



Activity Report 2015

# Project-Team STARS

Spatio-Temporal Activity Recognition Systems

RESEARCH CENTER  
Sophia Antipolis - Méditerranée

THEME  
Vision, perception and multimedia  
interpretation



## Table of contents

<b>1. Members</b> .....	<b>1</b>
<b>2. Overall Objectives</b> .....	<b>3</b>
2.1.1. Research Themes	3
2.1.2. International and Industrial Cooperation	5
<b>3. Research Program</b> .....	<b>5</b>
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
3.3. Semantic Activity Recognition	6
3.3.1. Introduction	6
3.3.2. High Level Understanding	7
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
<b>4. Application Domains</b> .....	<b>10</b>
4.1. Introduction	10
4.2. Video Analytics	10
4.3. Healthcare Monitoring	11
4.3.1. Research	11
4.3.2. Ethical and Acceptability Issues	11
<b>5. Highlights of the Year</b> .....	<b>11</b>
5.1.1. Perception for Activity Recognition	12
5.1.2. Semantic Activity Recognition	12
<b>6. New Software and Platforms</b> .....	<b>12</b>
6.1. CLEM	12
6.2. EGMM-BGS	13
6.3. MTS	13
6.4. Person Manual Tracking in a Static Camera Network (PMT-SCN)	13
6.5. PrintFoot Tracker	13
6.6. Proof of Concept Néosensys (Poc-NS)	14
6.7. SUP	14
6.8. VISEVAL	14
6.9. py_ad	14
6.10. py_ar	15
6.11. py_sup_reader	15
6.12. py_tra3d	15
6.13. sup_ad	15
<b>7. New Results</b> .....	<b>16</b>
7.1. Introduction	16
7.1.1. Perception for Activity Recognition	16
7.1.2. Semantic Activity Recognition	16
7.1.3. Software Engineering for Activity Recognition	16
7.2. Pedestrian Detection using Convolutional Neural Networks	17
7.3. Head Detection for Eye Tracking Application	18

7.4.	Minimizing Hallucination in Histogram of Oriented Gradients	19
7.5.	Hybrid Approaches for Gender estimation	21
7.6.	Automated Healthcare: Facial-expression-analysis for Alzheimer's Patients in Musical Mnemotherapy	23
7.7.	Robust Global Tracker Based on an Online Estimation of Tracklet Descriptor Reliability	24
7.8.	Optimizing People Tracking for a Video-camera Network	27
7.9.	Multi-camera Multi-object Tracking and Trajectory Fusion	28
7.10.	Person Re-identification in Real-World Surveillance Systems	29
7.11.	Human Action Recognition in Videos	30
7.12.	Evaluation of Event Recognition without Using Ground Truth	32
7.13.	Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition	33
7.14.	Semi-supervised Activity Recognition Using High-order Temporal-composite Patterns of Visual Concepts	35
7.15.	From Activity Recognition to the Assessment of Seniors' Autonomy	37
7.16.	Serious Games Interfaces Using an RGB-D Camera : Results and Perspectives	38
7.17.	Assistance for Older Adults in Serious Game Using an Interactive System	39
7.18.	Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition	41
7.19.	Run-time Adaptation of Video Systems	42
7.20.	Scenario Description Language	44
7.21.	Scenario Recognition	44
7.22.	The Clem Workflow	45
7.23.	Safe Composition in WCOMP Middleware for Internet of Things	45
7.24.	Design of UHD Panoramic Video Camera	47
7.25.	Brick & Mortar Cookies	48
7.26.	Monitoring Older People Experiments	48
<b>8.</b>	<b>Bilateral Contracts and Grants with Industry</b>	<b>49</b>
<b>9.</b>	<b>Partnerships and Cooperations</b>	<b>50</b>
9.1.	National Initiatives	50
9.1.1.	ANR	50
9.1.1.1.	MOVEMENT	50
9.1.1.2.	SafEE	50
9.1.2.	Investment of Future	50
9.1.3.	FUI	51
9.2.	European Initiatives	51
9.2.1.1.	CENTAUR	51
9.2.1.2.	PANORAMA	52
9.2.1.3.	DEM@CARE	52
9.3.	International Initiatives	53
9.3.1.1.	Informal International Partners	53
9.3.1.1.1.	Collaborations with Asia:	53
9.3.1.1.2.	Collaboration with U.S.A.:	53
9.3.1.1.3.	Collaboration with Europe:	53
9.3.1.2.	Other IIL projects	53
9.4.	International Research Visitors	53
9.4.1.	Visits of International Scientists	53
9.4.2.	Visits to International Teams	54
<b>10.</b>	<b>Dissemination</b>	<b>54</b>
10.1.	Promoting Scientific Activities	54
10.1.1.	Scientific events organisation	54
10.1.1.1.	General chair, scientific chair	54
10.1.1.2.	Member of the organizing committee	54

---

10.1.2. Scientific events selection	54
10.1.2.1. Member of the conference program committees	54
10.1.2.2. Reviewer	54
10.1.3. Journal	54
10.1.3.1. Member of the editorial boards	54
10.1.3.2. Reviewer - Reviewing activities	54
10.1.4. Invited talks	54
10.1.5. Scientific expertise	55
10.2. Teaching - Supervision - Juries	55
10.2.1. Teaching	55
10.2.2. Supervision	55
10.2.3. Juries	56
10.3. Popularization	56
<b>11. Bibliography</b> .....	<b>56</b>



# Project-Team STARS

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

## Keywords:

### Computer Science and Digital Science:

- 3.2.1. - Knowledge bases
- 3.3.2. - Data mining
- 3.4.1. - Supervised learning
- 3.4.2. - Unsupervised learning
- 3.4.5. - Bayesian methods
- 5.3.2. - Sparse modeling and image representation
- 5.3.3. - Pattern recognition
- 5.4.1. - Object recognition
- 5.4.2. - Activity recognition
- 8.1. - Knowledge
- 8.2. - Machine learning
- 8.3. - Signal analysis

### Other Research Topics and Application Domains:

- 1.3.2. - Cognitive science
- 2.1. - Well being
- 7.1.1. - Pedestrian traffic and crowds
- 8.1. - Smart building/home
- 8.4. - Security and personal assistance

## 1. Members

### Research Scientists

François Brémond [Team leader, Inria, Senior Researcher, HDR]  
Sabine Moisan [Inria, Researcher, HDR]  
Annie Ressouche [Inria, Researcher]

### Faculty Member

Jean Paul Rigault [Univ. Nice, Professor]

### Engineers

Manikandan Bakthavatchalam [Inria]  
Vasanth Bathrinathan [Inria, granted by FP7 DEM@CARE project]  
Antitza Dantcheva [Inria]  
Giuseppe Donatiello [Inria, until Mar 2015, granted by FP7 DEM@CARE project]  
Anais Ducoffe [Inria]  
Baptiste Fosty [Inria, granted by Caisse des Dépôts et Consignations]  
Furqan Muhammad Khan [Inria, from Jun 2015]  
Matias Marin [Inria, from Jul 2015]  
Thanh Hung Nguyen [Inria]  
Jacques Serlan [Inria, until May 2015]  
Remi Trichet [Inria, from Dec 2015]

**PhD Students**

Julien Badie [Inria, granted by FP7 DEM@CARE project, until Nov 2015]  
Carolina Garate Oporto [Inria, granted by FP7 DEM@CARE project, until Aug 2015]  
Auriane Gros [CHU Nice]  
Michal Koperski [Inria, granted by Toyota Motor Europe]  
Farhood Negin [Inria]  
Thi Lan Anh Nguyen [Inria]  
Minh Khue Phan Tran [Inria, granted by CIFRE with Genius]  
Ines Sarray [Inria, since Oct 2015]

**Post-Doctoral Fellows**

Julien Badie [Inria, granted by FP7 DEM@CARE project, since Dec 2015]  
Piotr Tadeusz Bilinski [Inria, granted by FP7 DEM@CARE project]  
Carlos Fernando Crispim Junior [Inria, granted by FP7 DEM@CARE project]  
Malik Souded [Inria, until Apr 2015, granted by Min. de l'Economie]  
Serhan Cosar [Inria, until Jul 2015]  
Slawomir Bak [Inria, until Fev 2015]  
Bernard Boulay [Neosensys engineer]  
Duc Phu Chau [Inria, until Sep 2015]

