



Activity Report 2013

Project-Team STARS

Spatio-Temporal Activity Recognition Systems

RESEARCH CENTER
Sophia Antipolis - Méditerranée

THEME
**Vision, perception and multimedia
interpretation**

Table of contents

1. Members	1
2. Overall Objectives	2
2.1. Presentation	2
2.1.1. Research Themes	2
2.1.2. International and Industrial Cooperation	4
2.2. Highlights of the Year	4
3. Research Program	5
3.1. Introduction	5
3.2. Perception for Activity Recognition	5
3.2.1. Introduction	5
3.2.2. Appearance models and people tracking	5
3.2.3. Learning shape and motion	6
3.3. Semantic Activity Recognition	6
3.3.1. Introduction	6
3.3.2. High Level Understanding	6
3.3.3. Learning for Activity Recognition	7
3.3.4. Activity Recognition and Discrete Event Systems	7
3.4. Software Engineering for Activity Recognition	7
3.4.1. Platform Architecture for Activity Recognition	8
3.4.2. Discrete Event Models of Activities	9
3.4.3. Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems	10
4. Application Domains	10
4.1. Introduction	10
4.2. Video Analytics	10
4.3. Healthcare Monitoring	11
5. Software and Platforms	11
5.1. SUP	11
5.2. ViSEval	12
5.3. Clem	15
6. New Results	15
6.1. Introduction	15
6.1.1. Perception for Activity Recognition	15
6.1.2. Semantic Activity Recognition	16
6.1.3. Software Engineering for Activity Recognition	17
6.2. Background Subtraction and People Detection in Videos	17
6.3. Tracking and Video Representation	19
6.4. Video segmentation with shape constraint	20
6.4.1. Video segmentation with growth constraint	20
6.4.2. Video segmentation with statistical shape prior	22
6.5. Articulating motion	25
6.6. Lossless image compression	25
6.7. People detection using RGB-D cameras	27
6.8. Online Tracking Parameter Adaptation based on Evaluation	28
6.9. People Detection, Tracking and Re-identification Through a Video Camera Network	29
6.9.1. People detection:	30
6.9.2. Object tracking:	30
6.9.3. People re-identification:	30
6.10. People Retrieval in a Network of Cameras	30

6.11. Global Tracker : an Online Evaluation Framework to Improve Tracking Quality	32
6.12. Human Action Recognition in Videos	34
6.13. 3D Trajectories for Action Recognition Using Depth Sensors	35
6.14. Unsupervised Sudden Group Movement Discovery for Video Surveillance	35
6.15. Group Behavior Understanding	37
6.16. Evaluation of an Activity Monitoring System for Older People Using Fixed Cameras	39
6.17. A Framework for Activity Detection of Older People Using Multiple Sensors	39
6.18. Walking Speed Detection on a Treadmill using an RGB-D Camera	41
6.19. Serious Game for older adults with dementia	42
6.20. Unsupervised Activity Learning and Recognition	44
6.21. Extracting Statistical Information from Videos with Data Mining	46
6.22. SUP	46
6.23. Model-Driven Engineering for Activity Recognition	50
6.23.1. Configuration Adaptation at Run Time	50
6.23.2. Run Time Components	51
6.24. Scenario Analysis Module	52
6.25. The Clem Workflow	52
6.26. Multiple Services for Device Adaptive Platform for Scenario Recognition	53
7. Bilateral Contracts and Grants with Industry	53
8. Partnerships and Cooperations	54
8.1. Regional Initiatives	54
8.2. National Initiatives	54
8.2.1. ANR	54
8.2.1.1. MOVEMENT	54
8.2.1.2. SafEE	55
8.2.2. Investment of future	55
8.2.3. Large Scale Inria Initiative	55
8.2.4. Other collaborations	55
8.3. European Initiatives	56
8.3.1. FP7 Projects	56
8.3.1.1. CENTAUR	56
8.3.1.2. SUPPORT	56
8.3.1.3. Dem@Care	56
8.3.1.4. VANAHEIM	57
8.3.2. Collaborations in European Programs, except FP7	57
8.4. International Initiatives	58
8.4.1. Inria International Partners	58
8.4.1.1. Collaborations with Asia	58
8.4.1.2. Collaboration with U.S.	58
8.4.1.3. Collaboration with Europe	58
8.4.2. Participation In other International Programs	58
8.5. International Research Visitors	58
9. Dissemination	60
9.1. Scientific Animation	60
9.2. Teaching - Supervision - Juries	60
9.2.1. Teaching	60
9.2.2. Supervision	61
9.2.3. Juries	61
9.3. Popularization	61
10. Bibliography	61

Project-Team STARS

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

Monique Thonnat [Senior Researcher, Inria, HdR]
François Brémond [Team leader, Senior Researcher, Inria, HdR]
Guillaume Charpiat [Researcher, Inria]
Sabine Moisan [Researcher, Inria, HdR]
Annie Ressouche [Researcher, Inria]
Daniel Gaffé [Associate Professor, UNS, until Feb 2013, since Sep 2013]

External Collaborators

Etienne Corvée [Linkcare Services, from Jan 2013]
Alexandre Derreumaux [CHU Nice, from Jan 2013]
Guido-Tomas Pusiol [Post doc Standfort University, from Mar 2013 until Sep 2013]
Jean-Paul Rigault [UNS, from Jan 2013]
Qiao Ma [Beihang University, from Jan 2013]
Luis-Emiliano Sanchez [Universidad Nacional del Centro de la Provincia de Buenos Aires, from Mar 2013 until Apr 2013]
Silviu-Tudor Serban [MB TELECOM LTD. S.R.L. - ROMANIA, from Jan 2013 until Sep 2013]
Jean-Yves Tigli [UNS and CNRS-I3S, from Jan 2013]

Engineers

Vasanth Bathrinarayanan [FP7 SUPPORT project, Inria, from Jan 2013]
Giuseppe Donatiello [FP7 SUPPORT project, Inria, from Mar 2013]
Hervé Falciani [FP7 VANAHEIM project, Inria, from Jan 2013 until Sep 2013]
Baptiste Fosty [FP7 VANAHEIM project, Inria, until Jul 2013; CDC AZ@GAME project, Inria, from Aug 2013]
Julien Gueytat [FP7 VANAHEIM project, Inria, from Jan 2013 until Jul 2013; FP7 SUPPORT project, Inria, from Aug 2013]
Srinidhi Mukanahallipatna Simha [AE PAL, Inria, until Jul 2013]
Jacques Serlan [Inria, from Nov 2013]

PhD Students

Malik Souded [CIFFRE Keeneo since Jun 2013; AE PAL from Jul 2013 until Nov 2013; FP7 SUPPORT, Inria, from Dec 2013]
Julien Badie [FP7 DEM@CARE project, Inria]
Carolina Garate Oporto [FP7 VANAHEIM project, Inria]
Minh Khue Phan Tran [CIFRE, Genius, from May 2013]
Ratnesh Kumar [FP7 DEM@CARE project, Inria]
Rim Romdhane [Inria, Inria, until Sep 2013]
Piotr Tadeusz Bilinski [FP7 SUPPORT project, Inria, until Feb 2013]

Post-Doctoral Fellows

Slawomir Bak [DGCIS PANORAMA project, Inria]
Duc Phu Chau [FP7 VANAHEIM project, Inria; DGCIS PANORAMA project, Inria]
Carlos-Fernando Crispim Junior [FP7 DEM@CARE project, Inria]
Anh-Tuan Nghiem [CDC AZ@GAME project, Inria]
Sofia Zaidenberg [FP7 VANAHEIM project, Inria, until Jul 2013; w Sep 2013 until Oct 2013]

Salma Zouaoui-Elloumi [FP7 VANAHEIM project, Inria, until Jun 2013; FP7 SUPPORT from Jul 2013]
Serhan Cosar [FP7 SUPPORT project, Inria, from Mar 2013]

Visiting Scientists

Vít Libal [Honeywell Praha, from Jul 2013 until Oct 2013]
Marco San Biagio [Italian Inst. of Tech. of Genova, from Apr 2013 until Sep 2013]
Kartick Subramanian [Nanyang Technological University, from Mar 2013 until Aug 2013]

Administrative Assistant

Jane Desplanques [Inria]

Others

Michal Koperski [;FP7 EIT ICT LABS GA project, Inria, from Apr 2013 until Sep 2013; CDC TOYOTA project, Inria, from Oct 2013]
Imad Rida [FP7 EIT ICT LABS GA project, Inria, from Mar 2013 until Aug 2013]
Abhineshwar Tomar [FP7 DEM@CARE project, Inria, until Apr 2013]
Vaibhav Katiyar [FP7 SUPPORT project, Asian Institute of Technology Khlong Luang Pathumtani, Thailand, until Jan 2013]
Mohammed Cherif Bergheul [EGIDE, Inria, from Apr 2013 until Sep 2013]
Stefanus Candra [EGIDE, Inria, from Aug 2013 until Dec 2013]
Agustín Caverzasi [EGIDE, Inria, from Aug 2013]
Sahil Dhawan [EGIDE, Inria, from Jan 2013 until Jul 2013]
Narjes Ghrairi [EGIDE, Inria, from Apr 2013 until Sep 2013]
Joel Wanza Weloli [EGIDE, Inria, from Jun 2013 until Aug 2013]

2. Overall Objectives

2.1. Presentation

2.1.1. Research Themes

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

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

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

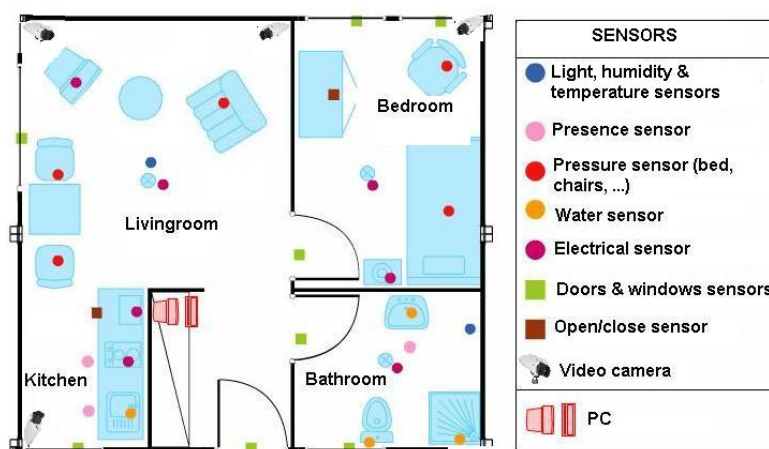


Figure 1. Homecare monitoring: the set of sensors embedded in an apartment

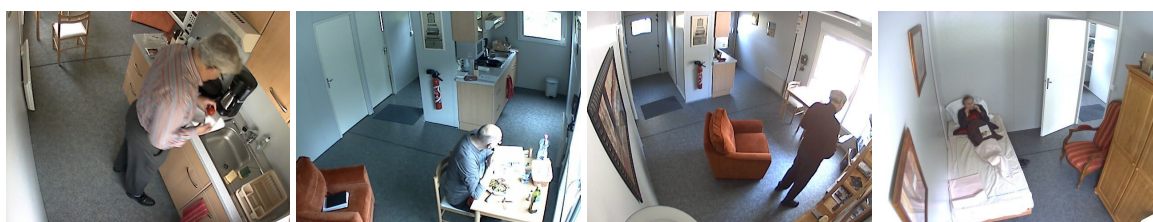


Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

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

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

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

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

2.1.2. International and Industrial Cooperation

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

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

2.2. Highlights of the Year

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

During this period, we have designed several novel algorithms for activity recognition systems. In particular, we have extended an efficient algorithm for tuning automatically the parameters of the people tracking algorithm.

We have designed a compact system for activity recognition running on a mini-PC which is easily deployable using RGBD video cameras. This algorithm has been tested on more than 70 videos of older adults performing 15 min of physical exercises and cognitive tasks. This evaluation has been part of a large clinical trial with Nice Hospital to characterize the behaviour profile of Alzheimer patients compared to healthy older people.

We have also been able to demonstrate the tracking and the recognition of group behaviours in live in Paris subway. We have stored efficiently in a huge database the meta-data (e.g. people trajectories) generated from the processing of 8 video cameras, each of them lasting several days. From these meta-data, we have automatically discovered few hundreds of rare events, such as loitering, collapsing, ... to display them on the screen of subway security operators.

Monique Thonnat has been at the head of the Inria Bordeaux Center since the first of November 2013. She is still working part-time in Stars team.

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

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

3.2. Perception for Activity Recognition

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

Computer Vision; Cognitive Systems; Learning; Activity Recognition.

3.2.1. Introduction

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

3.2.2. Appearance models and people tracking

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

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

Therefore, we work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). New 3D 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.

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

3.2.3. *Learning shape and motion*

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

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

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

3.3. Semantic Activity Recognition

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

Activity Recognition, Scene Understanding, Computer Vision

3.3.1. *Introduction*

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

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

3.3.2. *High Level Understanding*

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event).

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

3.3.3. Learning for Activity Recognition

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

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

3.3.4. Activity Recognition and Discrete Event Systems

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

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

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

3.4. Software Engineering for Activity Recognition

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

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

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

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

3.4.1. Platform Architecture for Activity Recognition

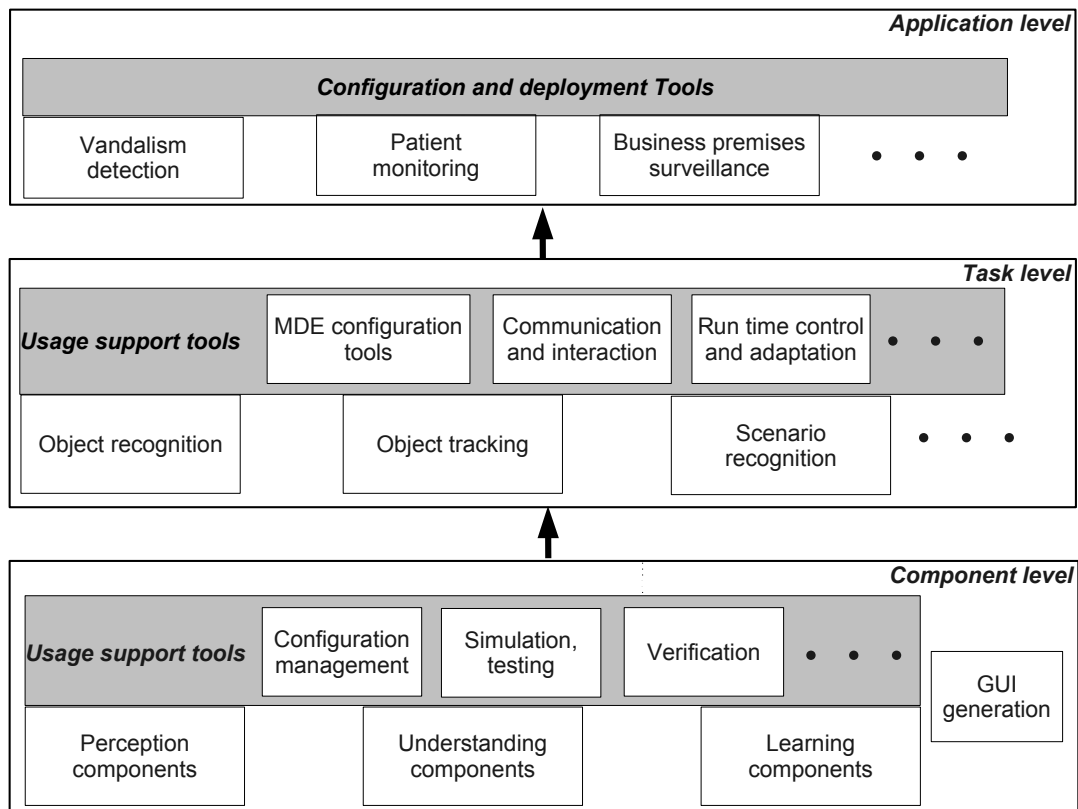


Figure 4. Global Architecture of an Activity Recognition The grey areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [15], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system or a cognitive system.

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

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

- Perception components contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
- Understanding components provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.
- Learning components implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

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

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

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

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

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

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

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

3.4.2. Discrete Event Models of Activities

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

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

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

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

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [57]. 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 [56], [85], [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 [7], [1], [9] (also referred to as visual surveillance) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

For instance, the Keeneo start-up was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in video analytics (VSIP library, which was a previous version of SUP). Keeneo has been bought by Digital Barriers in August 2011 and is now independent from Inria. However, Stars continues to maintain a close cooperation with Keeneo for impact analysis of 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 [66], [86] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

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

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

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

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

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

5. Software and Platforms

5.1. SUP

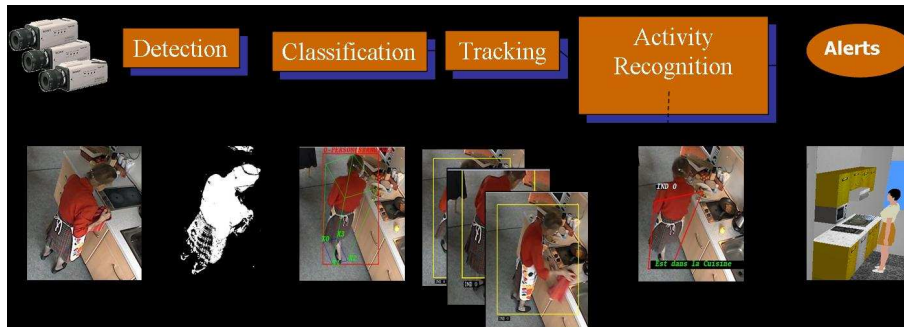


Figure 5. Tasks of the Scene Understanding Platform (SUP).

SUP is a Scene Understanding Software Platform written in C and C++ (see Figure 5). SUP is the continuation of the VSIP platform. SUP is splitting the workflow of a video processing into several modules, such as acquisition, segmentation, etc., up to activity recognition, to achieve the tasks (detection, classification, etc.) the platform supplies. Each module has a specific interface, and different plugins implementing these interfaces can be used for each step of the video processing. This generic architecture is designed to facilitate:

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

Currently, 15 plugins are available, covering the whole processing chain. Several plugins are using the Genius platform, an industrial platform based on VSIP and exploited by Keeneo.

