

[52] C. GARATE, S. ZAIDENBERG, J. BADIE, F. BREMOND. *Group Tracking and Behavior Recognition in Long Video Surveillance Sequences*, January 2014, VISAPP - 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, https://hal.inria.fr/hal-00879734

[53] A. KÖNIG, C. F. CRISPIM-JUNIOR, A. DERREUMAUX, G. BENSADOUN, P.-D. PETIT, F. BREMOND, R. DAVID, F. VERHEY, P. AALTEN, P. ROBERT. *Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients*, December 2014 [*DOI :* 10.3233/JAD-141767], https://hal.inria.fr/hal-01094093

[54] I. SARRAY. *Composition Adaptative et Vérification Formelle de Logiciel en Informatique Ubiquitaire*, Polytech'Nice, September 2014, https://hal.inria.fr/hal-01095219

References in notes

[55] 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

[56] 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

[57] K. S. ANTON ANDRIYENKO. *Multi-target Tracking by Continuous Energy Minimization*, in "CVPR", Providence,RI, IEEE, June 2011, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5995311

[58] 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, China, July 2012, pp. 487-493 [*DOI :* 10.1109/AVSS.2012.90], https://hal.inria.fr/hal-00761322

[59] S.-H. BAE, K.-J. YOON. *Robust Online Multi-Object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning*, in "CVPR", Columbus, IEEE, June 2014

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

[61] 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", Shanghai, China, April 2013, http://hal.inria.fr/hal-00806321

[62] D. P. CHAU, J. BADIE, F. BREMOND, M. THONNAT. *Online Tracking Parameter Adaptation based on Evaluation*, in "IEEE International Conference on Advanced Video and Signal-based Surveillance", Krakow, Poland, August 2013, https://hal.inria.fr/hal-00846920

[63] D. P. CHAU, F. BREMOND, M. THONNAT. *Online evaluation of tracking algorithm performance*, in "The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP)", London,UK, December 2009, https://hal.inria.fr/inria-00486479

[64] D. P. CHAU, M. THONNAT, F. BREMOND. *Automatic Parameter Adaptation for Multi-object Tracking*, in "International Conference on Computer Vision Systems (ICVS)", St Petersburg, Russia, Lecture Notes in Computer Science, Springer, July 2013, https://hal.inria.fr/hal-00821669

[65] 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, pp. 359-364, http://nusmv.fbk.eu/NuSMV/papers/cav02/ps/cav02.ps

[66] C. F. CRISPIM-JUNIOR, V. BATHRINARAYANAN, B. FOSTY, R. ROMDHANE, A. KONIG, M. THONNAT, F. BREMOND. *Evaluation of a Monitoring System for Event Recognition of Older People*, in "International Conference on Advanced Video and Signal-Based Surveillance 2013", Krakow, Poland, August 2013, pp. 165 - 170 [*DOI :* 10.1109/AVSS.2013.6636634], https://hal.inria.fr/hal-00875972

[67] 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, pp. 2387-2397

[68] 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

[69] D. GAFFÉ, A. RESSOUCHE. *Algebraic Framework for Synchronous Language Semantics*, in "Theoritical Aspects of Software Engineering", Birmingham, United Kingdom, L. FERARIU, A. PATELLI (editors), 2013 Symposium on Theoritical Aspects of Sofware Engineering, IEEE Computer Society, July 2013, pp. 51-58, https://hal.inria.fr/hal-00841559

[70] C. GARCIA CIFUENTES, M. STURZEL, F. JURIE, G. J. BROSTOW. *Motion Models that Only Work Sometimes*, in "British Machive Vision Conference", Guildford, United Kingdom, September 2012, 12 p. , https://hal.archives-ouvertes.fr/hal-00806098

[71] A. HEILI, A. LOPEX-MENDEZ, J.-M. ODOBEZ. *Exploiting Long-term Connectivity and Visual Motion in CRF-based Multi-person Tracking*, in "Idiap-RR", May 2014, http://publications.idiap.ch/downloads/reports/2014/Heili_Idiap-RR-05-2014.pdf