**Visiting Scientists**

Javier Ortiz [Inria, until Feb 2015]  
Salwa Baabou [Guest PhD, from Nov 2015]  
Siyuan Chen [Guest PhD, from Sep 2015]  
Jesse Hoey [Guest Professor, until Jun 2015]  
Adlen Kerboua [Guest PhD, until Nov 2015]  
Carola Strumia [Guest Student, until Apr 2015]

**Administrative Assistants**

Jane Desplanques [Inria, since May 2015]  
Nadezda Lacroix Coste [Inria, until May 2015]

**Others**

Ghada Bahloul [Inria, Neosensys intership, from Jul 2015 until Sep 2015]  
Marine Chabran [Inria, Internship, until Feb 2015]  
Carolina Da Silva Gomes Crispim [Inria, Neosensys intership, until Sep 2015]  
Kanishka Nithin Dhandapani [Inria, intership, from Jun 2015 since Nov 2015]  
Ramiro Leandro Diaz [Inria, Internship, from Jul 2015]  
Imane Khalis [Inria, Internship, from May 2015 until Sep 2015]  
Nazli Temur [Inria, Internship, from Jun 2015 until Nov 2015]  
Ujjwal Ujjwal [Inria, Internship, from Jun 2015 until Nov 2015]  
Véronique Madeleine Jeanne Bertrand [Inria, from Nov 2015]  
Vania Bogorny [External Collaborator]  
Etienne Corvee [External Collaborator, Linkcare Services]  
Alexandre Derreumaux [External Collaborator, CHU Nice]  
Dario Dotti [Internship, until Nov 2015]  
Daniel Gaffe [External Collaborator, Univ. Nice, Associate Professor, on secondment until Mar 2015]  
Alvaro Gomez Uria Covella [Internship, until Apr 2015]  
Rachid Guerchouche [Neosensys]  
Nathalie Hausermann [Neosensys, from Aug 2015]  
Herve Lartigue [Neosensys, from Mar 2015]  
Anne-Lise Le Maitre [Neosensys, from Aug 2015]  
Ngoc Hai Pham [External Collaborator]  
Yves Pichon [Neosensys]



Annunziato Polimeni [Neosensys]

Philippe Robert [External Collaborator, CHU Nice and COBTECK]

Jean Yves Tigli [External Collaborator, Associate Professor, Univ. Nice]

## 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 analytic.

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.