Goals of SUP are twofold:

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

5.2. ViSEval

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

The proposed evaluation tool (ViSEval, visualization and evaluation) respects three important properties:

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

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

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

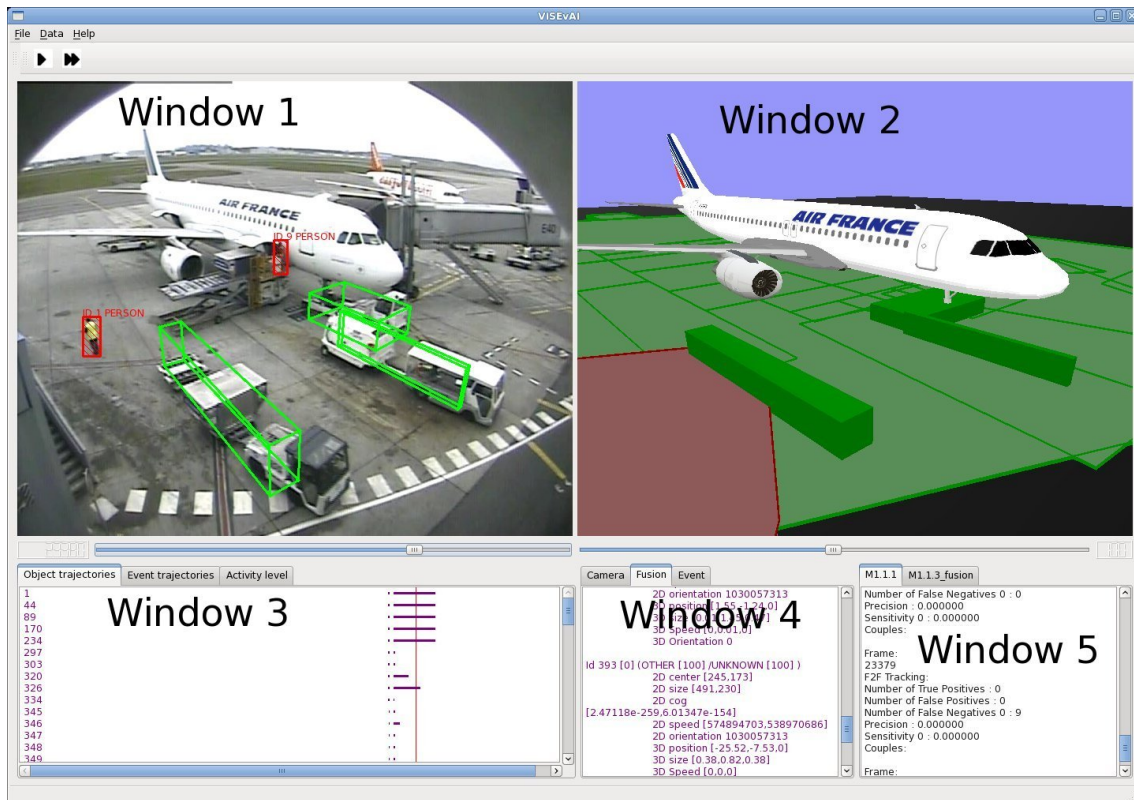


Figure 6. GUI of the ViSEvAI software

The GUI is composed of 3 different parts:

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

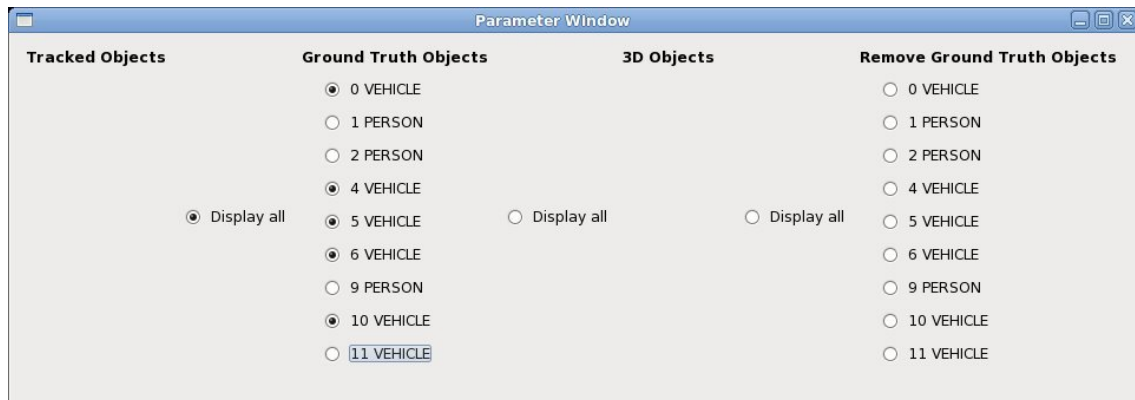


Figure 7. The object window enables users to choose the object to display

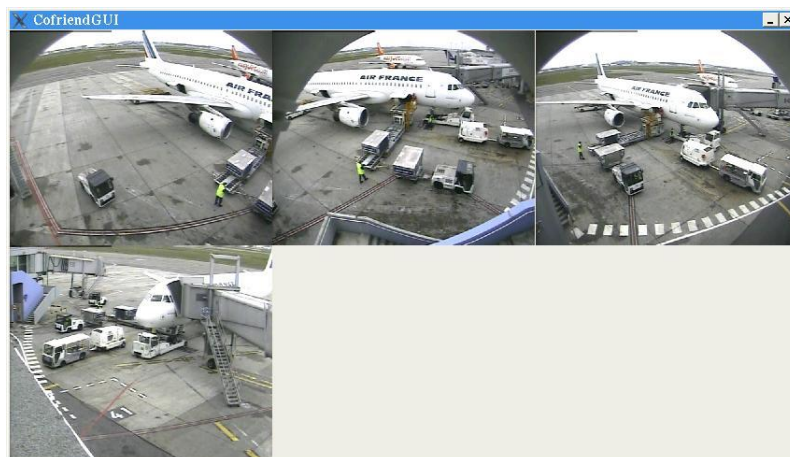


Figure 8. The multi-view window

- Window 2: the 3D virtual scene displays a 3D view of the scene (3D avatars for the detected and ground-truth objects, context, ...).
 - Window 3: the temporal information about the detected and ground truth objects, and about the recognized and ground-truth events.
 - Window 4: the description part gives detailed information about the objects and the events,
 - Window 5: the metric part shows the evaluation results of the frame metrics.
2. The object window enables the user to choose the object to be displayed (see Figure 7).
 3. The multi-view window displays the different points of view of the scene (see Figure 8).

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

The ViSEvAI 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. ViSEvAI 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/ViSEvAI_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 [19] [82] programs. LE is a synchronous language supporting a modular compilation. It also supports automata possibly designed with a dedicated graphical editor.

Each LE program is compiled later into lec and lea files. Then when we want to generate code for different backends, depending on their nature, we can either expand the lec code of programs in order to resolve all abstracted variables and get a single lec file, or we can keep the set of lec files where all the variables of the main program are defined. Then, the *finalization* will simplify the final equations and code is generated for simulation, safety proofs, hardware description or software code. Hardware description (Vhdl) and software code (C) are supplied for LE programs as well as simulation. Moreover, we also generate files to feed the NuSMV model checker [65] in order to perform validation of program behaviors.

6. New Results

6.1. Introduction

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

6.1.1. Perception for Activity Recognition

Participants: Julien Badie, Slawomir Bak, Vasanth Bathrinarayanan, Piotr Bilinski, François Brémond, Guillaume Charpiat, Duc Phu Chau, Etienne Corvée, Carolina Garate, Vaibhav Katiyar, Ratnesh Kumar, Srinidhi Mukanahallipatna, Marco San Biago, Silviu Serban, Malik Souded, Kartick Subramanian, Anh Tuan Nghiem, Monique Thonnat, Sofia Zaidenberg.

This year Stars has extended an algorithm for tuning automatically the parameters of the people tracking algorithm. We have evaluated the algorithm for re-identification of people through a camera network while taking into account a large variety of potential features together with practical constraints. We have designed several original algorithms for the recognition of short actions and validated its performance on several benchmarking databases (e.g. ADL). We have also worked on video segmentation and representation, with different approaches and applications.

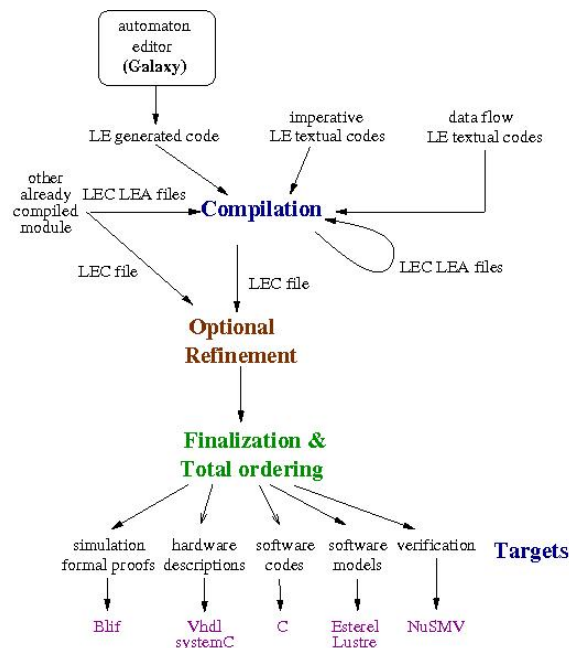


Figure 9. The Clem Toolkit

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

- Background Subtraction and People Detection in Videos (6.2)
- Tracking and Video Representation (6.3)
- Video segmentation with shape constraint (6.4)
- Articulating motion (6.5)
- Lossless image compression (6.6)
- People detection using RGB-D cameras (6.7)
- Online Tracking Parameter Adaptation based on Evaluation (6.8)
- People Detection, Tracking and Re-identification Through a Video Camera Network (6.9)
- People Retrieval in a Network of Cameras (6.10)
- Global Tracker : an Online Evaluation Framework to Improve Tracking Quality (6.11)
- Human Action Recognition in Videos (6.12)
- 3D Trajectories for Action Recognition Using Depth Sensors (6.13)
- Unsupervised Sudden Group Movement Discovery for Video Surveillance (6.14)
- Group Behavior Understanding (6.15)

6.1.2. Semantic Activity Recognition

Participants: Guillaume Charpiat, Serhan Cosar, Carlos -Fernando Crispim Junior, Hervé Falciani, Baptiste Fosty, Qiao Ma, Rim Romdhane.

During this period, we have thoroughly evaluated the generic event recognition algorithm using both sensors (RGB and RGBD video cameras). This algorithm has been tested on more than 70 videos of older adults performing 15 min of physical exercises and cognitive tasks. In Paris subway, we have been able to demonstrate the recognition in live of group behaviours. We have also been able to store the meta-data (e.g. people trajectories) generated from the processing of 8 video cameras, each of them lasting 2 or 3 days. From these meta-data, we have automatically discovered few hundreds of rare events, such as loitering, collapsing, ... to display on the screen of subway security operators.

Concerning semantic activity recognition, the contributions are :

- Evaluation of an Activity Monitoring System for Older People Using Fixed Cameras (6.16)
- A Framework for Activity Detection of Older People Using Multiple Sensors (6.17)
- Walking Speed Detection on a Treadmill using an RGB-D Camera (6.18)
- Serious Game for older adults with dementia (6.19)
- Unsupervised Activity Learning and Recognition (6.20)
- Extracting Statistical Information from Videos with Data Mining (6.21)

6.1.3. Software Engineering for Activity Recognition

Participants: François Brémond, Daniel Gaffé, Julien Gueytat, Sabine Moisan, Anh Tuan Nghiem, Annie Ressouche, Jean-Paul Rigault, Luis-Emiliano Sanchez.

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

The contributions for this research axis are:

- SUP (6.22)
- Model-Driven Engineering for Activity Recognition (6.23)
- Scenario Analysis Module (6.24)
- The Clem Workflow (6.25)
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.26)

6.2. Background Subtraction and People Detection in Videos

Participants: Vasanth Bathrinathan, Srinidhi Mukanahallipatna, Silviu Serban, François Brémond.

Keywords: Background Subtraction, People detection, Automatic parameter selection for algorithm

Background Subtraction Background subtraction is a vital real time low-level algorithm, which differentiates foreground and background objects in a video. We have thoroughly evaluated our Extended Gaussian Mixture model containing a shadows-removal algorithm, which performs better than other state of the art methods. Figure 10 shows the comparison of 13 background subtraction algorithms results on a challenging railway station monitoring video dataset from Project CENTAUR, which includes illumination change, shadows, occlusion and moving trains. Our algorithms performs the best in terms of result and with good processing speed too. Figure 11 is an example of our background subtraction algorithm's output on an indoor sequence of a surveillance footage from the Project SUPPORT.

Ongoing research include automatic parameter selection for this algorithm based on some learnt context. Since tuning the parameters is a daunting task for a non-experienced person, we try to learn some context information in a video like occlusion, contrast variation, density of foreground, texture etc. and map them to appropriate parameters of segmentation algorithm. Thus designing a controller to automatically adapt parameters of a algorithm as the scene context changes over time.

Algorithm	Speed	TP	FP	FN	Precision	Sensitivity	F-Score
DPAadaptiveMedian	13fps	79	951	2000	.076	.038	.050
Mixture of Gaussian	7.5fps	244	1217	1835	.167	.117	.137
DPPPratiMediod	6.4fps	266	2113	1813	.111	.127	.119
VuMeter	12.4fps	299	1311	1780	.185	.143	.162
T2FGMM_UV	3 fps	469	7161	1610	.061	.225	.096
DPWrenGABS	10.5fps	568	2603	1511	.179	.273	.216
MultiLayer(IDIAP)	2fps	713	1462	1366	.327	.342	.335

Algorithm	Speed	TP	FP	FN	Precision	Sensitivity	F-Score
Adaptive Background Learning	10.8fps	912	3410	1167	.211	.438	.285
DPZivkavoiGMM	8.5fps	943	2997	1136	.239	.453	.313
DPtexture	1 fps	1014	2430	1065	.294	.487	.367
PBAS	3 fps	1110	1413	969	.439	.534	.482
INRIA	14fps	1229	1788	850	.407	.591	.482
LBFuzzySOM	6fps	1161	2067	918	.359	.558	.437
GMG	5.79	1250	6671	829	.157	.60	.250

Figure 10. Background Subtraction result on confidential dataset from CENTAUR project

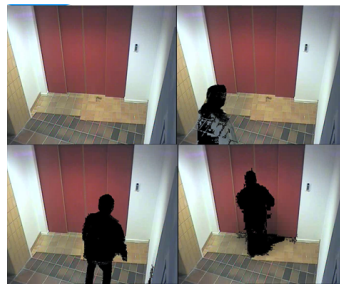


Figure 11. Background Subtraction result on a video to count number of people walking through the door after using their badge inside the terminal area (Project SUPPORT) - Autonomous Monitoring for Securing European Ports

People Detection

A new robust real-time person detection system was proposed [45], which aims to serve as solid foundation for developing solutions at an elevated level of reliability. Our belief is that clever handling of input data correlated with efficacious training algorithms are key for obtaining top performance. A comprehensive training method on very large training database and based on random sampling that compiles optimal classifiers with minimal bias and overfit rate is used. Building upon recent advances in multi-scale feature computations, our approach attains state-of-the-art accuracy while running at high frame rate.

Our method combines detection techniques that greatly reduce computational time without compromising accuracy. We use efficient LBP and MCT features which we compute on integral images for optimal retrieval of rectangular region intensity and nominal scaling error. AdaBoost is used to create cascading classifiers with significantly reduced detection time. We further refine detection speed by using the soft cascades approach and by transferring all important computation from the detection stage to the training stage. Figure 12 shows some output samples from various datasets which it was tested on.

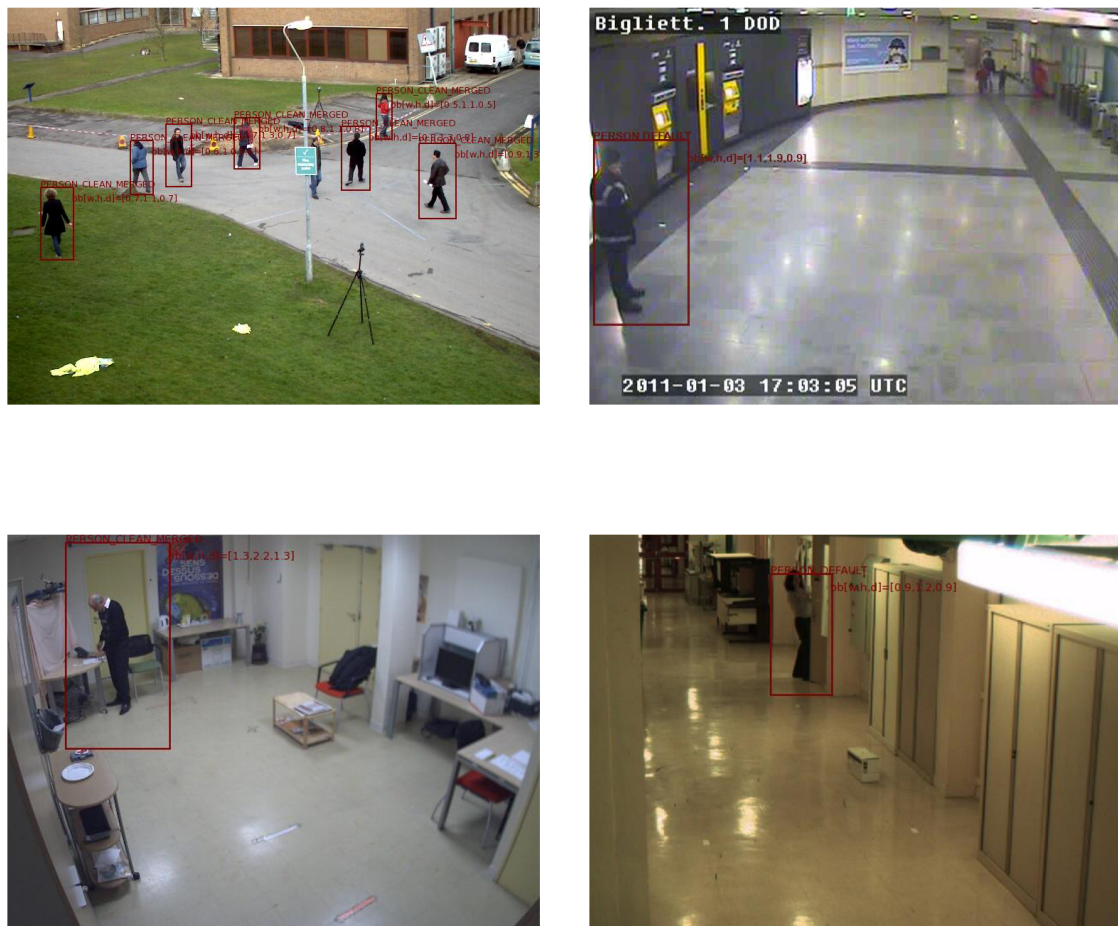


Figure 12. Detection output from PETS (left-top), VANAHEIM (right-top), Hospital (left-bottom) and ETISEO (right-bottom)

6.3. Tracking and Video Representation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

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

Multiple Object Tracking The objective is to find trajectories of objects (belonging to a particular category) in a video. To find possible occupancy locations, an object detector is applied to all frames of a video, yielding bounding boxes. Detectors are not perfect and may provide false detections; they may also miss objects sometimes. We build a graph of all detections, and aim at partitioning the graph into object trajectories. Edges in the graph encode factors between detections, based on the following :

- Number of common point tracks between bounding boxes (the tracks are obtained from an optical-flow-based point tracker)
- Global appearance similarity (based on the pixel colors inside the bounding boxes)
- Trajectory straightness : for three bounding boxes at different frames, we compute the Laplacian (centered at the middle frame) of the centroids of the boxes.
- Repulsive constraint : Two detections in a same frame cannot belong to the same trajectory.

We compute the partitions by using sequential tree re-weighted message passing (TRW-S). To avoid local minima, we use a label flipper motivated from the Iterative Conditional Modes algorithm.

We apply our approach to typical surveillance videos where object of interest are humans. Comparative quantitative results can be seen in Tables 1 and 2 for two videos. The evaluation metrics considered are : Recall, Precision, Average False Alarms Per Frame (FAF), Number of Groundtruth Trajectories (GT), Number of Mostly Tracked Trajectories, Number of Fragments (Frag), Number of Identity Switches (IDS), Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP).

This work has been submitted to CVPR' 14.

Table 1. Towncenter Video Output

Method	MOTA	MOTP	Detector
[59] (450-750)	56.8	79.6	HOG
Ours (450-750)	53.5	69.1	HOG

Table 2. Comparison with recent proposed approaches on PETS S2L1 Video

Method	Recall	Precision	FAF	GT	MT	Frag	IDS
[77]	96.9	94.1	0.36	19	18	15	22
Ours	95.4	93.4	0.28	19	18	42	13

Video Representation We continued our work from the previous year on Fiber-Based Video Representation. During this year we focused on obtaining competitive results with the state-of-the-art (Figure 13).

The usefulness of our novel representation is demonstrated by a simple video inpainting task. Here a user input of only 7 clicks is required to remove the dancing girl disturbing the news reporter (Figure 14).