[72] A. HEILI, J.-M. ODOBEZ. *Parameter estimation and contextual adaptation for a multi-object tracking CRF model*, in "PETS", 2013

[73] 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

[74] C.-H. KUO, C. HUANG, R. NEVATIA. *Multi-target tracking by on-line learned discriminative appearance models*, in "CVPR", 2010

[75] 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, pp. 175-194

[76] Y. LI, C. HUANG, R. NEVATIA. *Learning to associate: Hybridboosted multi-target tracker for crowded scene*, in "CVPR", 2009

[77] A. MILAN, K. SCHINDLER, S. ROTH. *Detection and Trajectory-Level Exclusion in Multiple Object Tracking*, in "CVPR", 2013

[78] 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

[79] A. T. NGHIEM, E. AUVINET, J. MEUNIER. *Head detection using Kinect camera and its application to fall detection*, in "Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on", 2012, pp. 164-169, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&tp=&arnumber=6310538

[80] 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, pp. 477-498

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

[82] 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

[83] 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, pp. 165-180, http://hal.inria.fr/inria-00605915/en/

[84] 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

[85] 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

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

[87] A. W. M. SMEULDERS, D. M. CHU, R. CUCCHIARA, S. CALDERARA, A. DEHGHAN, M. SHAH. *Visual Tracking: an Experimental Survey*, in "IEEE Transaction on Pattern Analysis and Machine Intelligence", July 2014, vol. 36, n° 7

[88] G. J. Szekely, M. L. Rizzo. *Brownian distance covariance*, in "The Annals of Applied Statistics", 2009, vol. 3, n° 4, pp. 1236–1265

[89] S. Tang, M. Andriluka, A. Milan, K. Schindler, S. Roth, B. Schiele. *Learning People Detectors for Tracking in Crowded Scenes*, in "IEEE International Conference on Computer Vision (ICCV)", December 2013, http://www.cv-foundation.org/openaccess/content_iccv_2013/html/Tang_Learning_People_Detectors_2013_ICCV_paper.html

[90] J.-Y. Tigli, S. Lavirotte, G. Rey, V. Hourdin, D. Cheung, E. Callegari, M. Riveill. *WComp middleware for ubiquitous computing: Aspects and composite event-based Web services*, in "Annals of Telecommunications", 2009, vol. 64, n° 3-4, ISSN 0003-4347 (Print) ISSN 1958-9395 (Online)

[91] J.-Y. Tigli, S. Lavirotte, G. Rey, V. Hourdin, M. Riveill. *Lightweight Service Oriented Architecture for Pervasive Computing*, in "IJCSI International Journal of Computer Science Issues", 2009, vol. 4, n° 1, ISSN (Online): 1694-0784 ISSN (Print): 1694-0814

[92] O. Tuzel, F. Porikli, P. Meer. *Region Covariance: A Fast Descriptor for Detection And Classification*, in "ECCV", 2006

[93] H. Wang, A. Kläser, C. Schmid, C.-L. Liu. *Action Recognition by Dense Trajectories*, in "IEEE Conference on Computer Vision & Pattern Recognition", Colorado Springs, United States, June 2011, pp. 3169-3176, http://hal.inria.fr/inria-00583818/en

[94] Y. Wu. *Mining Actionlet Ensemble for Action Recognition with Depth Cameras*, in "Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Washington, DC, USA, CVPR '12, IEEE Computer Society, 2012, pp. 1290–1297, http://dl.acm.org/citation.cfm?id=2354409.2354966

[95] I. A. K. Xu Yan, S. K. Shah. *What do I see? Modeling Human Visual Perception for Multi-person Tracking*, in "ECCV", Zurich, Germany, Sept 2014, http://www2.cs.uh.edu/~xuyan/docs/eccv2014.pdf

[96] A. R. Zamir, A. Dehghan, M. Shah. *GMCP-Tracker: Global Multi-object Tracking Using Generalized Minimum Clique Graphs*, in "ECCV", 2012