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

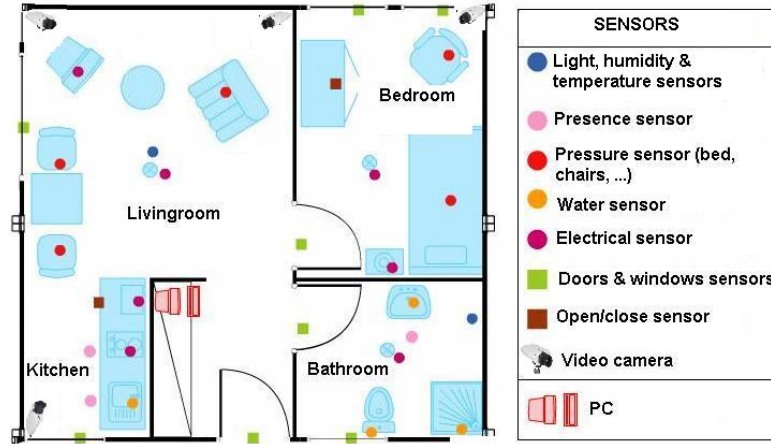


Figure 1. Homecare monitoring: the set of sensors embedded in an apartment

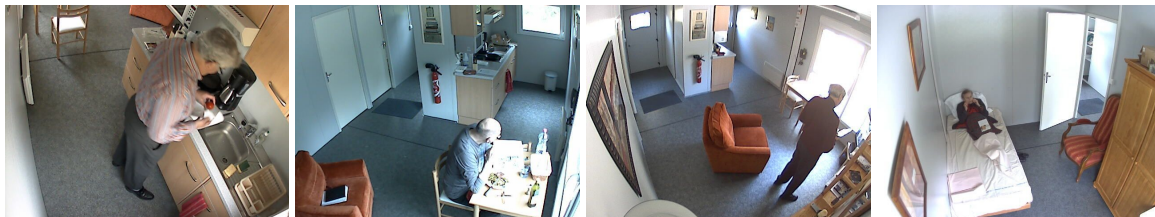


Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

```

Activity (PrepareMeal,
PhysicalObjects(
Components(
    (p : Person), (z : Zone), (eq : Equipment))
    (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

### 2.1.2. *International and Industrial Cooperation*

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

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

## 3. Research Program

### 3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved:** *the software engineering* research 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 analytic and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversely, 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:** 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. Modeling 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 modeling 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.

### 3.3. Semantic Activity Recognition

**Participants:** François Brémond, Sabine Moisan, Monique Thonnat.

Activity Recognition, Scene Understanding, Computer Vision

#### 3.3.1. Introduction

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

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

### 3.3.2. High Level Understanding

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

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

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

### 3.3.3. Learning for Activity Recognition

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

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

### 3.3.4. Activity Recognition and Discrete Event Systems

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

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

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

## 3.4. Software Engineering for Activity Recognition

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

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

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

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

### 3.4.1. Platform Architecture for Activity Recognition

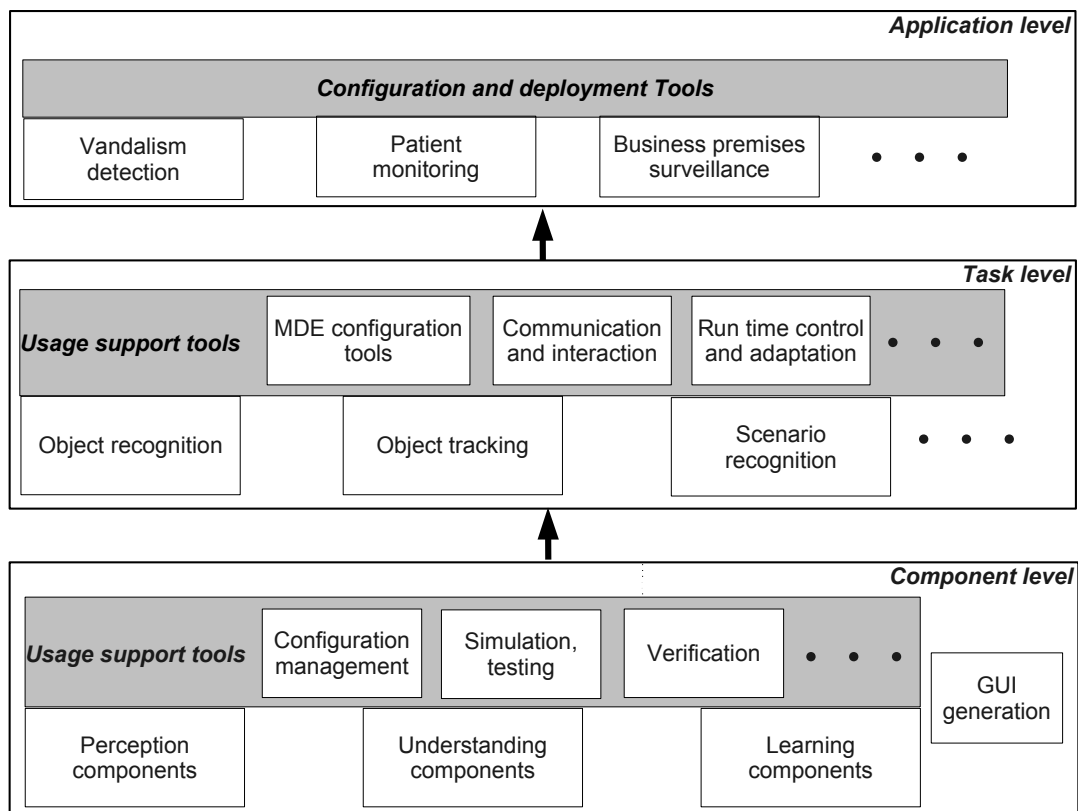


Figure 4. Global Architecture of an Activity Recognition The gray areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

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 [14], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system.

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

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.
  - *Perception components* contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
  - *Understanding components* provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.
  - *Learning components* implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

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

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

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

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

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

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

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

### 3.4.2. Discrete Event Models of Activities

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

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 [46]. 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 [71]. In particular, the constraints can be analyzed 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 [44], [80], [76].

## 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 analytic** and **healthcare monitoring**.

### 4.2. Video Analytics

Our experience in video analytic [6], [1], [8] (also referred to as visual surveillance) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

For instance, the Keeneo start-up was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in video analytic (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

Since 2011, we have initiated a strategic partnership (called CobTek) with Nice hospital [56], [82] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

#### 4.3.1. Research

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

For this domain, a goal for Stars is to allow people with dementia to continue living in a self-sufficient manner in their own homes or residential centers, away from a hospital, as well as to allow clinicians and caregivers remotely provide 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, behavioral patterns, lifestyle, as well as changes in them, indicating the progression of their condition.

#### 4.3.2. Ethical and Acceptability Issues

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

As presented in 3.1, Stars aims at designing cognitive vision systems with perceptual capabilities to monitor efficiently people activities. As a matter of fact, vision sensors can be seen as intrusive ones, even if no images are acquired or transmitted (only meta-data describing activities need to be collected). Therefore new communication paradigms and other sensors (e.g. accelerometers, RFID, and new sensors to come in the future) are also envisaged to provide the most appropriate services to the observed people, while preserving their privacy. To better understand ethical issues, Stars members are already involved in several ethical organizations. For instance, F. Brémond 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. Highlights of the Year

## 5.1. Highlights of the Year

This year Stars has proposed new algorithms in the domains of perception for activity recognition and semantic activity recognition.

### 5.1.1. Perception for Activity Recognition

For perception, the main achievements are:

- A new Re-Identification algorithm which outperforms the State-of-the-art algorithms while being adapted to real-world applications (i.e. it does not require the use of heavy manual annotations which is typical of metric learning algorithms). The remaining challenge is to be able to distinguish people who have similar appearance.
- A new generic action recognition algorithm which outperforms the State-of-the-art algorithms. This algorithm uses new action descriptors that enable finer gesture classification. An open issue is to get a real-time implementation with good enough performance. An extension of this algorithm has been devised for RGB-D cameras, which has been demonstrated in a real-life application, where a robot has to recognize people taking their meal (e.g. eating, drinking).
- New generic tracking algorithms, which can optimize the on-line tuning of tracking parameters and can operate at different temporal scales to recover from lost tracklets. These tracking algorithms have been validated on real world videos lasting more than a week. The utilization of such sophisticated algorithms is still complex and requires some more researches for their deployment in a large variety of applications.

### 5.1.2. Semantic Activity Recognition

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

- New tools to help modeling human activities of daily living. These tools enable to evaluate and improve activity recognition algorithms on long videos depicting the performance of older people living in a nursing home in Nice. The utilization of these tools by clinicians and medical doctors is an ongoing task.
- A new algorithm to recognize human activities, that can benefit from the fusion of events coming from camera networks and heterogeneous sensors.
- A new algorithm to discover human activities of daily living by processing in an unsupervised manner a large collection of videos. The generation of the event models does not require the use of heavy manual annotations which is typical of supervised activity recognition algorithms. However this algorithm still need to have well tracked people to be able to understand their behaviors with sufficient precision.