This work has been accepted for publication next year [41].

6.4. Video segmentation with shape constraint

Participant: Guillaume Charpiat.

keywords: video segmentation, graph-cut, shape growth, shape statistics, shape prior, dynamic time warping

6.4.1. Video segmentation with growth constraint

This is joint work with Yuliya Tarabalka (Ayin Inria team) and Björn Menze (ETH Zurich, also MIT and collaborator of Asclepios Inria team).

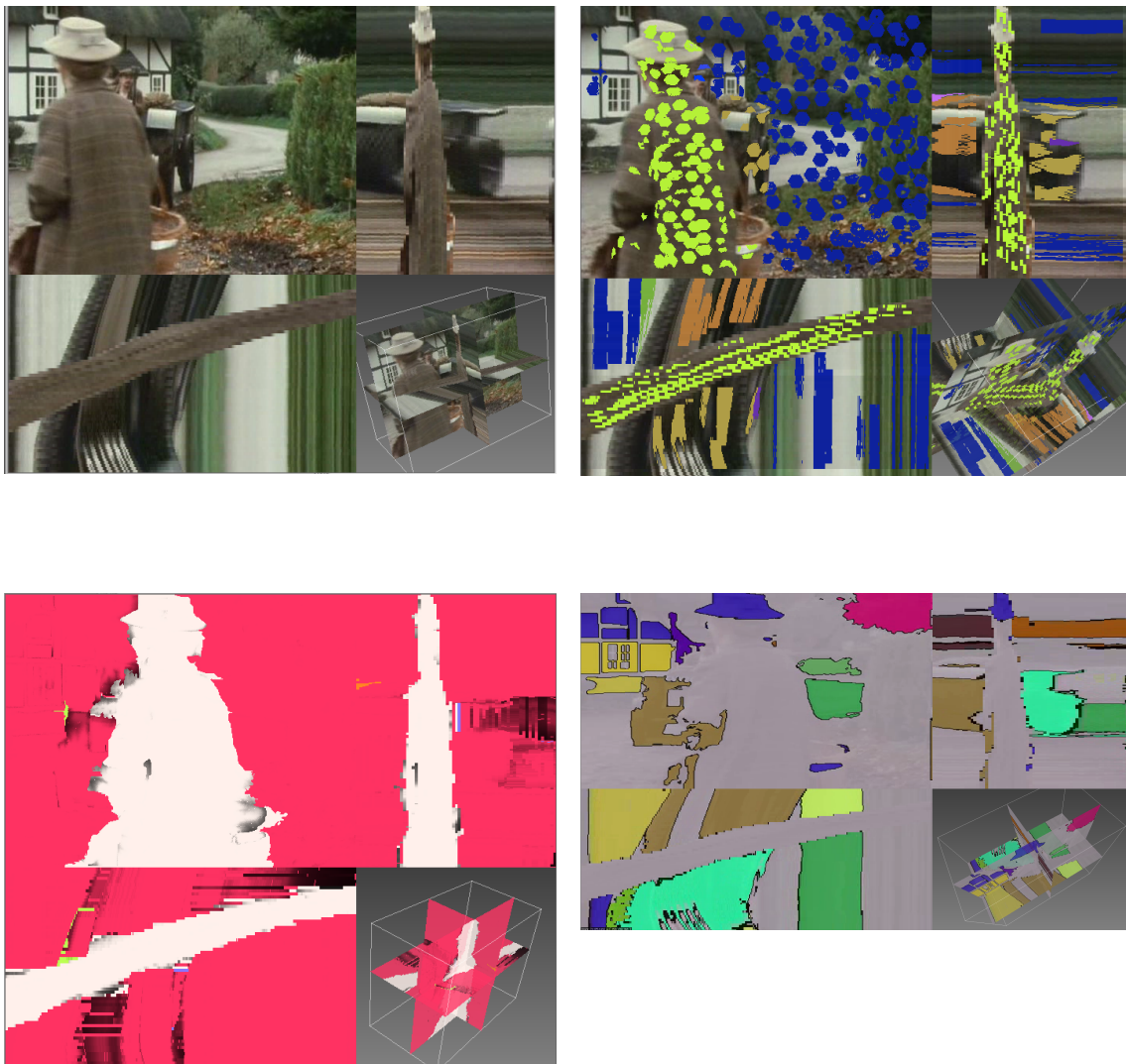


Figure 13. **Top Row:** Left image displays a sequence as a volumetric display. Right image displays all fibers found, clustered at a particular hierarchy. **Bottom Row :** Left Image displays the highest level of the hierarchical clustering, with fiber extension. Right Image shows the result obtained from [71]. **Our result demonstrates better long term temporal coherency.**



Figure 14. Inpainting task. **Left** : Original video (top) and xt slice (bottom) showing trajectories. **Right** : Our result. Clusters of fibers were computed and selected with only 7 mouse clicks to distinguish the disturbing girl from the reporter and background. The girl was removed and the hole was filled by extending the background fibers in time.

Context : One of the important challenges in computer vision is the automatic segmentation of objects in videos. This task becomes more difficult when image sequences are subject to low signal-to-noise ratio or low contrast between intensities of neighboring structures in the image scene. Such challenging data is acquired routinely, for example in medical imaging or in satellite remote sensing. While individual frames could be analyzed independently, temporal coherence in image sequences provides crucial information to make the problem easier. In this work, we focus on segmenting shapes in image sequences which only grow or shrink in time, and on making use of this knowledge as a constraint to help the segmentation process.

Approach and applications : We had proposed last year an approach based on graph-cut (see Figure 15), able to obtain efficiently (linear time in the number of pixels in practice), for any given video, its globally-optimal segmentation satisfying the growth constraint. This year we applied this method to three different applications :

- forest fires in satellite images,
- organ development in medical imaging (brain tumor, in multimodal MRI 3D volumes),
- sea ice melting in satellite observation, with a shrinking constraint instead of growth (see Figure 16).

The results on the first application were published in IGARSS (International Geoscience and Remote Sensing Symposium) [48], while the last two applications and the theory were published in BMCV [47]. A journal paper is also currently under review. A science popularization article was also published [53]. Not related but also with the Ayin Inria team was published the last of a series of articles about optimizers for point process models [40], introducing graph-cuts in the multiple birth and death approach in order to detect numerous objects that should not overlap.

6.4.2. Video segmentation with statistical shape prior

This is joint work with Maximiliano Suster (leader of the Neural Circuits and Behaviour Group at Bergen University, Norway).

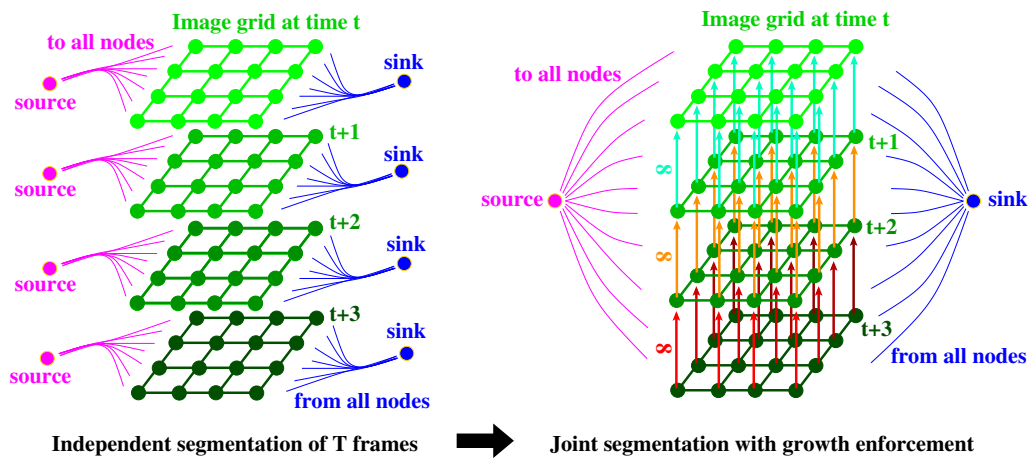


Figure 15. The approach : segmenting all frames together with a single graph-cut, with growth or shrinkage constraint, instead of segmenting independently each frame.

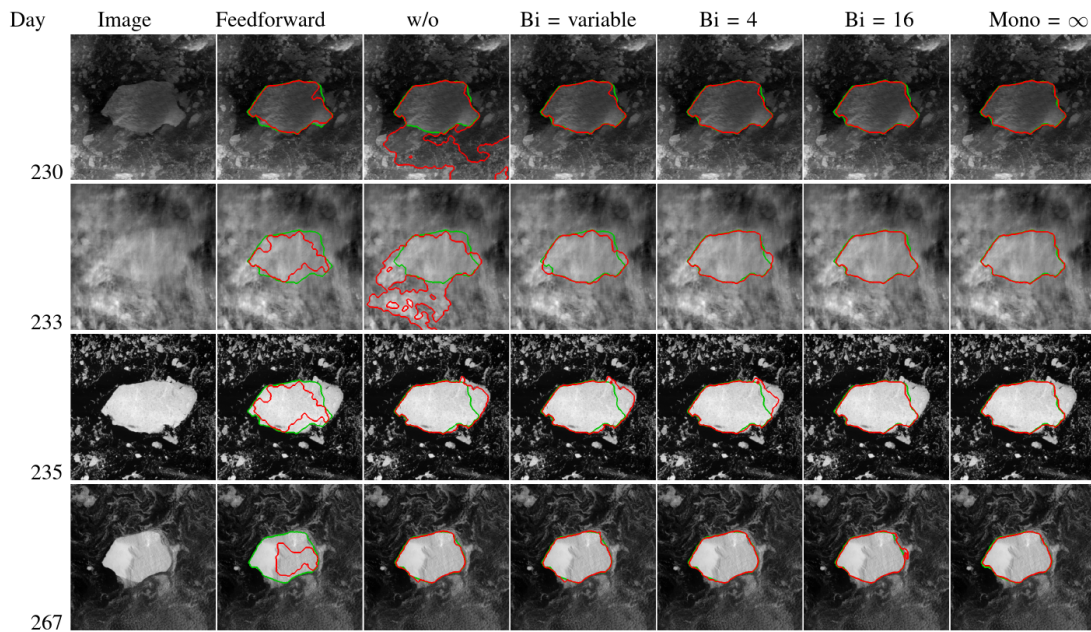


Figure 16. Example of a noisy, challenging sequence of melting sea ice, from satellite observations (first column). The frames were aligned beforehand, using inside/outside histograms of pixel intensities. The following columns show the results of varied approaches, ranging from frame-by-frame segmentation to techniques ensuring shape smoothness in time. Our approach (last column) has the right suitable prior (shape shrinkage) and thus performs the best.

Context : The zebrafish larva is a model organism widely used in biology to study genetics. Therefore, analyzing its behavior in video sequences is particularly important for this research field. For this, there is a need to segment the animal in the video, in order to estimate its speed, and also more precisely to extract its shape, in order to express for instance how much it is bent, how fast it bends, etc. However, as the animal is stimulated by the experimenter with a probe, the full zebrafish larva is not always visible because of occlusion.

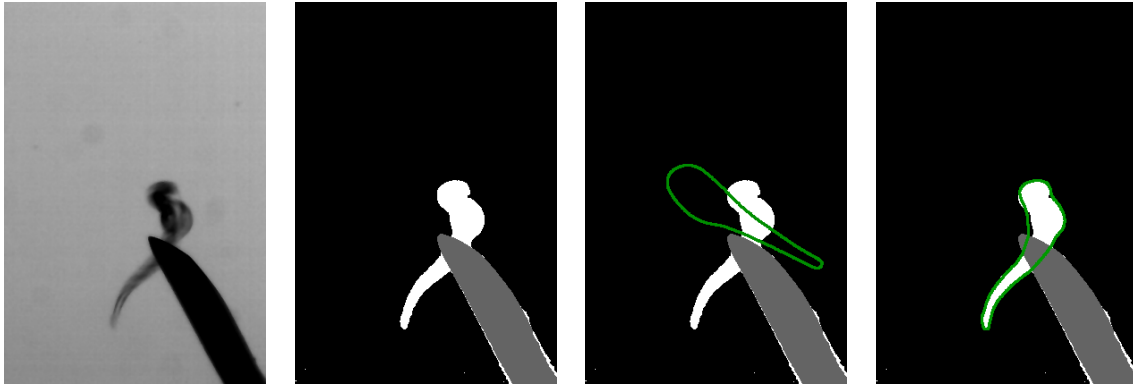


Figure 17. Example of a segmentation : initial image, processed image (based on video coherency), initialization of the active contour evolution, result. The total time spent per frame on average is reasonable for practical applications (magnitude order of 1 second).

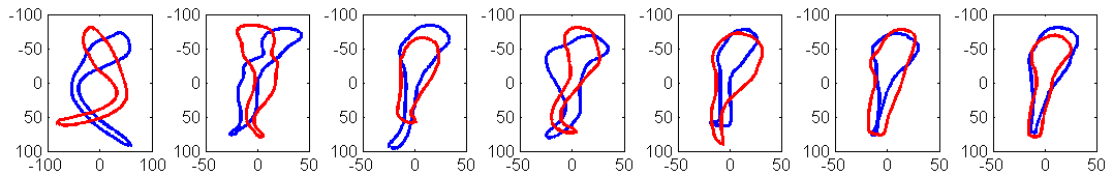


Figure 18. First deformation modes of the shape prior used in the segmentation above.

Approach : We build a shape prior based on a training set of examples of non-occluded shapes, and use it to segment new images where the animal is occluded. This is however not straightforward.

- **Building a training set of shape deformations :** Given a set of training images containing non-occluded animals, we extract their contours via multiple robust thresholdings and morphomathematical operations. For each contour, we then estimate automatically the location of the tip of the tail. We then compute point-to-point correspondences between all contours, using a modified version of Dynamic Time Warping, as well as the approximate tip location information. This is done in a translation- and rotation-invariant way.
- **Building the shape prior :** Based on these matchings, the mean shape is computed, as well as modes of deformation with PCA.
- **Segmenting occluded images :** Images with occluded shapes are pre-processed in a similar way to non-occluded ones; however, the resulted segmentation does not contain only the parts of the larva but also the probe, which has potentially similar colors and location, and is moving. To identify the

probe, whose shape depends on the video sequence, we make use of its rigidity and of temporal coherency. Then a segmentation criterion is designed to push an active contour towards the zones of interest (in a way that is robust to initialization), while keeping a shape which is feasible according to the shape prior.

Examples of data and results for a preliminary algorithm are shown in Figure 17, with the associated shape prior shown in Figure 18.

6.5. Articulating motion

Participant: Guillaume Charpiat.

keywords: shape evolution, metrics, gradient descent, Finsler gradient, Banach space, piecewise-rigidity, piecewise-similarity

This is joint work with Giacomo Nardi, Gabriel Peyré and François-Xavier Vialard (Ceremade, Paris-Dauphine University).

Context in optimization : A fact which is often ignored when optimizing a criterion with a gradient descent is that the gradient of a quantity depends on the metric chosen. In many domains, people choose by default the underlying L^2 metric, while it is not always relevant. Here we extend the set of metrics that can be considered, by building gradients for metrics that do not derive from inner products, with examples of metrics involving the L^1 norm, possibly of a derivative.

Mathematical foundations : This work introduces a novel steepest descent flow in Banach spaces. This extends previous works on generalized gradient descent, notably the work of Charpiat et al. [6], to the setting of Finsler metrics. Such a generalized gradient allows one to take into account a prior on deformations (e.g., piecewise rigid) in order to favor some specific evolutions. We define a Finsler gradient descent method to minimize a functional defined on a Banach space and we prove a convergence theorem for such a method. In particular, we show that the use of non-Hilbertian norms on Banach spaces is useful to study non-convex optimization problems where the geometry of the space might play a crucial role to avoid poor local minima.

Application to shape evolution : We performed some applications to the curve matching problem. In particular, we characterized piecewise-rigid deformations on the space of curves and we studied several models to perform piecewise-rigid evolutions (see Figure 19). We also studied piecewise-similar evolutions. Piecewise-rigidity intuitively corresponds to articulated motions, while piecewise-similarity further allows the elastic stretching of each articulated part independently. One practical consequence of our work is that any deformation to be applied to a shape can be easily and optimally transformed into an articulated deformation with few articulations, the number and location of the articulations being not known in advance. Surprisingly, this problem is actually convex.

An article was submitted to the journal *Interfaces and Free Boundaries* [52].

6.6. Lossless image compression

Participant: Guillaume Charpiat.

keywords: image compression, entropy coding, graph-cut

This is joint work with Yann Ollivier and Jamal Atif from the TAO Inria team.

Context : Understanding, modelling, predicting and compressing images are tightly linked, in that any good predictor can be turned into a good compressor via entropy coding (such as Huffman coding or arithmetic coding). Indeed, with such techniques, the more predictable an event E is, i.e. the higher its probability $p(E)$, the easier to compress it will be, with coding cost $-\log(p(E))$. Therefore we are interested in image compression, in order to build better models of images.

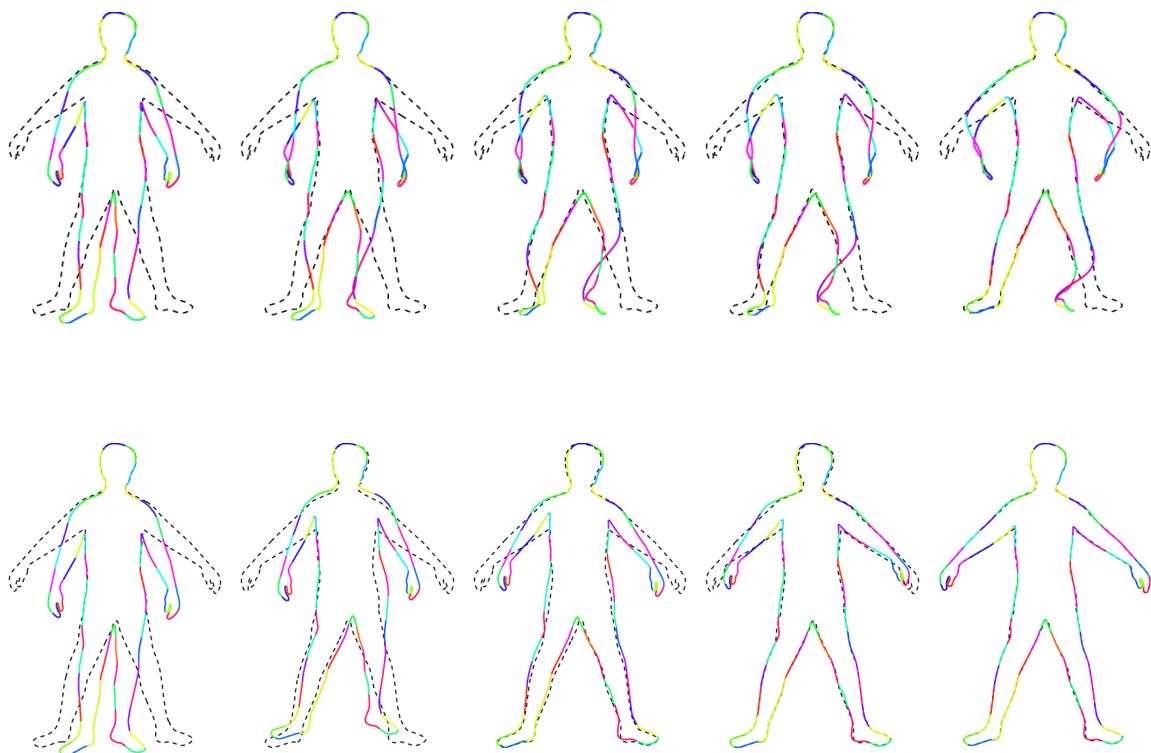


Figure 19. Example of use of the Finsler gradient for the piecewise-rigid evolution of curves. Given an initial shape S and a target shape T , as well as a shape dissimilarity measure $E(S) = \text{Dissim}(S, T)$, any classical gradient descent on $E(S)$ would draw the evolving shape S towards the target T . However the metric considered to compute the gradient changes the path followed. The top row is the evolution obtained with a Sobolev gradient H^1 , which has the property of smoothing spatially the flow along the curve, to avoid irregular deformations. This is however not sufficient. The bottom row makes use of the Finsler gradient instead, with a metric favoring piecewise-rigid deformations.