## 6. New Software and Platforms

### 6.1. CLEM

#### FUNCTIONAL DESCRIPTION

The Clem Toolkit is a set of tools devoted to design, simulate, verify and generate code for LE programs. LE is a synchronous language supporting a modular compilation. It also supports automata possibly designed with a dedicated graphical editor and implicit Mealy machine definition.

- Participants: Daniel Gaffé and Annie Ressouche
- Contact: Annie Ressouche
- URL: <http://www-sop.inria.fr/teams/pulsar/projects/Clem/>

## 6.2. EGMM-BGS

### FUNCTIONAL DESCRIPTION

This software implements a generic background subtraction algorithm for video and RGB-D cameras, which can take feedback from people detection and tracking processes. Embedded in a people detection framework, it does not classify foreground / background at pixel level but provides useful information for the framework to remove noise. Noise is only removed when the framework has all the information from background subtraction, classification and object tracking. In our experiment, our background subtraction algorithm outperforms GMM, a popular background subtraction algorithm, in detecting people and removing noise.

- Participants: Anh Tuan Nghiem, François Brémond and Vasanth Bathrinarayanan
- Contact: François Brémond

## 6.3. MTS

### FUNCTIONAL DESCRIPTION

This software consists of a retrieval tool for a human operator to select a person of interest in a network of cameras. The multi-camera system can re-identify the person of interest, wherever and whenever (s)he has been observed in the camera network. This task is particularly hard due to camera variations, different lighting conditions, different color responses and different camera viewpoints. Moreover, we focus on non-rigid objects (i.e. humans) that change their pose and orientation contributing to the complexity of the problem. In this work we design two methods for appearance matching across non-overlapping cameras. One particular aspect is the choice of the image descriptor. A good descriptor should capture the most distinguishing characteristics of an appearance, while being invariant to camera changes. We chose to describe the object appearance by using the covariance descriptor as its performance is found to be superior to other methods. By averaging descriptors on a Riemannian manifold, we incorporate information from multiple images. This produces mean Riemannian covariance that yields a compact and robust representation. This new software has made digital video surveillance systems a product highly asked by security operators, especially the ones monitoring large critical infrastructures, such as public transportation (subways, airports, and harbours), industrials (gas plants), and supermarkets.

- Participants: Slawomir Bak and François Brémond
- Contact: François Brémond

## 6.4. Person Manual Tracking in a Static Camera Network (PMT-SCN)

### FUNCTIONAL DESCRIPTION

This software allows tracking a person in a heterogeneous camera network. The tracking is done manually. The advantage of this software is to give the opportunity to operators in video-surveillance to focus on tracking the activity of a person without knowing the positions of the cameras in a considered area. When the tracked person leaves the field-of-view (FOV) of a first camera, and enters the FOV of a second one, the second camera is automatically showed to the operator. This software was developed conjointly by Inria and Neosensys.

- Participants: Bernard Boulay, Anaïs Ducoffe, Sofia Zaidenberg, Anais Ducoffe, Annunziato Polimeni and Julien Gueytat
- Partner: Neosensys
- Contact: Anaïs Ducoffe

## 6.5. PrintFoot Tracker

### FUNCTIONAL DESCRIPTION

This software implements a new algorithm for tracking multiple persons in a single camera. This algorithm computes many different appearance-based descriptors to characterize the visual appearance of an object and to track it over time. Object tracking quality usually depends on video scene conditions (e.g. illumination, density of objects, object occlusion level). In order to overcome this limitation, this algorithm presents a new control approach to adapt the object tracking process to the scene condition variations. More precisely, this approach learns how to tune the tracker parameters to cope with the tracking context variations. The tracking context, or video context, of a video sequence is defined as a set of six features: density of mobile objects, their occlusion level, their contrast with regard to the surrounding background, their contrast variance, their 2D area and their 2D area variance. The software has been experimented with three different tracking algorithms and on long, complex video datasets.

- Participants: Duc Phu Chau, François Brémont and Monique Thonnat
- Contact: François Brémont

## 6.6. Proof of Concept Néosensys (Poc-NS)

### FUNCTIONAL DESCRIPTION

This is a demonstration software which gathers different technologies from Inria and Neosensys: PMT-SCN, re-identification and auto-side switch. This software is used to approach potential clients of Neosensys.

- Participants: Bernard Boulay, Sofia Zaidenberg, Julien Gueytat, Slawomir Bak, François Brémont, Annunziato Polimeni and Yves Pichon
- Partner: Neosensys
- Contact: François Brémont

## 6.7. SUP

Scene Understanding Platform

KEYWORDS: Activity recognition - 3D - Dynamic scene

### FUNCTIONAL DESCRIPTION

SUP is a software platform for perceiving, analyzing and interpreting a 3D dynamic scene observed through a network of sensors. It encompasses algorithms allowing for the modeling of interesting activities for users to enable their recognition in real-world applications requiring high-throughput.

- Participants: François Brémont, Carlos Fernando Crispim Junior and Etienne Corvée
- Partners: CEA - CHU Nice - USC Californie - Université de Hamburg - I2R
- Contact: François Brémont
- URL: <https://team.inria.fr/stars/software>

## 6.8. VISEVAL

### FUNCTIONAL DESCRIPTION

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.

- Participants: Bernard Boulay and François Brémont
- Contact: François Brémont
- URL: [http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl\\_Description.html](http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl_Description.html)

## 6.9. py\_ad

py action detection

## FUNCTIONAL DESCRIPTION

Action Detection framework which allows user to detect action in video stream. It uses model trained in py\_ar.

- Participants: Michal Koperski and François Brémond
- Contact: Michal Koperski

## 6.10. py\_ar

py action recognition

## FUNCTIONAL DESCRIPTION

Action Recognition training/evaluation framework. It allows user do define action recognition experiment (on clipped videos). Train, test model, save the results and print the statistics.

- Participants: Michal Koperski and François Brémond
- Contact: Michal Koperski

## 6.11. py\_sup\_reader

## FUNCTIONAL DESCRIPTION

This is a library which allows to read video saved in SUP format in Python.

- Participant: Michal Koperski
- Contact: Michal Koperski

## 6.12. py\_tra3d

py trajectories 3d

## SCIENTIFIC DESCRIPTION

New video descriptor which fuse trajectory information with 3D information from depth sensor.

## FUNCTIONAL DESCRIPTION

3D Trajectories descriptor Compute 3D trajectories descriptor proposed in (<http://hal.inria.fr/docs/01/05/49/49/PDF/koperski-icip.pdf>)

- Participants: Michal Koperski and François Brémond
- Contact: Michal Koperski

## 6.13. sup\_ad

sup action detection

## SCIENTIFIC DESCRIPTION

This software introduces the framework for online/real-time action recognition using state-of-the-art features and sliding window technique.

## FUNCTIONAL DESCRIPTION

SUP Action Detection Plugin is a plugin for SUP platform which performs action detection using sliding window and Bag of Words. It uses an input data model trained in py\_ar project.

- Participants: Michal Koperski and François Brémond
- Contact: Michal Koperski

## 7. New Results

### 7.1. Introduction

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

#### 7.1.1. Perception for Activity Recognition

**Participants:** Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Duc Phu Chau, Etienne Corvée, Antitza Dancheva, Kanishka Nithin Dhandapani, Carolina Garate, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Javier Ortiz, Ujjwal Ujjwal.

The new results for perception for activity recognition are:

- Pedestrian Detection using Convolutional Neural Networks (see 7.2)
- Head detection for eye tracking application (see 7.3)
- Minimizing hallucination in Histogram of Oriented Gradients (see 7.4)
- Hybrid approaches for Gender estimation (see 7.5)
- Automated Healthcare: Facial-expression-analysis for Alzheimer’s patients in musical mnemotherapy (see 7.6)
- Robust Global Tracker based on an Online Estimation of Tracklet Descriptor Reliability (see 7.7)
- Optimizing people tracking for a video-camera network (see 7.8)
- Multi-camera Multi-object Tracking and Trajectory Fusion (see 7.9)
- Person Re-Identification in Real-World Surveillance Systems (see 7.10)
- Human Action Recognition in Videos (see 7.11)

#### 7.1.2. Semantic Activity Recognition

**Participants:** Vasanth Bathrinarayanan, François Brémond, Duc Phu Chau, Serhan Cosar, Alvaro Gomez Uria Covella, Carlos Fernando Crispim Junior, Ramiro Leandro Diaz, Giuseppe Donatiello, Baptiste Fosty, Carolina Garate, Alexandra Koenig, Michal Koperski, Farhood Negin, Thanh Hung Nguyen, Min Kue Phan Tran, Philippe Robert.

For this research axis, the contributions are :

- Evaluation of Event Recognition without using Ground Truth (see 7.12)
- Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition (see 7.13)
- Semi-supervised activity recognition using high-order temporal-composite patterns of visual concepts (see 7.14)
- From activity recognition to the assessment of seniors’ autonomy (see 7.15)
- Serious Games Interfaces using an RGB-D camera (see 7.16)
- Assistance for Older Adults in Serious Game using an Interactive System (see 7.17)
- Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition (see 7.18)

#### 7.1.3. Software Engineering for Activity Recognition

**Participants:** Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Imane Khalis, Nazli Temur, Daniel Gaffé, Rachid Guerchouche, Matias Marin, Etienne Corvée, Carolina Da Silva Gomes Crispim, Anais Ducoffe, Jean Yves Tigli, François Brémond.

The contributions for this research axis are:

- Run-time Adaptation of Video Systems (see 7.19)
- Scenario Description Language (see 7.20)
- Scenario Recognition (see 7.21)
- The Clem Workflow (see 7.22)
- Safe Composition in WComp Middleware for Internet of Things (see 7.23)
- Design of UHD panoramic video camera (see 7.24)
- Brick & Mortar Cookies (see 7.25)
- Monitoring Older People Experiments (see 7.26)

## 7.2. Pedestrian Detection using Convolutional Neural Networks

**Participants:** Ujjwal Ujjwal, François Brémond.

**Keywords:** Pedestrian detection, CNN

The objective of the work was to perform pedestrian detection in different settings. The settings corresponded to different types of camera-views as well as different types of camera settings (e.g- moving camera vs. static camera). The work followed a wide range of experiments using different public implementations of convolutional neural networks and on different types of datasets. We detail the experiments one by one in the following subsections :

### Experiments on CNN architectures

We started with an evaluation of different CNN architectures for pedestrian detection. Towards this end, we implemented three important and famous architectures - LeNet [72], AlexNet [69] and CifarNet [68]. For the purpose of training and validation we extracted patches from the public datasets of Inria [55], Daimler [58], TUD-Brussels [92], Caltech [57], ViPer [62], USC [93] and MIT [78]. The breakup of the dataset used for training was as shown in table 1.

Table 1. Training, testing and validation patch details for CNN training

	Pedestrian Patches	Non-Pedestrian Patches
Training	131,183	61,500
Validation	65,591	30,700
Testing	65,591	30,600

Implementation of all the three models for pedestrian detection which gave a very high accuracy ( 94.2% (LeNet), 98% (AlexNet) and 98.2% (CifarNet) ) for classification at patch level. Though these results were good at the patch level, more thorough understanding was needed to determine the effect of network architecture on classification. This was important because the three architectures vary greatly in terms of number of layers and other parameters. Moreover the practical problem in pedestrian detection chiefly involves detecting pedestrians in an image (i.e. when full-scale images instead of pre-defined patches are available).

The first set of experiments was done using sliding windows. This had to be abandoned soon, since for each image this was taking an impractical time ( $> 3minutes/image$ ). Moreover sliding window is less suited in its naive setting due to the fact that each candidate window had to be rescaled to meet the network input size and tested individually by extracting features over it. This was followed by efforts to understand and implement a wide range of other techniques for full-scale detection using CNN. This is still an open problem though some encouraging advancements through R-CNN [61] and OverFeat [83] have been made. A major difficulty lies in lack of robust implementations of CNN which allow for integrated training and testing with object localization. Moreover existing implementations are less flexible and often make it difficult to carry out modifications required to implement new techniques independently.

We settled with the R-CNN which uses region proposals extracted using selective search [87] to extract object proposals and then train a CNN using those proposals and subsequently classifying using a SoftMax classifier or a SVM.

The evaluation was done on both moving cameras and static cameras and the evaluation showed that the network was performing a little satisfactorily, though below the state-of-art performance standards. The performance metric was Average Miss Rate (AMR) Vs. False Positives Per Image (FP/I). A good detector must exhibit a very low AMR alongwith very low FP/I. Table 2 summarizes the detection results, with table 3 summarizing state-of-art results on different pedestrian detection datasets.

Table 2. Our R-CNN results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.27	0.36
DAIMLER	15K	0.38	0.40
Caltech	16K	0.46	0.43
USC	584	0.02	0
PETS 2009 S2.L1	5565	0.42	0.29
PETS 2009 S2.L2	1744	0.35	0.19

Table 3. State-of-art results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.14	0.1
DAIMLER	15K	0.29	0.21
Caltech	16K	0.12	0.1
PETS 2009 S2.L1	5565	0.22	0.1
PETS 2009 S2.L2	1744		

It was felt that more work is needed to organize a CNN library and subsequently work to improve the above results.