MDL approach : The state-of-the-art sequential prediction of time series based on the advice of various experts combines the different expert predictions, with weights depending on their individual past performance (cf. Gilles Stoltz and Peter Grünwald’s work). This approach originates from the Minimum Description Length principle (MDL). This work was however designed for 1D data such as time series, and is not directly applicable to 2D data such as images. Consequently, our aim has been to adapt such an approach to the case of image compression, where time series are replaced with 2D series of pixel colors, and where experts are predictors of the color of a pixel given the colors of neighbors.

New method and results : This year, we have focused on lossless greyscale image compression, and proposed to encode any image with two maps, one storing the choice of the expert made for each pixel, and one storing the encoding of the intensity of each pixel according to its expert. In order to compress efficiently the first map, we ask the choices of experts to be coherent in space, and then encode the boundaries of the experts’ areas. To find a suitable expert map, we optimize the total encoding cost explicitly, set as an energy minimization problem, solved with graph-cuts. An example of expert map obtained is shown in Figure 20. Preliminary results with a hierarchical ordering scheme already compete with standard techniques in lossless compression (PNG, lossless JPEG2000, JPEG-LS).

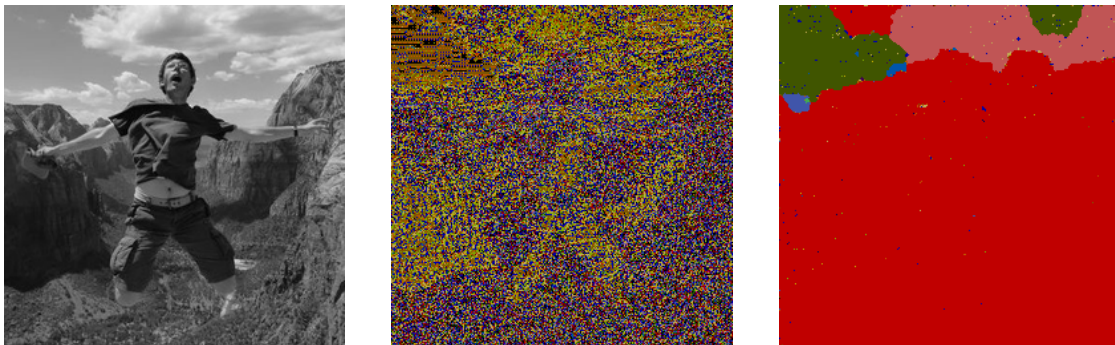


Figure 20. An image to encode; the map of the best experts, chosen independently for each pixel [each expert is represented by one color]; the expert map obtained with our approach enforcing spatial coherency in the expert choice.

6.7. People detection using RGB-D cameras

Participants: Anh-Tuan Nghiem, François Brémond.

keywords: people detection, HOG, RGB-D cameras

With the introduction of low cost RGB-D cameras like Kinect of Microsoft, video monitoring systems have another option for indoor monitoring beside conventional RGB cameras. Comparing with conventional RGB camera, reliable depth information from RGB-D cameras makes people detection easier. Besides that, constructors of RGB-D cameras also provide various libraries for people detection, skeleton detection or hand detection etc. However, perhaps due to high variance of depth measurement when objects are too far from the camera, these libraries only work when people are in the range of 0.5 to around 4.5 m from the cameras. Therefore, for our own video monitoring system, we construct our own people detection framework consisting of a background subtraction, a people classifier, a tracker and a noise removal component as illustrated in figure 21.

In this system, the background subtraction algorithm is designed specifically for depth data. Particularly, the algorithm employs temporal filters to detect noise related to imperfect depth measurement on some special surface.

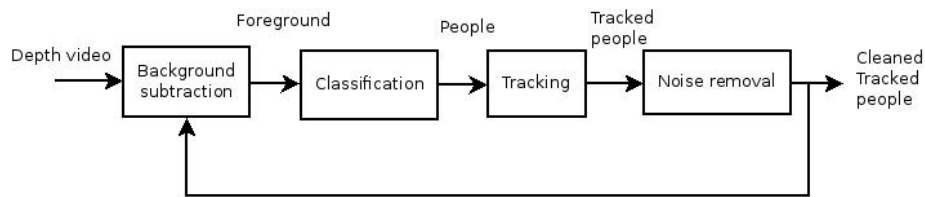


Figure 21. The people detection framework

The people classification part is the extension of the work in [79]. From the foreground region provided by the background subtraction algorithm, the classification first searches for people head and then extracts HOG like features (Histogram of Oriented Gradient on binary image) above the head and the shoulder. Finally, these features are classified by a SVM classifier to recognise people.

The tracker links detected foreground regions in the current frame with the ones from previous frames. By linking objects in different frames, the tracker provides useful history information to remove noise as well as to improve the sensitivity of the people classifier.

Finally, the noise removal algorithm uses the object history constructed by the tracker to remove two types of noise: noise detected by temporal filter at the background subtraction algorithm and noise from high variance of depth measurement on objects far from the camera. Figure 22 illustrates the performance of noise removal on the detection results.



Figure 22. The people detection framework

The overall performance of our people detection framework is comparable to the one provided by Primesense, the constructor of RGB-D camera Microsoft Kinect.

Currently, we are doing extensive evaluation of the framework and the results will be submitted to a conference in the near future.

6.8. Online Tracking Parameter Adaptation based on Evaluation

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

Keywords: Object tracking, parameter tuning, online evaluation, machine learning

Several studies have been proposed for tracking mobile objects in videos [50]. For example we have proposed recently a new tracker which is based on co-inertia analysis (COIA) of object features [44]. However the parameter tuning is still a common issue for many trackers. In order to solve this problem, we propose an online parameter tuning process to adapt a tracking algorithm to various scene contexts. The proposed approach brings two contributions: (1) an online tracking evaluation, and (2) a method to adapt online tracking parameters to scene contexts.

In an offline training phase, this approach learns how to tune the tracker parameters to cope with different contexts. Different learning schemes (e.g. neural network-based) are proposed. A context database is created at the end of this phase to support the control process of the considered tracking algorithm. This database contains satisfactory parameter values of this tracker for various contexts.

In the online control phase, once the tracking quality is evaluated as not good enough, the proposed approach computes the current context and tunes the tracking parameters using the learned values.

The experimental results show that the proposed approach improves the performance of the tracking algorithm and outperforms recent state of the art trackers. Figure 23 shows the correct tracking results of four people while occlusions happen. Table 3 presents the tracking results of the proposed approach and of some recent trackers from the state of the art. The proposed controller increases significantly the performance of an appearance-based tracker [63]. We obtain the best *MT* value (i.e. mostly tracked trajectories) compared to state of the art trackers.

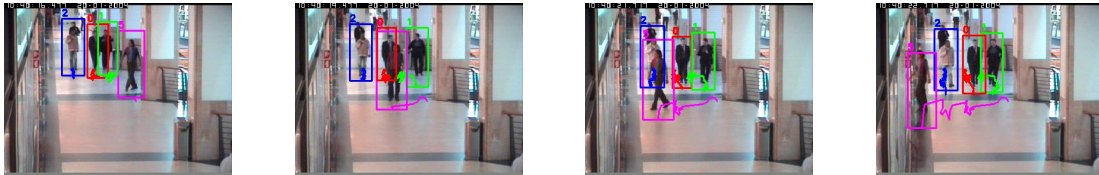


Figure 23. Tracking results of four people in the sequence *ShopAssistant2cor* (Caviar dataset) are correct, even when occlusions happen.

Table 3. Tracking results for the Caviar dataset. The proposed controller improves significantly the tracking performance. MT: Mostly tracked trajectories, higher is better. PT: Partially tracked trajectories. ML: Most lost trajectories, lower is better. The best values are printed in bold.

Approaches	MT (%)	PT (%)	ML (%)
Xing et al. [92]	84.3	12.1	3.6
Li et al. [76]	84.6	14.0	1.4
Kuo et al. [74]	84.6	14.7	0.7
D.P Chau et al. [63] without the proposed approach	78.3	16.0	5.7
D.P Chau et al. [63] with the proposed approach	85.5	9.2	5.3

This work has been published in [33], [34].

6.9. People Detection, Tracking and Re-identification Through a Video Camera Network

Participants: Malik Souded, François Brémond.

keywords: People detection, Object tracking, People re-identification, Region covariance descriptors, SIFT descriptor, LogitBoost, Particle filters.

This work aims at proposing a whole framework for people detection, tracking and re-identification through camera networks. Three main constraints have guided this work: high performances, real-time processing and genericity of the proposed methods (minimal human interaction/parametrization). This work is divided into three separate but dependent tasks:

6.9.1. People detection:

The proposed approach optimizes state-of-the-art methods [89], [93] which are based on training cascades of classifiers using the LogitBoost algorithm on region covariance descriptors. The optimization consists in clustering negative data before the training step, and speeds up both the training and detection processes while improving the detection performance. This approach has been published this year in [46]. The evaluation results and examples of detection are shown in Figures 24 and 25.

6.9.2. Object tracking:

The proposed object tracker uses a state-of-the-art background subtraction algorithm to initialize objects to track, with a collaboration of the proposed people detector in the case of people tracking. The object modelling is performed using SIFT features, detected and selected in a particular manner. The tracking process is performed at two levels: SIFT features are tracked using a specific particle filter, then object tracking is deduced from the tracked SIFT features using the proposed data association framework. A fast occlusion management is also proposed to achieve the object tracking process. The evaluation results are shown in Figure 26.

6.9.3. People re-identification:

A state-of-the-art method for people re-identification [67] is used as a baseline and its performance has been improved. A fast method for image alignment for multiple-shot case is proposed first. Then, texture information is added to the computed visual signatures. A method for people visible side classification is also proposed. Camera calibration information is used to filter candidate people who do not match spatio-temporal constraints. Finally, an adaptive feature weighting method according to visible side classification concludes the improvement contributions. The evaluation results are shown in Figure 27.

This work has been published in [28].

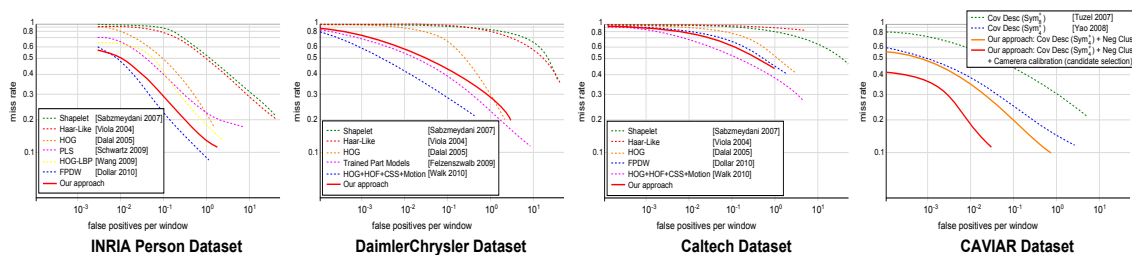


Figure 24. People detector evaluation and comparison on Inria, DaimlerChrysler, Caltech and CAVIAR datasets.

6.10. People Retrieval in a Network of Cameras

Participants: Sławomir Bąk, Marco San Biago, Ratnesh Kumar, Vasanth Bathrinathan, François Brémont.

keywords: Brownian statistics, re-identification, retrieval

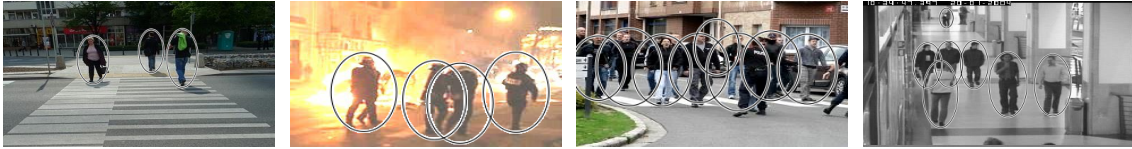


Figure 25. Some examples of detection using the proposed people detector.

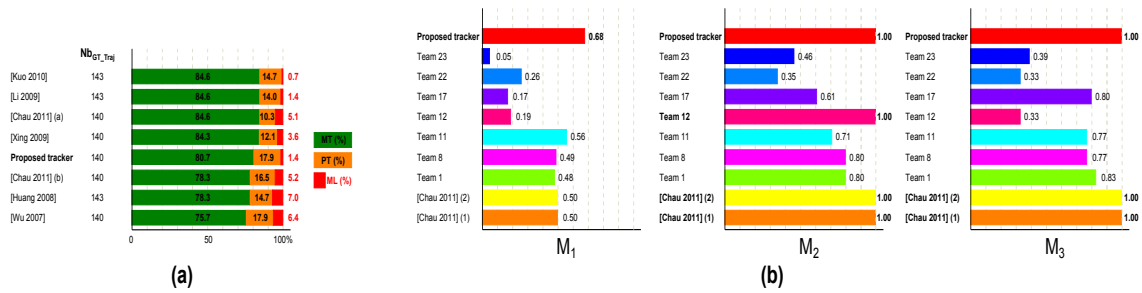


Figure 26. Object tracking evaluation on: (a) CAVIAR dataset using MT, PT and ML metrics. (b) ETI-VSI-BE-18-C4 sequence from ETISEO dataset, using ETISEO metrics.

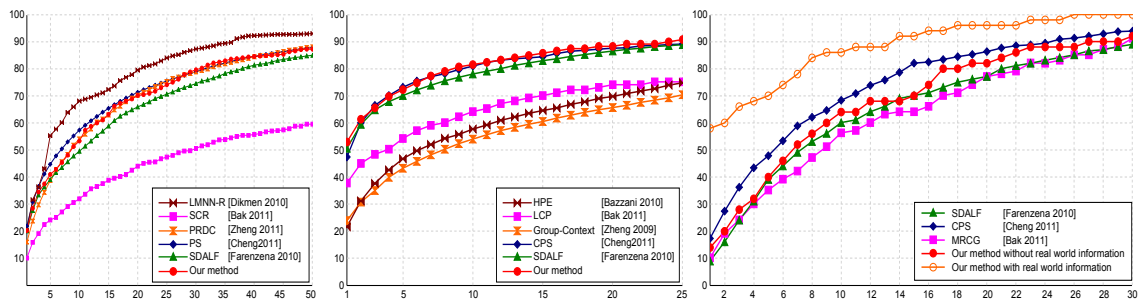


Figure 27. People re-identification evaluation on VIPeR (left), iLids-119 (middle) and CAVIAR4REID (right) datasets.

Task. Person re-identification (also known as multi-camera tracking) is defined as the process of determining whether a given individual has already appeared over a network of cameras. In most video surveillance scenarios, features such as face or iris are not available due to video low-resolution. Therefore a robust modeling of the global appearance of an individual (clothing) is necessary for re-identification. This problem is particularly hard due to significant appearance changes caused by variations in view angle, lighting conditions and different person pose. In this year, we focused on the two following aspects: new image descriptors and a design of a retrieval tool.

New image region descriptors. We have evaluated different image descriptors *w.r.t.* their recognition accuracy. As the covariance descriptor achieved the best results, we have employed this descriptor using different learning strategies to achieve the most accurate model for representing a human appearance [51]. We have also proposed a new descriptor based on recent advances in mathematical statistics related to Brownian motion [31]. This new descriptor outperforms the classical covariance in terms of matching accuracy and efficiency. We show that the proposed descriptor can capture richer characteristics than covariance, especially when fusing nonlinearly dependent features, which is often the case for images. The effectiveness of the approach is validated on three challenging vision tasks: object tracking & person re-identification [31] and pedestrian classification (the paper submitted to conference CVPR 2014). In all our experiments, we demonstrate competitive results while in person re-identification and tracking we significantly outperform the state-of-the-art.

New design of retrieval tool for a large network of cameras. Owing to the complexity of the re-identification problem, current state of the art approaches have relatively low retrieval accuracy, thus a fully automated system is still unattainable. However, we propose a retrieval tool [30], [29] that helps a human operator to solve the re-identification task (see Figure 28). This tool allows a human operator to browse images of people extracted from a network of cameras: to detect a person on one camera and to re-detect the same person few minutes later on another camera. The main stream is displayed on the left of the screen, while retrieval results are shown on the right. The results show lists of the most similar signatures extracted from each camera (green boxes indicate the correctly retrieved person). Below the main stream window a topology of the camera network is displayed. Detection and single camera tracking (see the main stream) are fully automatic. The human operator only needs to select a person of interest, thus producing retrieval results (right screen). The operator can easily see a preview of the retrieval results and can go directly to the original video content.

Perspectives. Currently, we are working not only on invariant image descriptors, which provide high recognition accuracy, but also on improving the alignment of the person pose, while matching appearance from cameras with significant difference in viewpoint. In addition to re-identification technology, we also work on designing an intuitive graphical interface, an important tool for the human operator analyzing retrieval results. Displaying retrieval results from a large camera network is still an issue, even after applying time-space constraints (the usage of topology of cameras).

Acknowledgements

This work has been supported by PANORAMA and CENTAUR European projects.

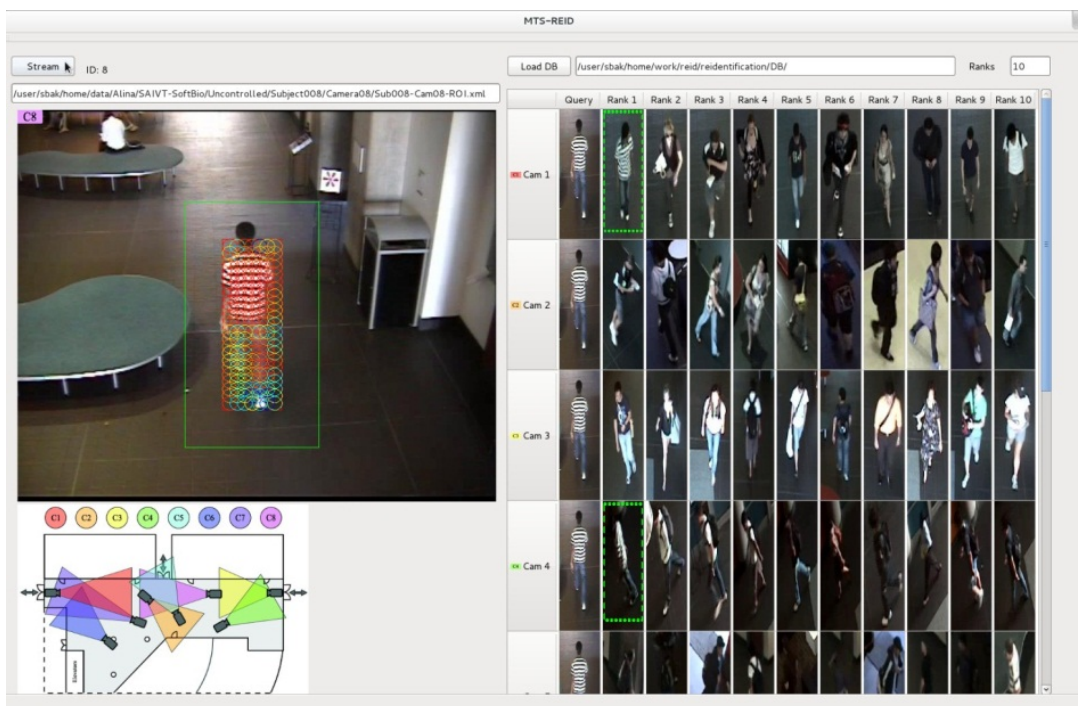
6.11. Global Tracker : an Online Evaluation Framework to Improve Tracking Quality

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

keywords: online quality estimation, improving tracking results

This work addresses the problem of estimating the quality of a tracking algorithm during runtime and correcting the anomalies found. Evaluating and tuning a tracking algorithm generally requires multiple runs and a ground truth. The proposed online evaluation framework, called global tracker, overcome these limitations by proposing a three-steps algorithm to improve tracking results in any kind of situations (mono-camera, camera network, 3D camera, ...).