### Conclusion

Pedestrian detection finds its applications in different settings. It is also highly influenced by a wide variety of variations which have many practical ramifications in areas such as surveillance. It is important to develop a robust and high-performance system for pedestrian detection that is able to take into account a very wide range of such variations such as occlusion and poor visibility. CNNs have shown great promise in object detection and recognition lately and this inspires it growing applications in pedestrian detection. While the current results of our R-CNN experiments do not match the state-of-art it has shown some promise by providing consistent numbers across datasets which shows that CNNs are a good way to transcend a system beyond dataset-specific restrictions. An important factor is the instance of moving cameras vis-a-vis static cameras. While the present experiments show that decent performance is obtained on moving camera databases, consistent and similar performance is also obtained in the context of static camera databases such as PETS. This shows that with better training and improved practices of dataset handling such as augmentation and dataset structuring by clustering based methods can help in pushing the performance to acceptable levels for applications in automated surveillance and driving applications.

**Further Work** We intend to take this study forward, by looking into novel approaches to gather more information about a pedestrian dataset from CNN, while further increasing the detection results.

## 7.3. Head Detection for Eye Tracking Application

**Participants:** Thanh Hung Nguyen, Antitza Dantcheva, François Brémond.



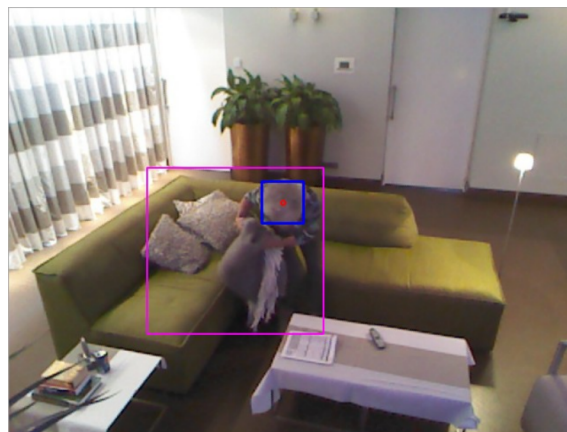
**Keywords:** computer vision, head detection, eye tracking

Head detection [77] uses RGBD sensor (Kinect 2 sensor) which is supported by SUP platform of STARS team. For the eye tracking, we use the open-source library (OpenBR) which has good performance in our test.

Until now, the head detector was working well when people were standing but not good enough when people were lying or bending as you can see on Figure 5 and Figure 6. This experiment was realised on simple datasets where mostly people was close to camera and walking. The main reason is the lack of samples in the learning process for the challenges cases (lying, bending). So at this moment we are collecting the head samples for it.

Number of true positives	209
Number of false positives	28
Number of false negatives	25

*Figure 5. Performance of head detection on simple dataset*



*Figure 6. An example of experiment*

## 7.4. Minimizing Hallucination in Histogram of Oriented Gradients

**Participants:** Javier Ortiz, Slawomir Bak, Michal Koperski, François Brémont.

**keywords:** computer vision, action recognition, re-identification,

**Challenges in histogram representation**

In computer vision very often histogram of values is used as a feature representation. For instance HOG descriptor is in fact histogram of gradient orientations. Also histogram of codewords in Bag of Words representation is a very popular action recognition representation. In such a case we are not interested in absolute values for given bin, but rather in shape of histogram to find the patterns. To make histogram representation independent from absolute values we use L1 or L2 normalization. In the normalization process we treat histogram as a vector and we transform it in the way that it should have unit length according to given norm (L1, L2). The drawback of such approach is that normalization process may amplify the noise for histograms where absolute values are very small (no pattern or only noise). Such histogram after normalization can be very similar to histogram with strong absolute values. This situation is showed in figure 7. Although original histogram in second row of figure 7 contains almost no information after normalization the values are amplified and the result is exactly the same as for histogram in the first row. Such behavior is called hallucinations.

### Proposed normalization method

We propose to add an artificial bin with given value which would prevent small noisy values from being amplified during normalization process. In figure 7 in first row we show artificial bin in pink color. Thanks to that histograms after normalization (last column) are different. If we analyze cumulative sum of histogram values across data (we sum values of all bins in whole histogram and we draw distribution), we can find that in some data-sets we obtain bi-nomial distribution see figure 8. The gap between two Gaussians indicates convenient border between noise and meaningful data. On the other hand many data-sets do not have this bi-nomial feature and for that we do not have formalization to find the value of artificial bin. This problem would be a subject of further studies.

The following method was successfully applied to person re-identification and action recognition problem. Further details can be found in the paper [42]

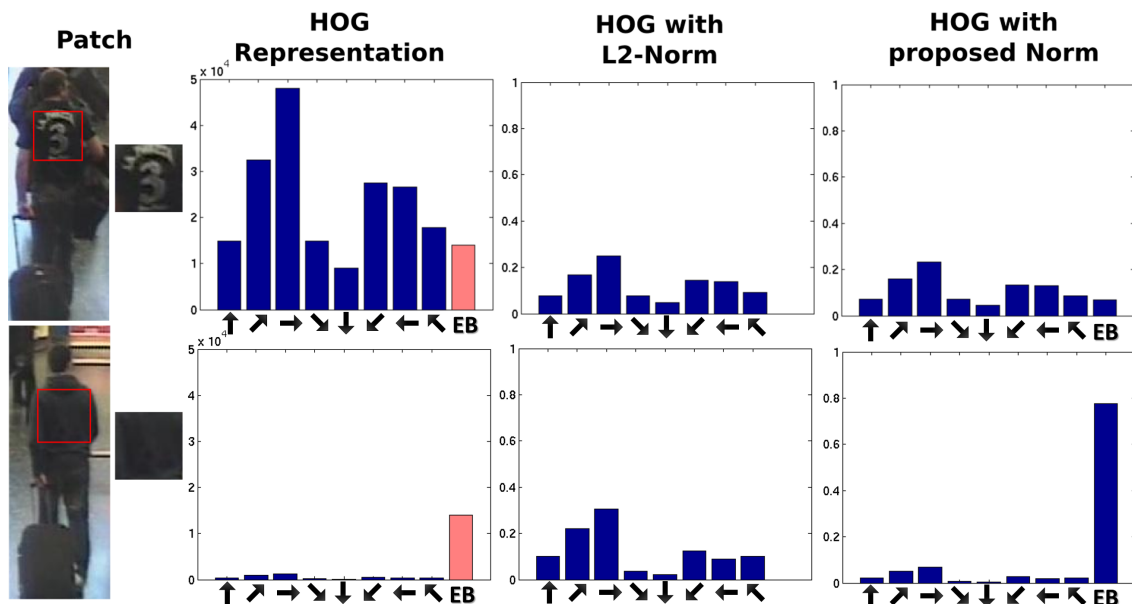


Figure 7. HOG representations of two patches with different amount of texture. Each bin in blue represents the sum of magnitudes of edges for a particular orientation. The bin marked as EB represents the proposed extra bin

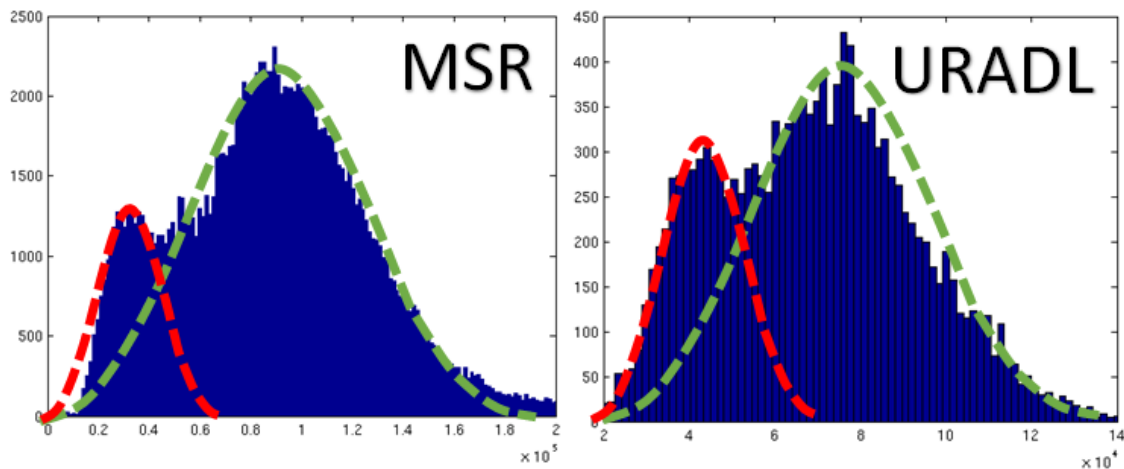


Figure 8. Distribution of the sum of bins for randomly sampled patches for URADL and MSRDailyActivity3D data-set.

## 7.5. Hybrid Approaches for Gender estimation

**Participants:** Antitza Dantcheva, François Brémond, Philippe Robert.

**keywords:** gender estimation, soft biometrics, biometrics, visual attributes

Automated gender estimation has numerous applications including video surveillance, human computer-interaction, anonymous customized advertisement and image retrieval. Gender estimation remains a challenging task, which is inherently associated with different biometric modalities including fingerprint, face, iris, voice, body shape, gait, as well as clothing, hair, jewelry and even body temperature [31]. Recent work has sought to further the gains of single-modal approaches by combining them, resulting into hybrid cues that offer a more comprehensive gender analysis, as well as higher resilience to degradation of any of the single sources.

### Can a smile reveal your gender?

In this work we propose a novel method for gender estimation, namely the use of dynamic features gleaned from smiles and show that (a) facial dynamics can be used to improve appearance-based gender-estimation, (b) that while for adults appearance features are more accurate than dynamic features, for subjects under 18 years old facial dynamics outperform appearance features. While it is known that sexual dimorphism concerning facial appearance is not pronounced in infants and teenagers, it is interesting to see that facial dynamics provide already related clues. The proposed system, fusing a state-of-the-art appearance and dynamics-based features (see Figure 9), improves the appearance based algorithm from 78.0% to 80.8% in video-sequences of spontaneous smiles and from 80% to 83.1% in video-sequences of posed smiles for subjects above 18 years old (see Table 4). These results show that smile-dynamics include pertinent and complementary information to appearance gender information.

While this work studies video sequences capturing frontal faces expressing human smiles, we can envision that additional dynamics, such as other facial expressions or head and body movements carry gender information as well.

**Distance-based gender prediction: What works in different surveillance scenarios?**

Table 4. True gender classification rates. Age given in years.

Age	< 10	10 – 19	20 – 29	30 – 39	40 – 49	> 49
Subj. amount	48	95	60	49	72	33
OpenBR	58.33%	50.53%	81.67%	75.51%	75%	81.82%
<b>Combined Age-Groups</b>	< 20		> 19			
Subj. amount	143		214			
OpenBR	52.45%		78.04%			
Dynamics (SVM+PCA)	60.1%		69.2%			
Dynamics (AdaBoost)	59.4%		61.7%			
OpenBR + Dynamics (Bagged Trees)	60.8%		80.8%			

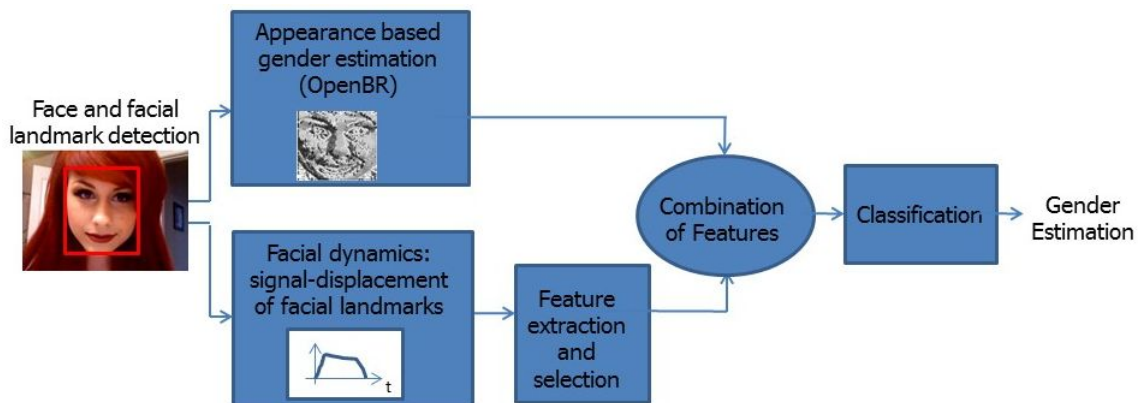


Figure 9. General Scheme of the facial appearance and dynamics framework.