The first step aims at correcting small tracking errors (when detections in consecutive frames are missing from an object trajectory) by interpolating the detected object data.



(a)

Figure 28. Re-identification tool

The second step aims at detecting and correcting any anomaly found in the output of the tracking algorithm. For each detected object on each frame, we compute three different sets of features : the features that are computed using only data from the object (eg. appearance, size, ...), the features that measure the level of interaction between two objects (eg. occlusion level, density) and the features that measure the level of interaction between the object and the environment (eg. occlusion with background element, entering or leaving zones, ...). By studying the evolution of the coefficients of variation of each feature, some anomalies can be detected. Then, by minimizing an energy function based on the object-only features, we correct the output of the tracking algorithm.

The final step uses re-acquisition and re-identification methods to merge detected objects representing the same real object. This algorithm aims at correcting ID change errors when an object leaves the scene and reappears in another camera or when the object reappears after a long-term occlusion. The method used here is a constrained clustering algorithm that create groups of detections representing the same real object.

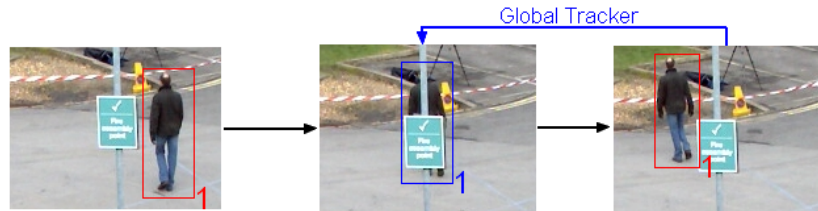


Figure 29. An example of the global tracker goal : offline learning of occlusion zones to manage online occlusion risks and optimize object trajectory.

This approach has been tested on several datasets (PETS 2009, Caviar, TUD-Stadtmitte). The results show that the global tracker, even associated with a tracking algorithm that does not have good results, can perform nearly as well as the state of the art and even exactly as well when a good tracker is used. On the Caviar dataset, the global tracker is even able to slightly outperform the result of the state of the art.

Table 4. Comparison of the tracking results using CLEAR metrics on the sequence S2.L1 of the PETS 2009 dataset with and without the global tracker

Method	MOTA	MOTP	\bar{M}
Berclaz et al. [60]	0.80	0.58	0.69
Shitrit et al. [58]	0.81	0.58	0.70
Henriques et al. [72]	0.85	0.69	0.77
Chau et al. [33] without global tracker	0.62	0.63	0.63
Chau et al. [33] with global tracker	0.85	0.71	0.78

A part of this approach is described in an article published in AVSS 2013 [33]. This article shows how a tracking algorithm can use the global tracker results to tune its parameters and improve its results. This work was also used to improve the tracking results in 2 papers [38] [54].

6.12. Human Action Recognition in Videos

Participants: Piotr Bilinski, Etienne Corvée, Slawomir Bak, François Brémond.

keywords: action recognition, tracklets, head detection, relative tracklets, bag-of-words.

In this work we address the problem of recognizing human actions in video sequences for home care applications.

Recent studies have shown that approaches which use a bag-of-words representation reach high action recognition accuracy. Unfortunately, these approaches have problems to discriminate similar actions, ignoring spatial information of features.

We propose a feature representation for action recognition based on dense point tracklets, head position estimation, and a dynamic coordinate system. Our main idea is that action recognition ought to be performed using a dynamic coordinate system corresponding to an object of interest. Therefore, we introduce a relative tracklet descriptor based on relative positions of a tracklet according to the central point of our dynamic coordinate system. As a center of our dynamic coordinate system, we choose the head position, providing description invariant to camera viewpoint changes. We use the bag-of-words approach to represent a video sequence and we capture global distribution of tracklets and relative tracklet descriptors over a video sequence. The proposed descriptors introduce spatial information to the bag-of-words model and help to distinguish similar features detected at different positions (*e.g.* to distinguish similar features appearing on hands and feet). Then we apply the Support Vector Machines with exponential chi-squared kernel to classify videos and recognize actions.

We report experimental results on three action recognition datasets (publicly available KTH and ADL datasets, and our locally collected dataset). Our locally collected dataset has been created in cooperation with the CHU Nice Hospital. It contains people performing daily living activities such as: standing up, sitting down, walking, reading a magazine, *etc.* Consistently, experiments show that our representation enhances the discriminative power of the tracklet descriptors and the bag-of-words model, and improves action recognition performance.

Sample video frames with extracted tracklets and estimated head positions are presented in Figure 30.

This work has been published in [32].

6.12.1. Acknowledgments

This work was supported by the Région Provence-Alpes-Côte d’Azur. However, the views and opinions expressed herein do not necessarily reflect those of the financing institution

6.13. 3D Trajectories for Action Recognition Using Depth Sensors

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], [62], [32], [90]) for Human Action Recognition to take advantage of the depth information from 3D sensors.

We propose to add depth information to trajectory based algorithms ([32], [90]). 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 information map. Such computed 3D trajectories, combined with other appearance features (HOG, HOF), are subject to a Bag of Words model and SVM classifier.

The evaluation of our method was conducted on the "Microsoft Daily Activity3D" data set [91] which consist of 16 actions (drink, eat, read book, call cellphone, write on a paper, use laptop *etc.*) performed by 10 subjects. The experiments showed that adding depth information to Dense Trajectories descriptor [90] gave gain in efficiency 57.72% to 64.12%. The mentioned work is going to be submitted in December 2013.

6.14. Unsupervised Sudden Group Movement Discovery for Video Surveillance

Participants: Sofia Zaidenberg, Piotr Bilinski, François Brémond.

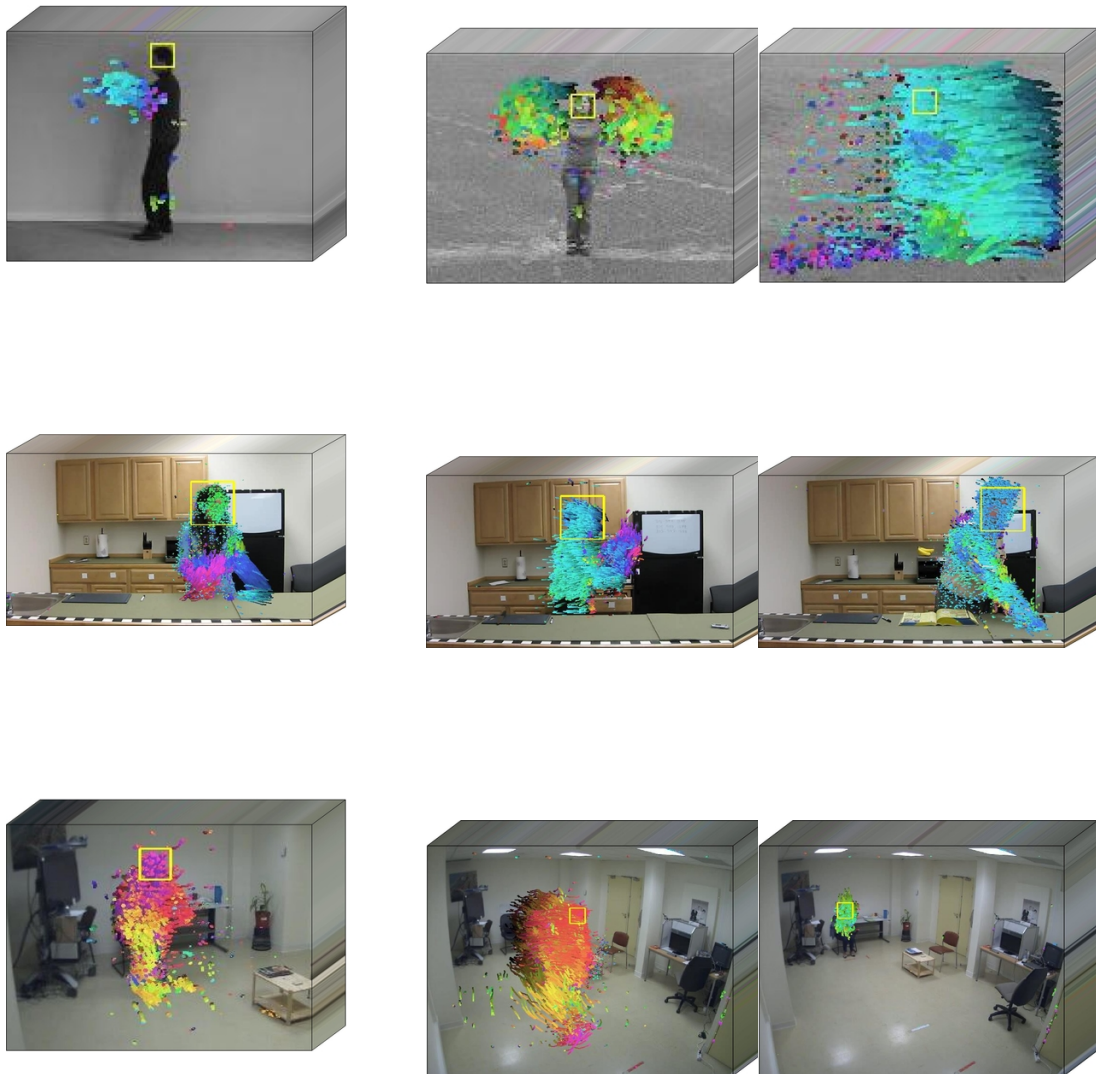


Figure 30. Sample video frames with extracted tracklets and estimated head positions for the KTH (first row), ADL (second row) and our locally collected dataset (third row).

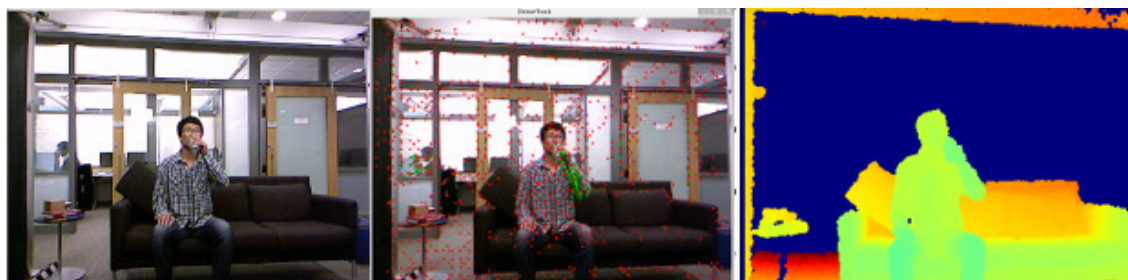


Figure 31. Visualization of MSR Daily Activity 3D data set. Left : video input frame; Middle : frame with detected trajectories (red = static points, green = detected trajectories); Right : corresponding depth map.

keywords: Sudden Group Movement Discovery, Video Surveillance.

In this work we address the problem of discovering “sudden” movements in video surveillance videos. We propose an unsupervised approach which 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 provides a semantic interpretation of the scene. We have tested our approach on a dataset of nearly eight hours of videos recorded from two cameras in the Parisian subway for a European Project. For evaluation we annotated one hour of sequences containing 50 sudden movements. Our system, if parametrized to a high sensitivity, detects 100% of what the annotator considered as sudden potentially dangerous events, with a false positive rate of 21.2%. Setting the sensitivity to lower values we decrease the false positive rate to only 5.3% but we also decrease the success rate to 76%. An example of an unusual sudden movement annotated by a human and detected by our approach is presented in Figure 32. This work has been published in [49].

6.14.1. Acknowledgments

This work was supported by the Région Provence-Alpes-Côte d’Azur and by the European Community’s Seventh Framework Programme FP7/2007-2013 - Challenge 2 - Cognitive Systems, Interaction, Robotics - under grant agreement number: 248907-VANAHEIM. However, the views and opinions expressed herein do not necessarily reflect those of the financing institution.

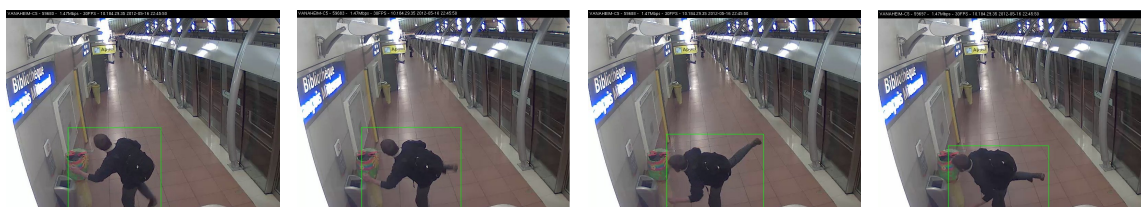


Figure 32. Example of an unusual sudden movement detected by our approach.

6.15. Group Behavior Understanding

Participants: Carolina Gárate, Sofia Zaidenberg, Julien Badie, François Brémond.

The goal is to recognize group behavior from videos. Dangerous and criminal behaviors are mostly observed within groups of people. The idea is to detect potentially dangerous situations while they are happening in the context of underground railway station security.

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

This research work considers a process consisting of 5 consecutive steps for video processing. The steps are: 1) segmentation, 2) blob detection, 3) physical objects tracking, 4) group tracking and 5) behavior recognition. Here, we are focussing on the last two phases: group tracking and behavior recognition.

The group tracking approach characterizes a group through three features: the average of the intra-object distance, the average standard deviations of speed and direction. The input for this algorithm is a set of trajectories for the physical objects (output of the stage 3: physical objects tracking) tracked by the algorithm described in [64]. The trajectories are processed using Mean-Shift clustering to create more reliable groups, see Figure 33.

The behavior recognition approach identifies 2 steps: knowledge modeling and the event recognition algorithm. The ontology is implemented with the ScReK declarative language [94]. The grammar describes the objects and events using the extended BNF (Backus Naur Form) representation.

We process large amounts of long video surveillance data from Paris and Turin underground railway station to perform statistical analysis. This analysis automatically brings forward data about the usage of the station and the various behaviors of groups for different hours of the day. We present the results and interpretation of one month of processed data from a video surveillance camera in Turin subway.



Figure 33. Illustration of a still group: in the right entrance zone a group of subway security staff is standing still.

One of the measures obtained in the experimentation is the agitation level which is represented by the variation of the size of the bounding box of a group. We consider 3 categories from *no agitation* ("Calm_Group", having a bounding box with stable size) to *little agitation* ("Active_Group") to *high agitation* ("Lively_Group", the bounding box's size varies a lot, meaning that group members move around more often). Figure 34 shows that most of the time, this middle category predominates. Groups are neither too calm, nor too agitated. Moreover, it is more common for a group to be lively rather than calm. The rest of the results obtained were presented in [54].

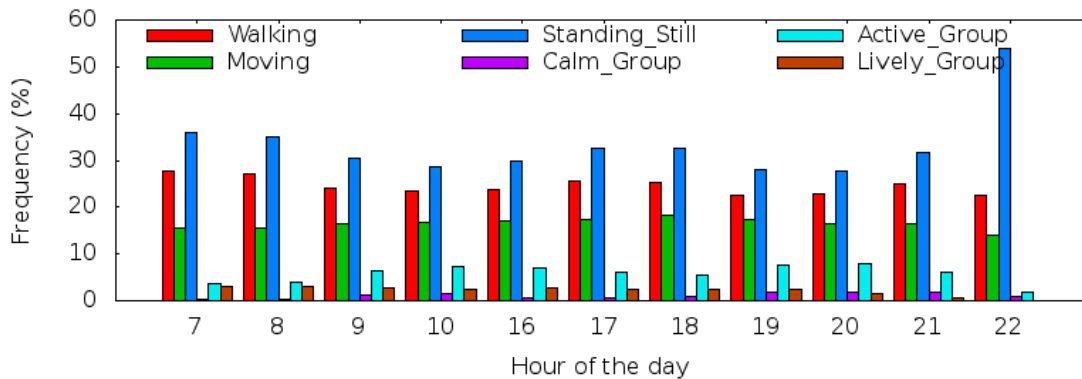


Figure 34. Comparison of speed of groups and variation of group's bounding boxes during the day.

6.16. Evaluation of an Activity Monitoring System for Older People Using Fixed Cameras

Participants: Carlos F. Crispim-Junior, Baptiste Fosty, Vasanth Bathrinarayanan, Salma Zouaoui-Elloumi, Monique Thonnat, François Brémond.

keywords: 2D-RGB cameras, RGB-D cameras, model-based activity recognition, older people

We have continued the evaluation of our model-based algorithm for complex activity recognition, now extending it to a larger dataset containing 38 older people participants undertaking instrumental activities of daily living (IADL) during 15 minutes (570 min. in total). The recordings have taken place in the observation room of the Memory Center of Nice hospital. Figure 35 presents the algorithm performance based on data obtained from a 2D-RGB video camera. A summary of the recognized activities (e.g., duration, frequency) is produced at the end of the event recognition task to be provided to doctors as a basis for the assessment of patient performance on IADL. This approach description and the evaluation results are published in 2013 AVSS Conference (see details in [36]). Figure 36 illustrates an example of a patient being monitored. Blue dots illustrates previous positions of the person in the scene.

The proposed approach has been also evaluated using a RGB-D camera, as this camera increases the robustness of the monitoring system against environment illumination changes and also eases the deployment of the system by providing real 3-D information on the scene. The evaluation of the RGB-D-based activity monitoring system has been published in [38]. A live demonstration of this system has also been presented and applied in the scope of Dem@care project (a FP7 project devoted to multi-sensor older people monitoring) in the exhibition held in November 2013 in conjunction with the 2013 edition of the ICT (Information Communication Technologies) Conference in Vilnius - Lithuania.

6.17. A Framework for Activity Detection of Older People Using Multiple Sensors

Participants: Carlos F. Crispim-Junior, Qiao Ma, Baptiste Fosty, François Brémond, Monique Thonnat.

keywords: model-based activity recognition, multi-sensor, Dempster-Shafer, Evidence Theory, older people.

Event	Sensitivity (%)	Precision (%)
Using Phone	77.77	82.89
Watching TV	84.00	67.74
Using_Office Desk	94.54	50.98
Using Coffee Machine	90.36	69.44
Using Pharmacy Basket	100.00	78.72
Watering Plant	100.00	60.65
Reading	71.42	88.23
Average Performance	88.30	71.24

Figure 35. Evaluation of Algorithm performance on 38 patient video recordings using a RGB camera

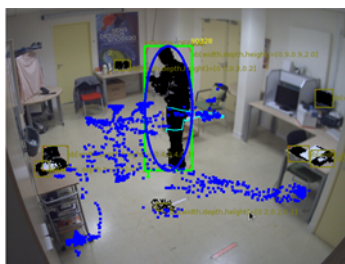


Figure 36. Example of a patient been monitored by the described system

We have extended our framework for multi-sensor activity detection by proposing a probabilistic approach for mutually exclusive activity conflict scenario. The proposed approach pre-learned a coefficient of reliability of each sensor with respect to each activity. The combination of the activities detected by multiple sensors is performed using the Dempster-Shafer evidence theory with an adapted combination rule based on runtime data from the sensor and the pre-computed coefficients of reliability. The combination of activities detected by multiple sensors can be performed in an iterative fashion taking into account several sensor contributions (see Fig.37). Tab. 5 presents the early results of the proposed probabilistic method at combining activities detected using RGB and RGB-D cameras. Results are presented individually for each camera and for the proposed approach.

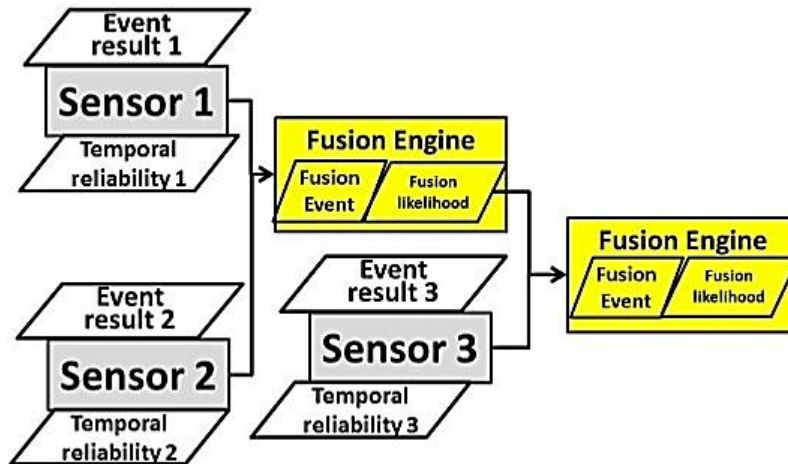


Figure 37. Probabilistic Approach Combination Scheme

The proposed fusion scheme performs better than the camera individual process in most of cases even in the presence of noise (see the RGB-D individual result of sensitivity for standing posture detection). The complete evaluation of the proposed approach is published in [43]. The developed probabilistic approach is now integrated into our previously developed framework for multi-sensor activity modeling. The new framework version allows experts to precisely define the sensors which will be used to detect each activity, or to automatically (and then probabilistic) combine multiple instances of (conflicting) activities detected by different sensors. The new framework proposal for multi-sensor activity detection has been published in [37].

Table 5. Performance of the proposed probabilistic approach on posture detection

Posture	Standing		Sitting	
Sensor	Precision	Sensitivity	Precision	Sensitivity
RGB	84.29	69.41	79.82	91.58
RGB-D	100.00	36.47	86.92	97.89
Fusion	82.35	91.30	91.04	95.31

6.18. Walking Speed Detection on a Treadmill using an RGB-D Camera

Participants: Baptiste Fosty, François Brémond.

keywords: RGB-D camera analysis, older people, serious games

Within the context of the Az@Game project, we have studied the potential of the RGB-D camera (Red Green Blue + Depth) for the control of a serious game dedicated to older people suffering from Alzheimer disease. Within this game, the patient is invited to perform some physical and cognitive tasks (walking on a treadmill at different speeds, performing gestures to control his/her character in the game, managing money) in order to assess the evolution of the disease, to stimulate them and improve their abilities. In this context, one of our goal is the computation of the walking speed of a person on a treadmill.

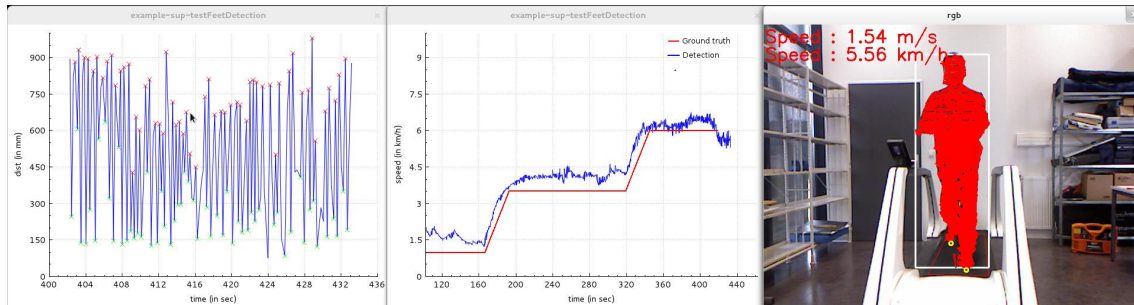


Figure 38. Walking speed computation on a treadmill. The left graph is the representation of the distance between the feet as a function of time. The middle graph is the representation of the speed of the person as a function of time. The right picture is the RGB-D camera view with the people detection and current speed. The two yellow circles show the positions of the detected feet.

The proposed solution is divided into three distinct steps :

- people detection and tracking using a background subtraction algorithm.
- feet detection : based on the cloud of 3D points of the person, and more particularly on the lower body part, the axis of each leg is computed. The corresponding foot is then defined as the projection of the lowest point of the leg on the leg axis (see Fig. 38, right picture).
- speed computation : computed from the successive positions of the feet, more precisely from the distances between the feet (see Fig. 38, left graph representing the distance between the feet). Each time this distance reaches a local maximum (corresponding to each step), the current speed is computed as the maximum over the time since the last step. The speed is then averaged with the previous computed speed to smooth the values (see Fig. 38, middle graph representing the speed in function of time).

Concerning the results, the first experimentation of the algorithm shows that, qualitatively, the computed walking speed is proportional to the real speed. Future work will focus on experimenting the proposed system on a larger scale (different people, location, etc.) in order to validate the approach. We will also focus on trying to detect arm gestures to have more control on the serious game.

About the Dem@Care project and the previous work on the activity recognition system developed to extract automatically and objectively evidences of early symptoms of Alzheimer's disease for older people, this contribution has been published at ASROB 2013, Workshop on Assistance and Service Robotics in a Human Environment (see [38]).

6.19. Serious Game for older adults with dementia

Participants: Minh Khue Phan Tran, François Brémond.

keywords: human-machine interaction, serious games, Alzheimer, activity recognition

Serious Games is carried out within the framework of the *Az@GAME project*. This project is to create games offering patient-oriented scenarios so as to measure their health progress, improve their physical fitness, stimulate their cognitive abilities, and help maintain their social skills. The main objective is to design a system interacting with older adults suffering from Alzheimer's or Alzheimer's related diseases. The three challenges in designing the system are:

- **perception's precision** : how does the system choose the "best moments" to interact with a patient ?
- **attractive-visualization** : how does the system make the patients comfortable?
- **ease of interaction** : how can it optimize the interaction with the patients ? In what way ?

The first prototype is under development. The system consists of two parts: Recognition and Interaction . Each component requires a 3D camera (Microsoft Kinect for the recognition component and Asus Xtion Pro Live Camera for the interaction component). The recognition part consists in observing the scene and deciding the best time to interact with a patient via the Asus camera. It uses the SUP framework. Afterwards, the interactive system tries to engage the patient via an interface and through Microsoft Kinect, the patient can interact with the interface using voice or gesture. The interface is designed with the Unity 3D game engine.

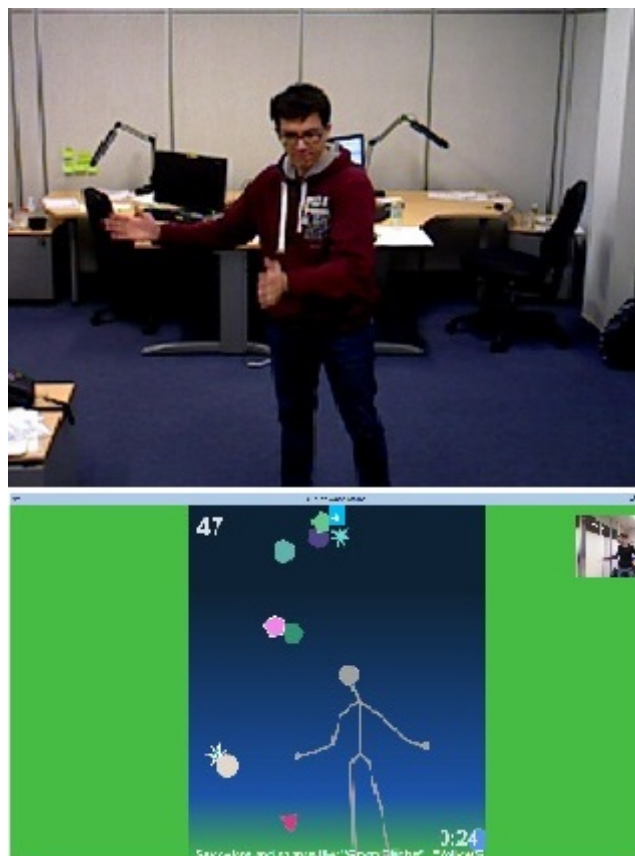


Figure 39. Example of a game and its gesture interface.

The first experiment will be conducted in the coffee area. The aim is to test the functionality of the system and measure its accuracy and effectiveness. The system will observe the scene and invite people who are

getting coffee or taking a break to play the game. Depending on the interaction with the person, the system will offer different scenarios. Videos will be recorded, with the consent of the subject, in order to evaluate the effectiveness of system. The recorded videos and meta-data provided by SUP will be evaluated to determine the accuracy of the system.

6.20. Unsupervised Activity Learning and Recognition

Participants: Serhan Cosar, Salma Zouaoui-Elloumi, François Brémond.

keywords: Unsupervised activity learning, hierarchical activity models, monitoring older people activities

The aim of this work is to monitor older people activities at hospital or at home environment in an unsupervised manner. We have extended the work in [81] that was initially based on user interface to label activities and proposed a new strongly unsupervised framework. It enables the discovery, modeling, and recognition of activities without user interaction. One advantage of this approach is that the framework learns individual behavioral patterns in unstructured scenes without restraining people to act based on a manually pre-defined model. The Figure 40-(a) presents the off-line learning steps of this framework. It takes as input a set of videos pre-processed to obtain trajectory information of people in the scene. Using the trajectory information (global position and pixel tracklets of body parts) of each person, zones of interest, where the person performs an activity, are learned. As in [81], we obtain three levels of zones using k -means clustering for different k values. The obtained zones are used to create different levels of events from the coarser to the finer ones. Based on the three levels of events, a hierarchical model of activities is learned to represent each action (Figure 40-(a)). For each new video, an on-line recognition process is performed by using the previously learned zones and models of activities (Figure 40-(b)).

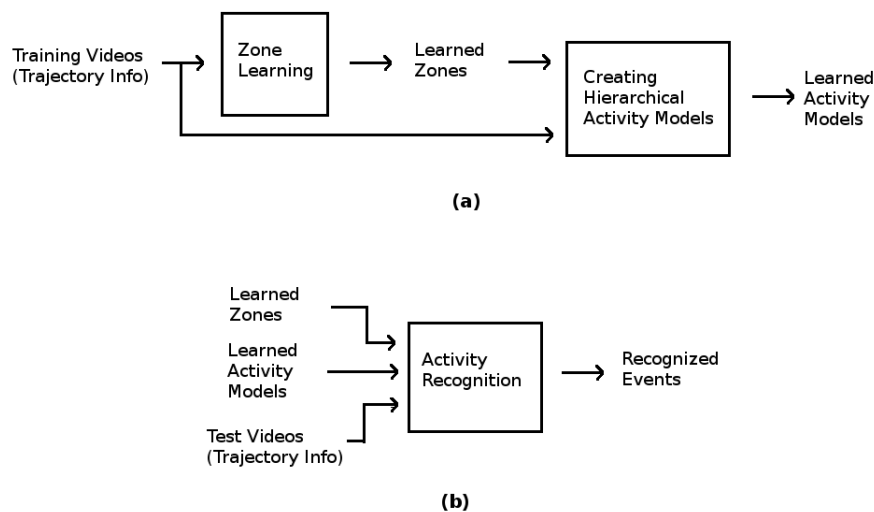


Figure 40. The flow diagram of the unsupervised activity recognition framework: (a) off-line learning phase and (b) on-line recognition phase

We have evaluated the performance of the unsupervised algorithm for RGB-D and 2D camera using 8 videos and 10 videos, respectively. Half of the videos are used for learning zones and models of activities. Videos are recorded in CHU Nice hospital while older people are visiting their doctors and include the following actions: “talking on the phone”, “preparing drugs”, “sitting at the table”, “preparing tea”, “looking at the bus map”, “watching TV” and “paying bill”. The trajectory information for 2D camera is obtained using the method

in [81]. For RGB-D camera, we have used the person detection algorithm in [79] and tracking algorithm in [33]. The results obtained for both cameras are presented in Table 6 and Table 7, respectively. We have used the following metrics to evaluate the framework: TP: True positive, FP: False positive, FN: False Negative, Sensitivity and Precision. According to the trajectory information, sometimes k -means clustering produces zones that are actually union of more than one zones. For such cases, we have combined the actions and presented as one single action.

Table 6. The recognition results obtained by using the 2D camera.

Actions	Instances	TP	FP	FN	Sensitivity(%)	Precision (%)
Paying bill	13	5	0	8	38.46	100
Preparing drugs	7	5	5	2	71.42	50
Looking at bus map+Watching TV	21	6	3	15	28.57	66.66
Sitting at the table	18	6	10	12	33.33	37.5
Talking on the phone	23	17	1	6	73.91	94.44
Preparing tea	23	11	3	12	47.82	78.57

Table 7. The recognition results obtained by using the RGB-D camera.

Actions	Instances	TP	FP	FN	Sensitivity(%)	Precision (%)
Paying bill + Watching TV	13	12	8	1	92.3	60
Preparing drugs	5	5	0	0	100	100
Looking at bus map	9	9	10	0	100	47.36
Sitting at the table	8	4	34	4	50	10.52
Talking on the phone	14	13	1	1	92.85	92.85
Preparing tea	16	9	5	7	56.25	64.28

As it can be seen in the tables, we obtain higher recognition rates by using the information coming from RGB-D camera.

Table 6 shows that for “talking on the phone” and “preparing drugs” actions occurring in two distant zones, using 2D camera gives high recognition rates (higher than 70 %). However, the actions “looking at bus map”, “watching TV” and “sitting at the table” are misclassified (low TP and high FP). Since the zones of these actions are very close to each other, the actions occurring in the borders are not well recognized. The reason of high FN is due to the problems in detection and tracking with 2D video cameras. The process of trajectory extraction described in [81] sometimes fails to track people. Because of the inadequate trajectory information, we have many FNs. Therefore, a better detection can considerably enhance the recognized actions.

By using the information coming from RGB-D camera, except for “sitting at the table” and “preparing tea” actions, we achieve high level of recognition rates (Table 7). However, similar to 2D camera, the recognition of “sitting at the table”, “paying bill” and “watching TV” actions fails because the learned zones in the scene are very close to each other. Hence, we have many false positives (FP) and false negatives (FN) for “sitting at the table” and “preparing tea” actions.

In the light of the preliminary experimental results, we can say that this unsupervised algorithm has a potential to be used for automated learning of behavioral patterns in unstructured scenes, for instance in home care environment for monitoring older people. Since the current framework does not require the user interaction to label activities, an evaluation process on big datasets will be easily performed. The proposed framework gives one action at each zone in an unsupervised way. We are currently focusing on refining the actions for each zone by using the pixel tracklets of the person's body parts. This will be achieved by performing clustering among activity models. As an example, the action of "sitting at the table" will be decomposed to "reading newspaper while sitting at the table" and "distributing cards while sitting at the table".

6.21. Extracting Statistical Information from Videos with Data Mining

Participants: Giuseppe Donatiello, Hervé Falciani, Duc Phu Chau, François Brémond.

keywords: video data mining, activity recognition, clustering techniques

Objective

Manual video observation is becoming less practical due to growing size of data. To tackle this problem, we have built a system to retrieve videos of interest thanks to an index based on activities recognized in an automated manner. We automatically detect activities in videos by combining data mining and computer vision to synthesize, analyze and extract valuable information from video data.

Approach

Our research introduces a new method for extracting statistical information from a video. Specifically, we focus on context modeling by developing an algorithm that automatically learns the zones in a scene where most activities occur by taking as input the trajectories of detected mobiles. Using K-means clustering, we define activity zones characterizing the scene dynamics, we can extract then people activities by relating their trajectories to the learned zones.

Results

To evaluate our system we have extended the OpenJUMP framework, an open source for Geographic Information System (GIS). The end user can have an overview of all activities of a large video, with the possibility of extracting and visualizing activities classified as usual or unusual. We have tested our approach on several videos recorded in subways in Turin (Italy) and Paris, as shown below, some examples of unusual activities (Figures 41, 42 and 43). The system has been showed in a live demonstration at RATP company in Paris for the European project Vanaheim (<http://www.vanaheim-project.eu/>).

6.22. SUP

Participants: Julien Gueytat, François Brémond.

keywords: SUP, Software, Video Processing

Presentation

SUP is a Scene Understanding Software Platform writtent in C++ designed for analyzing video content. (see Figure 44) SUP is splitting the workflow into several modules, such as acquisition, segmentation, etc., up to activity recognition. Each module has a specific interface, and different plugins implementing these interfaces can be used for each step of the video processing.

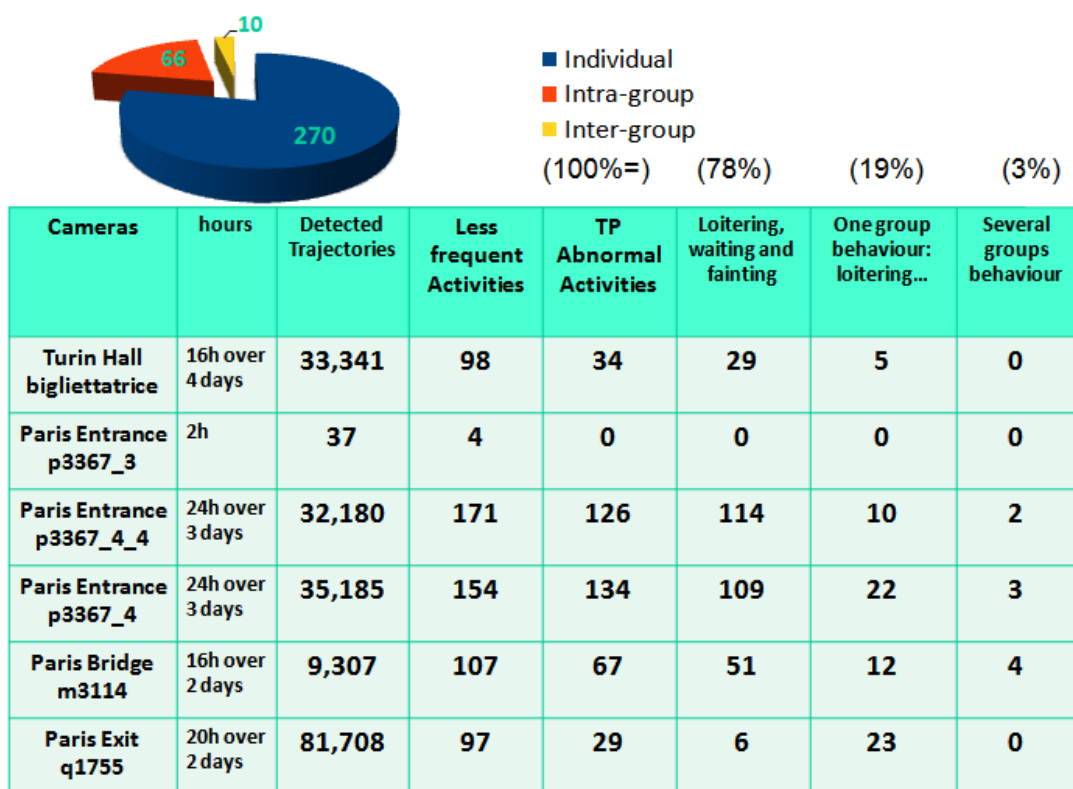


Figure 41. Results : three categories of discovered abnormal behaviors.