In this work we fuse features extracted from face, as well as from body silhouette towards gender estimation. Specifically, for face, a set of facial features from the OpenBR library, including histograms of local binary pattern (LBP) and scale-invariant feature transform (SIFT) are computed on a dense grid of patches. Subsequently, the histograms from each patch are projected onto a subspace generated using PCA in order to obtain a feature vector, followed by a Support Vector Machine (SVM) used for the final face-based-gender decision. Body based features include geometric and color based features, extracted from body silhouettes, obtained by background subtraction, height normalization and SVM-classification for the final body-based-gender-decision. We present experiments on images extracted from video-sequences, emphasizing on three distance-based settings: close, medium and far from the TunnelDB dataset (see Figure 10). As expected, while face-based gender estimation performs best in the close-distance-scenario, body-based gender estimation performs best when the full body is visible - in the far-distance-scenario (see Table 5). A decision-level-fusion of face and body-based features channels the benefits of both approaches into a hybrid approach, providing results that demonstrate that our hybrid approach outperforms the individual modalities for the settings medium and close.

Table 5. True gender classification rates. Age given in years.

System Distance	FAR	MEDIUM	CLOSE
FBGE	57.14	79.29	89.29
BBGE 89.3 85 79.3			
Fusion BBGE & FGBE	82.9	88.6	95



Figure 10. Three distance-based settings in the TunnelDB dataset: far, medium and close.

While the dataset is relatively unconstrained in terms of illumination, body and face are captured facing relatively frontally towards the camera. Future work will involve less constraints also towards the pose of humans.

## 7.6. Automated Healthcare: Facial-expression-analysis for Alzheimer's Patients in Musical Mnemotherapy

**Participants:** Antitza Dantcheva, François Brémond, Philippe Robert.

**keywords:** automated healthcare, healthcare monitoring, expression recognition

In this work we seek to apply computer vision towards increasing the life quality of patients with Alzheimer's disease (AD), and particularly in applying computer vision towards interventions to delay functional decline and to decrease the burden of the most disturbing behavioral symptoms. Towards this we design a smart interaction tool, that "reads" emotions of AD patients. This approach is becoming necessary now, because the increasing prevalence of chronic disorders and its impact on functional decline is challenging the sustainability

of healthcare systems. Firstly, we have assembled a dataset of video-sequences acquired in the Alzheimer's Disease - clinique Fondation GSF Noisiez. Multiple patients and sessions have been captured during musical mnemotherapy. We then have annotated several sequences per one of four facial expressions, that occur in the recorded dataset including: neutral, talking, smile and sad. We then proceed to classify these expressions for 10 patients based on two approaches, that we study individually, as well as fused. The first approach contains face detection, facial landmark localization and signal displacement analysis for different facial landmarks, which are ranked based on categorization-pertinence, fused and classified into one of the four expression-categories (see Figure 11). In the second approach, we use face detection, eyes-detection, face normalization and HOG-features, which we classify into one of the four expression-categories.

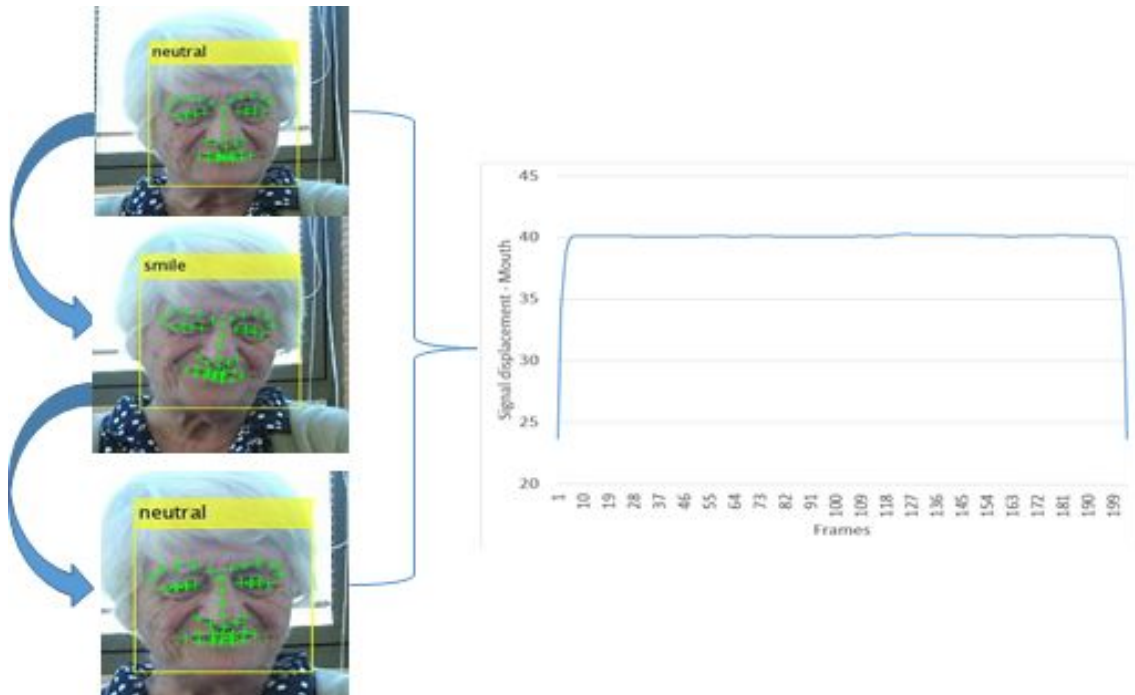


Figure 11. Smile detection based on signal displacement in the mouth region in the dataset collected at the Fondation GSF Noisiez in Biot, France.

The here used real-world-data challenges, as expected, all utilized computer vision algorithms (from face detection - due to no constraints of pose and illumination, to classifiers - due to a large intra-class-variation of facial expressions). Nevertheless, we obtain promising results that we envision improving by analyzing 2D and depth, as well as infrared data.

## 7.7. Robust Global Tracker Based on an Online Estimation of Tracklet Descriptor Reliability

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

**Keywords:** Tracklet fusion, Multi-object tracking

Object tracking - the process of locating a moving object (or multiple objects) in one camera or in a camera network over time - is an important part in surveillance video processing. However, the video context variation requires trackers to face plenty of challenges. For example, objects change their movement direction or their appearances, poses; objects are occluded by other objects or background; illumination is changed... In order to overcome above challenges, calculating the object appearance model overtime to adapt tracker to context variation is a necessary work.

In the state of the art, some online learning approaches [52], [48] have been proposed to track objects in various video scenes in each frame. These approaches learn online discriminative object descriptors to the current background as in [52] or learn an object appearance model which discriminates objects overtime as in [48]. However, the limitation of these approaches is that the reliability of object descriptor computed on only current frame is sensitive. False positives can reduce tracking quality. Furthermore, these algorithms try to find the discriminative descriptors or signatures of one object compared to its neighborhood but not considering to the correlation of this object with its can-match candidates. Meanwhile, global tracking methods [91], [98] show their dominant ability over previous methods in noisy filtering. The approach in [98] proposes an algorithm that recovers fragmentation of object trajectories by using enhanced covariance-based signatures and an online threshold learning. The approach in [91] proposes a hierarchical relation hypergraph based tracker. These global tracking algorithms have significant results in matching short trajectories and filtering some noise. However, object descriptor weights are fixed for the whole video. Therefore, their tracking performances can be reduced if the scene changes.

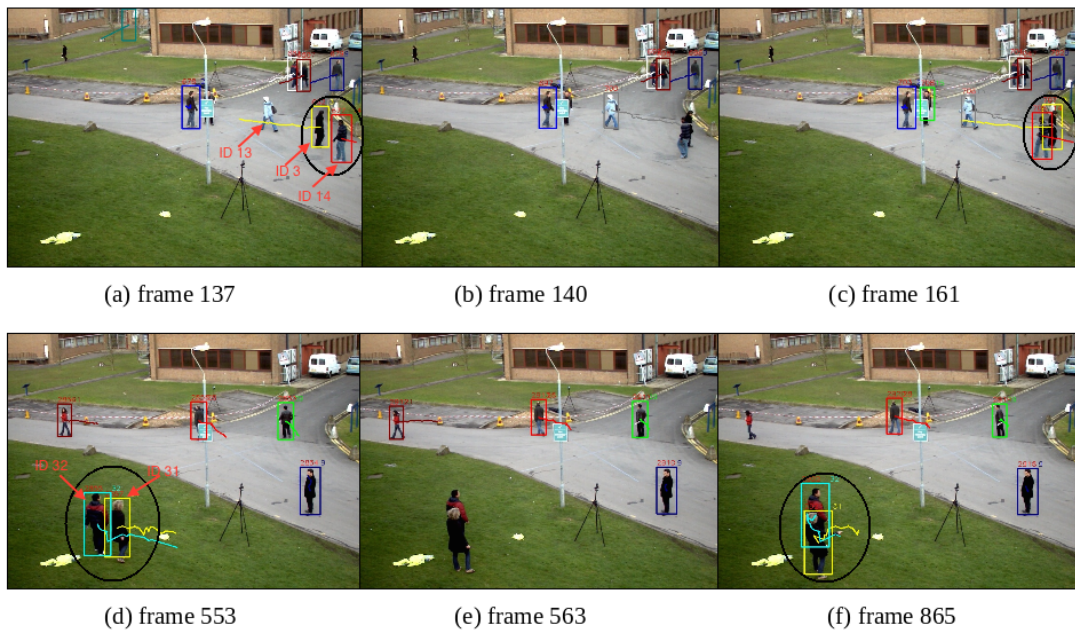


Figure 12. PETs2009 dataset: The online computation of discriminative descriptor weights depending on each video scene.

In this work, we propose a new approach to improve the tracking quality by a global tracker which merges all tracklets belonging to an object in the whole video. Particularly, we compute online descriptor reliability over time based on their discrimination. Based on the computed discriminative descriptor weights, the global matching score over descriptors of 2 tracklets is given. Then, we apply Hungarian algorithm to optimize

Table 6. Tracking performance. The best values are printed in bold, the second best values are printed in italic.

Dataset	Method	MOTA	MOTP	GT	MT	PT	ML	FG
PETS2015	Chau <i>et Al.</i> [53]	–	–	2	0.0	100.0	<b>0.0</b>	2
	<b>Ours</b> (Proposed approach + [53])	–	–	2	<b>100.0</b>	0.0	<b>0.0</b>	<b>1</b>
PETS2009	Chau <i>et Al.</i> [53]	0.62	0.63	21	–	–	–	8
	Bae <i>et Al.</i> with all [48]	0.83	0.69	23	<b>100</b>	0	<b>0.0</b>	<b>4</b>
	<b>Zamir et Al.</b> [95]	<b>0.90</b>	0.69	21	–	–	–	–
	<b>Bae et Al.</b> -global association [48]	0.73	0.69	23	<b>100</b>	0	<b>0.0</b>	12
	<b>Badie et Al.</b> [47]	<b>0.90</b>	<b>0.74</b>	21	–	–	–	–
	<b>Badie et Al.</b> [47] + [53]	0.85	0.71	21	66.6	23.9	9.5	6
	<b>Ours</b> (Proposed approach + [53])	<i>0.86</i>	<i>0.72</i>	21	76.2	14.3	9.5	<b>4</b>
TUD-Stadtmitte	Milan <i>et Al.</i> [74]	<b>0.71</b>	<b>0.65</b>	9	<b>70.0</b>	20.0	<b>0.0</b>	–
	Yan <i>et Al.</i> [94]	–	–	10	<b>70.0</b>	30.0	<b>0.0</b>	–
	Chau <i>et Al.</i> [53]	0.45	0.62	10	60.0	40.0	<b>0.0</b>	13
	<b>Ours</b> (Proposed approach + [53])	0.47	<b>0.65</b>	10	<b>70.0</b>	30.0	<b>0.0</b>	<b>7</b>
TUD-Crossing	Tang <i>et Al.</i> [84]	–	–	–	<b>53.8</b>	38.4	7.8	–
	Chau <i>et Al.</i> [53]	0.69	0.65	11	46.2	53.8	<b>0.0</b>	14
	<b>Ours</b> (Proposed approach + [53])	<b>0.72</b>	<b>0.67</b>	11	<b>53.8</b>	46.2	<b>0.0</b>	<b>8</b>