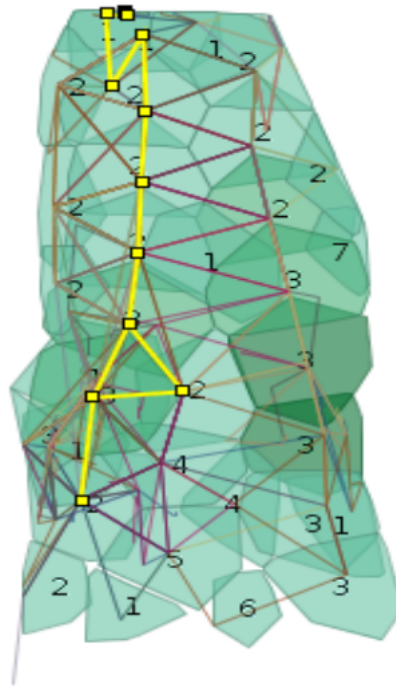


Figure 42. Trajectories and zones from camera M3114 in Paris (the yellow line is the trajectory selected by the user; the numbers represent the weight belonging to each segment : the higher, the more the people go through it).



Figure 43. Left : Person sitting. Middle : Person standing for a long time. Right : Unusual path.

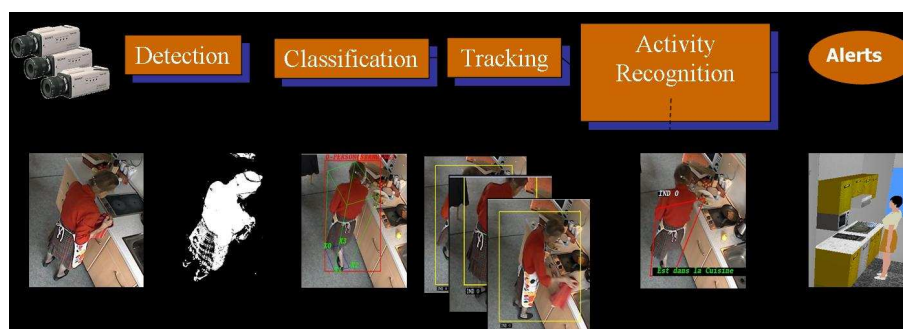


Figure 44. SUP workflow

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;
- languages : scenario description, empty 3D scene model description, video processing and understanding operator description;
- knowledge bases : scenario models and empty 3D scene models;
- algorithms of 2D & 3D visualisation of simulated temporal scenes and of real scene interpretation results;
- algorithms for evaluation of object detection, tracking and event recognition;
- learning techniques for event detection and human behaviour recognition;
- algorithms for image acquisition (RGB and RGBD cameras) and storage;
- algorithms for video processing supervision;
- algorithms for data mining and knowledge discovery;
- algorithms for image/video indexation and retrieval.

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)
- European partners: Lulea University of Technology, Dublin City University,...
- Industrial partners: Toyota, LinkCareServices, Digital Barriers

And new sites are coming: EHPAD Valrose, Institut Claude Pompidou, Delvalle and Biot.

Improvements

Our team focuses on developing a Scene Understanding Platform (SUP). This platform has been designed for analyzing video content. SUP is able to recognize events such as 'falling', 'walking' of a person. We can easily build new analyzing systems thanks to a set of algorithms also called plugins. The order of those plugins and their parameters can be changed at run time and the result visualized 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. All those advantages are available since we are working together with the Inria software developer team SED. Many Inria teams are pushing together to improve a common Inria development toolkit DTK. Our SUP framework is one of the DTK-like framework developed at Inria.

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

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 software developer team SED.

6.23. Model-Driven Engineering for Activity Recognition

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

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

The domain of video surveillance (VS) offers an ideal training ground for Software Engineering studies, because of the huge variability in both the surveillance tasks, the video analysis algorithms and the context. Such systems require run time adaptation of their architecture to react to external events and changing conditions in their context of execution.

The feature model formalism is widely used to capture variability, commonalities and configuration rules of software systems. We thus use feature modeling to represent the variability of both the specification and component views of video surveillance systems. We also include cross-tree constraints that formalize extra feature dependencies.

Based on this feature model, we can both select an initial system configuration at deployment time and dynamically adapt the current configuration at run time. This year we focused on runtime adaptation, from feature model to running components.

6.23.1. Configuration Adaptation at Run Time

In the continuation of our work on metrics on feature models, we have integrated a configuration selection algorithm in our feature model manager. Context changes or user interactions imply to dynamically reconfigure the model (selecting or deselecting features). Following *model at run time* techniques, we are able to determine the set of valid configurations to apply in a new execution context. Since only one configuration can be applied at a given time, the role of the selection algorithm is to select the "best" one.

To this end we enriched our feature representation with a set of *quality attributes* that correspond to a monotonic quantification of interesting aspects of the system quality. Examples are response time, accuracy, availability, performance, component switching time, etc. The configuration selection algorithm optimises a cost function, a linear weighted combination of the quality attributes. Thus we can rank the possible valid configurations and choose an optimal one. Our algorithm is a variant of the Best-First Search algorithm, a heuristic graph search technique. It starts with the set of valid configurations, which is a feature model where some features are *unselected*. Then it performs a systematic search in a graph where nodes are configurations and edges are selections or deselections of unselected features. The goal is to obtain a full configuration (one without unselected features) optimizing the cost function. The algorithm is parameterized with different strategies and associated heuristics with different optimality and efficiency characteristics.

Search strategies decide which node to visit next. We choose two well-known informed strategies that rely on heuristic functions as choice criteria. First we used a variant of the A* algorithm, BF*, but with a node-cost function instead of a path-cost one; it favors optimality over efficiency. Second, we implemented a Greedy Best-First Search (GBFS) strategy, where the next visited node is the best successor of the current one; it favors efficiency over optimality.

Computing the exact value of the cost function for a partial configuration is too expensive. We thus use heuristics to obtain a quick *estimate*. We have tested two sorts of heuristics. The simplest one, H_A , ignores the differences between the various sorts of groups (AND, OR, XOR) in the feature model and does not consider cross-tree constraints; it is fast but not very accurate. The second one, H_B , just drops the cross-tree constraints; it is thus more accurate, yet at a higher cost.

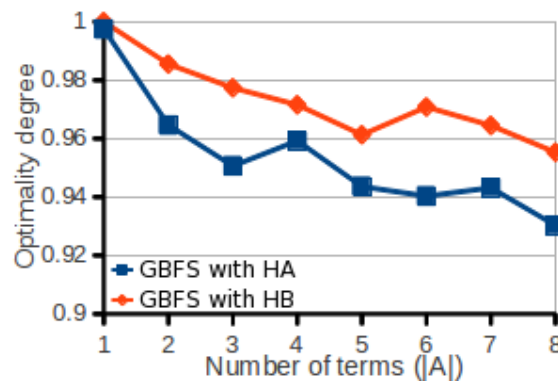


Figure 45. Optimality comparison of heuristics using GBFS strategy with respect to the number of attributes in the cost function

We have run experiments using large (randomly generated) feature models and compared completeness, optimality and efficiency of the selection algorithm, with different combinations of strategies and heuristics [42]. From our experiments, the GBFS strategy with heuristics H_B appears as the ideal option for real time systems that have to adapt in bounded time. This strategy ensures polynomial time complexity and guarantees optimality over 90%, which is good enough for our purpose (see figure 45). On the other hand, BF* strategy with heuristics H_B is ideal for offline decisions, such as defining the initial configuration of a system. Although this search strategy takes a significant time to compute, this is acceptable at deployment time to obtain the optimal configuration.

6.23.2. Run Time Components

When a configuration has been chosen, we must implement it with real components. We consider a configuration of a video-surveillance processing chain as a set of running components, that can be tuned, removed, added, or replaced dynamically, in response to events. To apply such configuration changes, we need a way to represent and dynamically manipulate the components themselves.

In a first attempt, we used an OSGi-like C++ framework (SOF, Service Oriented Framework). However, SOF did not really fulfill our needs. First, SOF is the only C++ OSGi framework that we could find and its C++ implementation deserves some improvement. Moreover, like OSGi, it relies on the notion of “service”, as can be found in Web applications, but which does not really fit our real time requirements. This notion of service is not our concern and makes programming more complicated than necessary.

Thus, we decided to define our own component module and to integrate it in a multi-threaded layer, easy to use for our end-users who are video system developers. Each component runs by default in its own

thread and communicates with other components through standardized communication channels. Our goal is to provide end-users with simple patterns to package their video codes into components. Thus we hide as much as possible the technical details such as threading synchronization, data exchange, and mechanisms for component management (replacement, tuning...) ensuring a continuous process.

We are currently setting up this framework on a simple video detection pipeline with OpenCV-based components. Then we shall integrate it within our Model at Run Time architecture.

6.24. Scenario Analysis Module

Participants: Annie Ressousche, Daniel Gaffé, Narjes Ghrairi, Sabine Moisan, Jean-Paul Rigault.

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.

Since last year, we relied on CLEM (see section 6.25) synchronous language to express the automata semantics of scenario models as Boolean equation systems. This year, we continue our research in this direction and we are studying a specific semantics of SAM language operators that translates any SAM program into Boolean equation system. Therefore, we will benefit from CLEM compilation technique to generate recognizer for each scenario model.

This year we focus on the definition of an execution machine able to transform asynchronous events coming from SUP or other devices into synchronous significant events feeding recognition engines generated by SAM. The execution machine can listen three types of asynchronous events: SUP events, Boolean sensors, sampled sensors and pulse train sensors. According to the sampling period of each sensor, the execution machine builds the significant events defining the synchronous logical instants which trigger the reaction of the scenario recognition engine. Thanks to the synchronous approach, scenario recognition engines are able to dynamically express the expected synchronous events of the next step; the execution machine takes into account of this information to filter relevant events. We perform several tests with real SUP data sets and the execution machine has a convincing behaviour (see [55]). To complement this work, we will integrate a notion of incompatible events which will make the execution machine more efficient and robust.

6.25. The Clem Workflow

Participants: Annie Ressousche, Daniel Gaffé, Joel Wanza Weloli.

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 9) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. This year, we mainly work on the implementation of new theoretical results concerning the foundation of LE semantics. We also design 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). In [69], we chosen an algebra which is a bilattice and we show that it is well suited to solve our problem. It is a new application of general bilattice theory [70]. This year, this approach has been improved, validated in CLEM compiler and published in [39]. Our compilation technique needs to represent Boolean equation systems with Binary Decision Diagrams (BDD) and we study and design a specific BDD library well suited to ours needs. From a practical point of view, we integrate new operators in LE language (sustain until, no emit, strong abort). We focus on automata extension which can consider now three types of transition: weak transition, strong transition and normal termination transition.

Second, 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 design the CLES simulator which interprets LEC. The actual version don't integrate the data part of the language and we plan to do this integration.

6.26. Multiple Services for Device Adaptive Platform for Scenario Recognition

Participants: Annie Ressousche, Daniel Gaffé, Mohammed Cherif Bergheul, 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 [88], [87], [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 behavioural model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

In [84], [83], 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 supply GALAXY automata editor to express Mealy machines and we extend AUTOM2CIRCUIT compiler to generate the internal code of WComp (C#). AUTOM2CIRCUIT is a tool developed by D. Gaffé since several years which compiles an explicit representation of automata into Boolean Mealy machine and generate a large and opened set of targets. This work is a preliminary study to integrate this generation of C# into CLEM.

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 base on RGBD sensors with very low rate of false alarms.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Collaborations

- Stars has a strong collaboration with the CobTek team (CHU Nice).
- G. Charpiat works with Yuliya Tarabalka (AYIN team) and with Björn Menze (Computer Vision Laboratory at ETH Zurich, Medical Vision group of CSAIL at MIT, and collaborator of Asclepios team) on the topic of shape growth/shrinkage enforcement for the segmentation of time series.
- G. Charpiat worked with former members from the Ariana team: Ahmed Gamal Eldin (LEAR team), Xavier Descombes (MORPHEME team) and Josiane Zerubia (AYIN team) on the topic of multiple object detection.
- A. Ressouche has a strong collaboration with the Rainbow team (I3S, UNS).

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. MOVEMENT

Program: ANR CSOSG

Project acronym: MOVEMENT

Project title: AutoMatic BiOmetric Verification and PersonEl 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.

8.2.1.2. 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.

8.2.2. Investment of future

8.2.2.1. 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 Genius

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.

8.2.3. Large Scale Inria Initiative

8.2.3.1. 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.

8.2.4. 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.

8.3. European Initiatives

8.3.1. FP7 Projects

8.3.1.1. CENTAUR

Title: Crowded ENvironments moniToring for Activity Understanding and Recognition

Type: POEPLÉ

Defi: Computer Vision

Instrument: Industry-Academia Partnerships and Pathway

Duration: January 2013 - December 2016

Coordinator: Honeywell (CZE)

Other partners: 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.

8.3.1.2. SUPPORT

Title: Security UPgrade for PORTs

Type: SECURITE

Defi: Port Security

Instrument: Industry-Academia Partnerships and Pathway

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.

8.3.1.3. Dem@Care

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

Type: ICT

Defi: Cognitive Systems and Robotics

Instrument: Industry-Academia Partnerships and Pathway

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.

8.3.1.4. VANAHEIM

Title: Autonomous Monitoring of Underground Transportation Environment

Type: ICT

Defi: Cognitive Systems and Robotics

Instrument: Industry-Academia Partnerships and Pathway

Duration: February 2010 - November 2013

Coordinator: Multitel (Belgium)

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

Inria contact: François Brémond

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

Abstract: The aim of this project is to study innovative surveillance components for the autonomous monitoring of multi-Sensory and networked Infrastructure such as underground transportation environment.

8.3.2. Collaborations in European Programs, except FP7

8.3.2.1. PANORAMA

Program: ENIAC

Project acronym: PANORAMA

Project title: Ultra Wide Context Aware Imaging

Duration: April 2012 - March 2015

Coordinator: Philips Healthcare (NL)

Other partners :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).

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

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.

8.4. International Initiatives

8.4.1. Inria International Partners

8.4.1.1. Collaborations with Asia

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

8.4.1.2. Collaboration with U.S.

Stars collaborates with the University of Southern California.

8.4.1.3. Collaboration with Europe

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

8.4.2. Participation In other International Programs

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

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

8.5. International Research Visitors

8.5.1. Visits of International Scientists

8.5.1.1. Internships

Narjes Ghrairi

Subject: Primitive Event Generation in an Activity Recognition Platform

Date: from Apr 2013 until Sep 2013

Institution: Ecole Nationale d'Ingénieurs de Tunis (Tunisia)

Mohammed Cherif Bergheul

Subject: Adaptive composition and formal verification of software in ubiquitous computing. Application to ambient health care systems.

Date: from Apr 2013 until Sep 2013

Institution: Ecole Polytech'Nice Cairo (Egypt)

Kartick Subramanian

Subject: People Tracking

Date: from Mar 2013 until Aug 2013

Institution: Nanyang Technological University, Singapore

Augustin Caverzasi

Subject: Trajectory fusion of multi-camera RGB-Depth tracking in partial overlapped scenes.

Date: from Aug 2013 until Dec 2013

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

Stefanus Candra

Subject: Evaluation of activity recognition system using RGB-Depth camera (e.g. Kinect)

Date: from Aug 2013 until Dec 2013

Institution: University of California, Berkeley CA, Usa

Sahil Dhawan

Subject: Assment of people detection using RGB-Depth sensors (e.g. Kinect), for apathetic patients to improve activity recognition systems.

Date: from Jan 2013 until Jul 2013

Institution: Birla institute of technology and Science, Pilani , India

Marco San Biagio

Subject: People detection using the Brownian descriptor.

Date: from Apr 2013 until Sep 2013

Institution: Italian Inst. of Tech. of Genova

Michal Koperski

Subject: 3D Trajectories for Action Recognition Using Depth Sensors

Date: from Apr 2013 until Dec 2013

Institution: Wroclaw University of Technology

9. Dissemination

9.1. Scientific Animation

- François Brémond was reviewer for the conferences : CVPR'13-14, ICCV'13, AVSS'13, ICVS'13, ICDP'13, ICRA'13, ACM MM'13;
- François Brémond was handling editor of the international journal "Machine Vision and Application";
- François Brémond was reviewer for the journal "IEEE Computing Now";
- François Brémond was program committee member of the conferences and workshops: IEEE Workshop on Applications of Computer Vision (WACV 2013-14), ICME AAMS-PS 2013, RIVF 2013, ASROB-2013, ICCV'13 Workshop on Action Recognition (THUMOS), on Re-Identification (Re-Id 2013) and on Object Tracking VOT'2013, LACNEM-13, ACM MIIRH 2013;
- François Brémond was speaker in two thematic schools: S5 Second Spring Short School on Surveillance in Modena Italy, May 7-9 2013 and Dem@Care School on Ambient Assisted Living (DemAAL 2013), Chania, Crete, Greece (16-20 September 2013);
- François Brémond was organizer of the SPHERE workshop on Ambient Assisted Living, at Inria Sophia Antipolis, 9-10 October 2013 and a PANORAMA special session, part of VISAPP Lisbon, Portugal, 5-8 January 2014;
- François Brémond was an expert in review of research projects for France Alzheimer Association (25th May 2013), for NTU Singapore (19th August 2013) and for INSA de Lyon (4th November 201);
- 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 (ERNICIP) - Video Analytics and surveillance Group, at European Commission's Joint Research Centre in Ispra, Italy, 29 April 2013, 28-29 October, 2013;
- François Brémond was area chair of the IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS'13, Krakow, Poland on August 27-30, 2013;
- François Brémond gave invited talks at:
 - the Symposium of World Congress of Gerontology and Geriatrics (IAGG) 2013, 23-27 June, Coex, Seoul, Korea (Title: "ICT in Dementia: from assessment to stimulation.);
 - the Conference "Les Echos sur la dépendance, quelles réponses pour faire reculer la dépendance ? " in March 2013;
 - the IX Journée de la Fédération CMRR in Nice on the 29 March 2013;
 - the Journée pédo psychiatrie, Nice on the 11th October 2013 (invited by F. Azkenazy);
 - University of Central Florida, 14-15 October 2013 (invited by M. Shah);
- Guillaume Charpiat reviewed for the conferences CVPR, ICCV and Grets;
- Guillaume Charpiat reviewed for the journals CVIU (Computer Vision and Image Understanding), SIAM Journal on Imaging Sciences (SIIMS), RIA (Revue d'intelligence artificielle);
- Guillaume Charpiat gave invited talks :
 - in the TAO team (8 October 2013) about lossless image compression,
 - and in the GALEN team (9 October 2013) about energies minimizable with graph cuts.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : Annie Ressousche, Critical Software Verification- Introduction to SCADE, 5h, niveau (L3), Mines Paris Tech, FR

Master : Annie Ressousche, Critical Software Verification and application to WComp Middleware , 10h, niveau (M2), Polytech'Nice School of Nice University, FR

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

9.2.2. Supervision

PhD: Rim Romdhane, Event Recognition in Video Scenes with Uncertain Knowledge, 30th September 2013, François Brémond and Monique Thonnat;

PhD: Malik Souled : Suivi d'Individu à travers un Réseau de Caméras Vidéo, 20th December 2013, François Brémond ;

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

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

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

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

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

9.2.3. Juries

- François Brémond was jury member of the following PhD theses:
 - Boris Meden, Lab Vision and Content Engineering, CEA Saclay – nano innov, 15 January 2013;
 - Wenjuan Gong, PhD VIVA at Computer Vision Center in Barcelona, 10th May 2013;
 - Miriam Redi, EURECOM, 29th May 2013;
 - Lyazid Sabri, Université Paris Est Créteil (UPEC), 1 July 2013;
 - Francis Martinez, Vision VLSI/System ISIR/UPMC & CNRS UMR, Université Pierre et Marie Curie (UPMC), Jussieu, Paris, 9 July 2013;
 - Hajer Fradi PhD defense, EURECOM, 28th January 2014.

9.3. Popularization

- François Brémond participated in ERCIM News 95, Special theme: "Image Understanding", Sept 2013;
- François Brémond was interviewed by magazine 01net/itv Jul 2013;
- François Brémond was interviewed on Autisme by Monaco Info TV, September 2013;
- François Brémond gave a talk at Forum de la Recherche, Nice, 22 November 2013;
- Guillaume Charpiat is part of the MASTIC commitee;
- Guillaume Charpiat wrote with Yuliya Tarabalka a popularization article about wild fire mapping from satellite images in the magazine Earthzine : <http://www.earthzine.org/2013/09/16/exploiting-temporal-coherence-for-fire-mapping-from-modis-spaceborne-observations/> [53].

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^o 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] G. CHARPIAT, P. MAUREL, J.-P. PONS, R. KERIVEN, O. FAUGERAS. *Generalized Gradients: Priors on Minimization Flows*, in "International Journal of Computer Vision", 2007
- [7] N. CHLEQ, F. BRÉMOND, M. THONNAT. , *Advanced Video-based Surveillance Systems*, Kluwer A.P. , Hangham, MA, USA, November 1998, pp. 108-118
- [8] F. CUPILLARD, F. BRÉMOND, M. THONNAT. , *Tracking Group of People for Video Surveillance*, Video-Based Surveillance Systems, Kluwer Academic Publishers, 2002, vol. The Kluwer International Series in Computer Vision and Distributed Processing, pp. 89-100
- [9] F. FUSIER, V. VALENTIN, F. BREMOND, M. THONNAT, M. BORG, D. THIRDE, J. FERRYMAN. *Video Understanding for Complex Activity Recognition*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 167-188
- [10] B. GEORIS, F. BREMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 189-205
- [11] C. LIU, P. CHUNG, Y. CHUNG, M. THONNAT. *Understanding of Human Behaviors from Videos in Nursing Care Monitoring Systems*, in "Journal of High Speed Networks", 2007, vol. 16, pp. 91-103
- [12] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", December 2004, vol. 16, n^o 1, pp. 33-40
- [13] V. MARTIN, J.-M. TRAVERE, F. BREMOND, V. MONCADA, G. DUNAND. *Thermal Event Recognition Applied to Protection of Tokamak Plasma-Facing Components*, in "IEEE Transactions on Instrumentation and Measurement", Apr 2010, vol. 59, n^o 5, pp. 1182-1191, <http://hal.inria.fr/inria-00499599>
- [14] S. MOISAN. *Knowledge Representation for Program Reuse*, in "European Conference on Artificial Intelligence (ECAI)", Lyon, France, July 2002, pp. 240-244
- [15] S. MOISAN. , *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*, Université de Nice-Sophia Antipolis, April 1998, Habilitation à diriger les recherches

- [16] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Blocks, a Component Framework with Checking Facilities for Knowledge-Based Systems*, in "Informatica, Special Issue on Component Based Software Development", November 2001, vol. 25, n^o 4, pp. 501-507
- [17] J. PATINO, H. BENHADDA, E. CORVEE, F. BREMOND, M. THONNAT. *Video-Data Modelling and Discovery*, in "4th IET International Conference on Visual Information Engineering VIE 2007", London, UK, 25th - 27th July 2007
- [18] J. PATINO, E. CORVEE, F. BREMOND, M. THONNAT. *Management of Large Video Recordings*, in "2nd International Conference on Ambient Intelligence Developments AmI.d 2007", Sophia Antipolis, France, 17th - 19th September 2007
- [19] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, in "Software Engineering Research, Management and Applications", R. LEE (editor), Studies in Computational Intelligence, Springer, 2008, vol. 150, pp. 157-171, selected as one of the 17 best papers of SERA'08 conference
- [20] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Revue des sciences et technologies de l'information, série Théorie et Science Informatique", June 2011, vol. 4, n^o 30, pp. 441-471, <http://hal.inria.fr/inria-00524499/en>
- [21] M. THONNAT, S. MOISAN. *What Can Program Supervision Do for Software Re-use?*, in "IEE Proceedings - Software Special Issue on Knowledge Modelling for Software Components Reuse", 2000, vol. 147, n^o 5
- [22] M. THONNAT. , *Vers une vision cognitive: mise en oeuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images.*, Université de Nice-Sophia Antipolis, October 2003, Habilitation à diriger les recherches
- [23] M. THONNAT. *Special issue on Intelligent Vision Systems*, in "Computer Vision and Image Understanding", May 2010, vol. 114, n^o 5, pp. 501-502, <http://hal.inria.fr/inria-00502843>
- [24] A. TOSHEV, F. BRÉMOND, M. THONNAT. *An A priori-based Method for Frequent Composite Event Discovery in Videos*, in "Proceedings of 2006 IEEE International Conference on Computer Vision Systems", New York USA, January 2006
- [25] V. VU, F. BRÉMOND, M. THONNAT. *Temporal Constraints for Video Interpretation*, in "Proc of the 15th European Conference on Artificial Intelligence", Lyon, France, 2002
- [26] V. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)", 9-15 September 2003
- [27] N. ZOUBA, F. BREMOND, A. ANFOSSO, M. THONNAT, E. PASCUAL, O. GUERIN. *Monitoring elderly activities at home*, in "Gerontechnology", May 2010, vol. 9, n^o 2, <http://hal.inria.fr/inria-00504703>

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [28] M. SOUDED. , *Détection, Suivi et Ré-identification de Personnes à Travers un Réseau de Caméras Vidéo*, Institut National de Recherche en Informatique et en Automatique - Inria, December 2013, <http://hal.inria.fr/tel-00913072>

Articles in Non Peer-Reviewed Journals

- [29] S. BAK, F. BREMOND. *Person Re-identification*, in "ERCIM News", October 2013, n^o 95, <http://hal.inria.fr/hal-00907397>

International Conferences with Proceedings

- [30] 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, <http://hal.inria.fr/hal-00907455>
- [31] 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, <http://hal.inria.fr/hal-00905588>
- [32] 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, IEEE, April 2013, pp. 1-7 [DOI : 10.1109/FG.2013.6553699], <http://hal.inria.fr/hal-00806321>
- [33] 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, <http://hal.inria.fr/hal-00846920>
- [34] D. P. CHAU, M. THONNAT, F. BREMOND. *Automatic Parameter Adaptation for Multi-object Tracking*, in "International Conference on Computer Vision Systems (ICVS)", St Petersburg, Russian Federation, Springer, July 2013, <http://hal.inria.fr/hal-00821669>
- [35] S. COSAR, M. CETIN. *A Sparsity-Driven Approach to Multi-camera Tracking in Visual Sensor Networks*, in "Workshop on Activity Monitoring by Multiple Distributed Sensing (AMMDS) in conjunction with 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance", Krakow, Poland, August 2013, <http://hal.inria.fr/hal-00919946>
- [36] 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], <http://hal.inria.fr/hal-00875972>
- [37] C. F. CRISPIM-JUNIOR, Q. MA, B. FOSTY, R. ROMDHANE, F. BREMOND, M. THONNAT. *Combining Multiple Sensors for Event Recognition of Older People*, in "MIIRH - 1st ACM MM Workshop on Multimedia Indexing and information Retrieval for Healthcare - 2013", Barcelona, Spain, ACM, October 2013, pp. 15-22, MIRRH, held in conjunction with ACM MM 2013 [DOI : 10.1145/2505323.2505329], <http://hal.inria.fr/hal-00907033>
- [38] B. FOSTY, C. F. CRISPIM-JUNIOR, J. BADIE, F. BREMOND, M. THONNAT. *Event Recognition System for Older People Monitoring Using an RGB-D Camera*, in "ASROB - Workshop on Assistance and Service Robotics in a Human Environment", Tokyo, Japan, November 2013, <http://hal.inria.fr/hal-00904002>

- [39] D. GAFFÉ, A. RESSOUCHE. *Algebraic Framework for Synchronous Language Semantics*, in "Theoretical Aspects of Software Engineering", Birmingham, United Kingdom, L. FERARIU, A. PATELLI (editors), IEEE Computer Society, July 2013, pp. 51-58, <http://hal.inria.fr/hal-00841559>
- [40] A. GAMAL ELDIN, G. CHARPIAT, X. DESCOMBES, J. ZERUBIA. *An efficient optimizer for simple point process models*, in "SPIE, Computational Imaging XI", Burlingame, California, United States, C. A. BOUMAN, I. POLLAK, P. J. WOLFE (editors), SPIE Proceedings, SPIE, February 2013, vol. 8657 [DOI : 10.1117/12.2009238], <http://hal.inria.fr/hal-00801448>
- [41] R. KUMAR, M. THONNAT, G. CHARPIAT. *Hierarchical Representation of Videos with Spatio-Temporal Fibers*, in "IEEE Winter Conference on Applications of Computer Vision", Colorado, United States, March 2014, <http://hal.inria.fr/hal-00911012>
- [42] S. LUIS EMILIANO, S. MOISAN, J.-P. RIGAULT. *Metrics on Feature Models to Optimize Configuration Adaptation at Run Time*, in "International Conference on Software Engineering (ICSE'2013) - CMSBSE - Combining Modelling and Search-Based Software Engineering - 2013", San Francisco, United States, IEEE, May 2013, pp. 39 - 44 [DOI : 10.1109/CMSBSE.2013.6604435], <http://hal.inria.fr/hal-00877387>
- [43] Q. MA, B. FOSTY, C. F. CRISPIM-JUNIOR, F. BREMOND. *FUSION FRAMEWORK FOR VIDEO EVENT RECOGNITION*, in "The 10th IASTED International Conference on Signal Processing, Pattern Recognition and Applications", Innsbruck, Austria, The International Association of Science and Technology for Development (IASTED), February 2013, <http://hal.inria.fr/hal-00784725>
- [44] 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 2014)", Lisbon, Portugal, January 2014, <http://hal.inria.fr/hal-00909566>
- [45] 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, <http://hal.inria.fr/hal-00909124>
- [46] M. SOUDED, F. BREMOND. *Optimized Cascade of Classifiers for People Detection Using Covariance Features*, in "International Conference on Computer Vision Theory and Applications (VISAPP)", Barcelona, Spain, February 2013, <http://hal.inria.fr/hal-00794369>
- [47] Y. TARABALKA, G. CHARPIAT, L. BRUCKER, B. MENZE. *Enforcing Monotonous Shape Growth or Shrinkage in Video Segmentation*, in "BMVC - British Machine Vision Conference", Bristol, United Kingdom, September 2013, <http://hal.inria.fr/hal-00856634>
- [48] Y. TARABALKA, G. CHARPIAT. *A Graph-Cut-Based Method for Spatio-Temporal Segmentation of Fire from Satellite Observations*, in "IEEE IGARSS - International Geoscience and Remote Sensing Symposium", Melbourne, Australia, July 2013, <http://hal.inria.fr/hal-00845691>
- [49] 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 - 2014", Lisbon, Portugal, S. BATTIATO (editor), SCITEPRESS Digital Library, January 2014, <http://hal.inria.fr/hal-00878580>

Conferences without Proceedings

- [50] D. P. CHAU, F. BREMOND, M. THONNAT. *Object Tracking in Videos: Approaches and Issues*, in "The International Workshop "Rencontres UNS-UD" (RUNSUD)", Danang, Viet Nam, April 2013, <http://hal.inria.fr/hal-00815046>

Scientific Books (or Scientific Book chapters)

- [51] S. BAK, F. BREMOND. *Re-identification by Covariance Descriptors*, in "Person Re-Identification", S. GONG, M. CRISTANI, Y. SHUICHENG, C. C. LOY (editors), Advances in Computer Vision and Pattern Recognition, Springer, December 2013, <http://hal.inria.fr/hal-00907335>

Research Reports

- [52] G. CHARPIAT, G. NARDI, G. PEYRÉ, F.-X. VIALARD. , *Finsler Steepest Descent with Applications to Piecewise-regular Curve Evolution*, July 2013, <http://hal.inria.fr/hal-00849885>

Scientific Popularization

- [53] Y. TARABALKA, G. CHARPIAT. *Exploiting Temporal Coherence for Fire Mapping from MODIS Spaceborne Observations*, in "Earthzine", September 2013, <http://hal.inria.fr/hal-00915245>

Other Publications

- [54] 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, <http://hal.inria.fr/hal-00879734>
- [55] N. GHRAIRI. , *Génération des événements primitifs dans une plateforme de reconnaissance d'activités*, Ecole nationale d'ingénieurs de SousseSousse(Tunisie), October 2013, 66 p. , <http://hal.inria.fr/hal-00871836>

References in notes

- [56] M. ACHER, P. COLLET, F. FLEUREY, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Modeling Context and Dynamic Adaptations with Feature Models*, in "Models@run.time Workshop", Denver, CO, USA, October 2009, <http://hal.inria.fr/hal-00419990/en>
- [57] M. ACHER, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Tackling High Variability in Video Surveillance Systems through a Model Transformation Approach*, in "ICSE'2009 - MISE Workshop", Vancouver, Canada, May 2009, <http://hal.inria.fr/hal-00415770/en>
- [58] H. BEN SHITRIT, J. BERCLAZ, F. FLEURET, P. FUA. *Tracking multiple people under global appearance constraints*, in "IEEE International Conference on Computer Vision (ICCV)", 2011, pp. 137-144
- [59] B. BENFOLD, I. REID. *Stable multi-target tracking in real-time surveillance video*, in "CVPR '11", 2011 [DOI : 10.1109/CVPR.2011.5995667], <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5995667>
- [60] J. BERCLAZ, E. TURETKEN, F. FLEURET, P. FUA. *Multiple Object Tracking using K-Shortest Paths Optimization*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2011

- [61] P. BILINSKI, F. BREMOND. *Contextual Statistics of Space-Time Ordered Features for Human Action Recognition*, in "9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)", Beijing, Chine, September 2012, <http://hal.inria.fr/hal-00718293>
- [62] P. BILINSKI, F. BREMOND. *Statistics of Pairwise Co-occurring Local Spatio-Temporal Features for Human Action Recognition*, in "4th International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR), in conjunction with 12th European Conference on Computer Vision (ECCV)", Florence, Italie, A. FUSIELLO, V. MURINO, R. CUCCHIARA (editors), Lecture Notes in Computer Science, Springer, October 2012, vol. 7583, pp. 311-320, This work was supported by the Region Provence-Alpes- Côte d'Azur [DOI : 10.1007/978-3-642-33863-2_31], <http://hal.inria.fr/hal-00760963>
- [63] D. P. CHAU, F. BREMOND, M. THONNAT. *A multi-feature tracking algorithm enabling adaptation to context variations*, in "The International Conference on Imaging for Crime Detection and Prevention (ICDP)", London, Royaume-Uni, November 2011, <http://hal.inria.fr/inria-00632245/en/>
- [64] D. P. CHAU, F. BREMOND, M. THONNAT, E. CORVEE. *Robust Mobile Object Tracking Based on Multiple Feature Similarity and Trajectory Filtering*, in "The International Conference on Computer Vision Theory and Applications (VISAPP)", Algarve, Portugal, March 2011, This work is supported by the PACA region, The General Council of Alpes Maritimes province, France as well as The ViCoMo, Vanaheim, Video-Id, Cofriend and Support projects, <http://hal.inria.fr/inria-00599734/en/>
- [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 "Proceeding CAV", Copenhagen, Danmark, E. BRINKSMA, K. G. LARSEN (editors), LNCS, Springer-Verlag, July 2002, n° 2404, pp. 359-364, <http://nusmv.fbk.eu/NuSMV/papers/cav02/ps/cav02.ps>
- [66] 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
- [67] M. FARENZENA, L. BAZZANI, A. PERINA, V. MURINO, M. CRISTANI. *Person re-identification by symmetry-driven accumulation of local features*, in "Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on", 2010, pp. 2360–2367
- [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. , *Algebras and Synchronous Language Semantics*, Inria, November 2012, n° RR-8138, 107 p. , <http://hal.inria.fr/hal-00752976>
- [70] M. GINSBERG. *Multivalued Logics: A Uniform Approach to Inference in Artificial Intelligence*, in "Computational Intelligence", 1988, vol. 4, pp. 265–316
- [71] M. GRUNDMANN, V. KWATRA, M. HAN, I. ESSA. *Efficient hierarchical graph-based video segmentation*, in "CVPR '10", http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5539893
- [72] J. F. HENRIQUES, R. CASEIRO, J. BATISTA. *Globally optimal solution to multi-object tracking with merged measurements*, in "IEEE International Conference on Computer Vision", 2011

- [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. KUO, C. HUANG, R. NEVATIA. , *Multi-target tracking by online learned discriminative appearance models*, 2010, In CVPR
- [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*, 2009, The International Conference on Computer Vision and Pattern Recognition (CVPR)
- [77] A. MILAN, K. SCHINDLER, S. ROTH. *Detection-and Trajectory-Level Exclusion in Multiple Object Tracking*, in "CVPR '13", <http://www.gris.tu-darmstadt.de/~aandriye/files/cvpr2013/cvpr2013-poster.pdf>
- [78] S. MOISAN, J.-P. RIGAUT, 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 [DOI : 10.1109/ISSPA.2012.6310538]
- [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] G.-T. PUSIOL. , *Discovery of human activities in video*, Institut National de Recherche en Informatique et en Automatique (Inria), May 2012
- [82] A. RESSOUCHE, D. GAFFÉ, V. ROY. , *Modular Compilation of a Synchronous Language*, Inria, 01 2008, n^o 6424, <http://hal.inria.fr/inria-00213472>
- [83] A. RESSOUCHE, J.-Y. TIGLI, O. CARRILLO. , *Composition and Formal Validation in Reactive Adaptive Middleware*, Inria, February 2011, n^o RR-7541, <http://hal.inria.fr/inria-00565860/en>
- [84] 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/>
- [85] L. M. ROCHA, S. MOISAN, J.-P. RIGAUT, 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>
- [86] 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>

- [87] J.-Y. TIGLI, S. LAVIROTTE, G. REY, V. HOURDIN, D. CHEUNG, E. CALLEGARI, M. RIVEILL. *WComp middleware for ubiquitous computing: Aspects and composite event-based Web services*, in "Annals of Telecommunications", 2009, vol. 64, n^o 3-4, ISSN 0003-4347 (Print) ISSN 1958-9395 (Online)
- [88] J.-Y. TIGLI, S. LAVIROTTE, G. REY, V. HOURDIN, M. RIVEILL. *Lightweight Service Oriented Architecture for Pervasive Computing*, in "IJCSI International Journal of Computer Science Issues", 2009, vol. 4, n^o 1, ISSN (Online): 1694-0784 ISSN (Print): 1694-0814
- [89] O. TUZEL, F. PORIKLI, P. MEER. *Human detection via classification on riemannian manifolds*, in "In Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition", 2007, pp. 1–8
- [90] 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>
- [91] 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>
- [92] J. XING, H. AI, S. LAO. , *Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses*, 2009, In CVPR
- [93] J. YAO, J.-M. ODOBEZ. *Fast Human Detection from Videos Using Covariance Features*, in "The Eighth International Workshop on Visual Surveillance - VS2008", 2008
- [94] S. ZAIDENBERG, B. BOULAY, F. BREMOND. *A generic framework for video understanding applied to group behavior recognition*, in "9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2012)", Beijing, Chine, Advanced Video and Signal Based Surveillance, IEEE Conference on, IEEE Computer Society, September 2012, pp. 136 -142, <http://hal.inria.fr/hal-00702179>