tracklet matching. On the other hand, a motion model is also combined with appearance descriptors in a flexible way to improve the tracking quality. Figure 12 shows the visual explanation. In frame 137, two objects have similar appearance but move with different direction. In this case, motion descriptor is more reliable. Inversely, in frame 553, two objects go consistently together but their coat and hair's colors are different. Therefore, the appearance descriptors are more reliable than motion one.

The proposed approach gets results of tracker in [53] as input and is tested on challenge datasets. The comparable results of this tracker with other trackers from the state of the art are shown in Table 6. This paper is accepted in PETs workshop [41].

## 7.8. Optimizing People Tracking for a Video-camera Network

**Participants:** Julien Badie, François Brémont.

**Keywords:** tracking quality estimation, error recovering, tracklet matching

This work addresses the problem of improving tracking quality during runtime. Most state-of-the-art tracking or high-level algorithms such as event recognition have difficulties to handle erroneous inputs. This framework detects and repairs detection or tracking errors. It works in an online situation and even in the case where prior knowledge of the scene (such as contextual information or training data) is not available.

The Global Tracker (figure 13) uses tracking results (tracklets) as input and produces corrected tracklets as output.

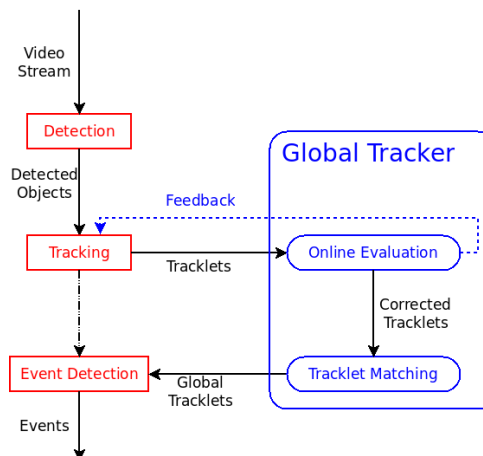


Figure 13. The Global Tracker framework, combining online evaluation and tracklet matching to improve tracking results.

The Global Tracker framework is divided into two main modules:

- Online evaluation of tracking results:** the quality of the tracking results is computed by analyzing the variation of each tracklet feature. A significant feature variation is considered as a potential error, an anomaly. To determine if this anomaly is a real error or a natural phenomenon, we use information given by the object neighborhood and the context. Finally the errors are corrected either by removing the erroneous nodes (basic approach) or by sending a signal to the tracking algorithm in order to tune its parameters for the next frames (feedback approach).

- **Tracklets matching over time:** tracklets representing the same object are merged together in a four-step algorithm. First we select key frames (frames that are close to the mean value of the features) for each tracklet. Then a visual signature is computed based on these key frames. The distance between each pair of signature is then computed. Finally the tracklet merging is done using unsupervised learning and a constrained clustering algorithm where all tracklets representing the same object are put in the same cluster.

This approach has been tested on several datasets such as PETS2009 (table 7), CAVIAR (table 8), TUD, I-LIDS and VANAHEIM and with different kinds of scenarios (tracking associated with a controller, 3D camera, camera network with overlapping or distant cameras). In each case, we are able to reach or outperform the state-of-the-art results.

Table 7. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset

Methods	MOTA	MOTP
Zamir et al. [95]	<b>0.90</b>	0.69
Milan et al. [75]	<b>0.90</b>	<b>0.74</b>
<b>Online evaluation</b>	0.90	0.74
<b>Tracklet matching</b>	0.83	0.68
<b>Global Tracker</b>	<b>0.92</b>	<b>0.76</b>

Table 8. Tracking results on the Caviar dataset

Methods	MT (%)	PT (%)	ML (%)
Li et al. [73]	84.6	14.0	1.4
Kuo et al. [70]	84.6	14.7	<b>0.7</b>
<b>Online Evaluation</b>	82.6	11.7	5.7
<b>Tracklet matching</b>	84.6	9.5	5.9
<b>Global Tracker</b>	<b>86.4</b>	<b>8.3</b>	5.3

This approach is described more in detail in the PhD manuscript [27].

## 7.9. Multi-camera Multi-object Tracking and Trajectory Fusion

**Participants:** Kanishka Nithin Dhandapani, Thi Lan Anh Nguyen, Julien Badie, François Brémont.

**Keywords:** Multicamera, Tracklet association, Trajectory fusion, Object Tracking.

In spite of number of solutions that exist for multi-object tracking, it is still considered most challenging and unsolved computer vision problems, mainly due to inter and intra-occlusions, inferior visibility in crowded scenes, object re-entry, abrupt movement of object, placement of cameras and other detection inaccuracies that occur in single camera. Such drawbacks in single-camera multi-target tracking can be solved to an extent by obtaining more visual information on the same scene (more cameras). Few works done in the past years are [50], [79], [65], [59]. However they have their own problems such as not real run time performance, complex optimizations, hypothesizing 3D reconstruction and data association together might lead to suboptimal solutions.

We present a multi-camera tracking approach that associates and performs late fusion of trajectories in a centralized manner from distributed cameras. We use multiple views of the same scene to recover information that might be missing in a particular view. For detection we use background subtraction followed by discriminatively Trained Part Based Models. For object tracking, we use an object appearance-based tracking algorithm introduced by Chau et al [54] that combines a large set of appearance features such as 2D size, 3D displacement, colour histogram, and dominant colour to increase the robustness of the tracker to manage occlusion cases. Each camera in the network runs the detection and tracking chain independent of each

other in a distributed manner. After a batch of frames, the data from each camera is gathered to a central node by projecting the trajectories of people to the camera with the most inclusive view through a planar homography technique and then global association and fusion are performed. Unlike the temporally local (frame to frame) data-association method, global data association has ability to deal with challenges posed by noisy detections. Global association also increases the temporal stride under optimization, therefore more stable and discriminative properties of targets can be used. Trajectory similarities are calculated as heuristically weighted combination of individual features based on geometry, appearance and motion. Association is modeled as a complete K-partite graph (all pairwise relationships inside the temporal window are taken into account) K corresponds to number of cameras in network. For simplicity purpose, we use K=2. Since we use complete K-partite graph, we have an optimal solution. Whereas methods that model association as complex multivariate optimization, upon scaling, face the problem of being stuck at local minima and may provide sub-optimal solutions. Fusion is performed using adaptive weighting method. Where the weights are derived from reliability attribute of each tracker. This enables correct and consistent trajectories after fusion even if the individual trajectories have inherent noises, occlusion and false positives

Our approach is evaluated on the publicly available PETS2009 dataset. PETS2009 is a challenging dataset due to its low FPS and interobject occlusions. We choose View1, View3 and View5 in S2.L1 scenario to evaluate. The results can be seen in Figure 14

The results are encouraging and are very raw and preliminary with lot of scope and room for improvement. With more fine tuning, error rate can be improved. However too significant errors in people detection to build on top of it. Thus, we need training detector on specific datasets to improve the approach. As future work, we will study if we can improve the optimization stage with a more complex optimization using minimal graph flow would improve the results drastically.

Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS 2009 S2.L1	Berclaz et al. [1]	1,3,5,6,8	82	56	-	-	-
	Leal-Taixe et al. [2]	1,5	76	60	-	-	-
	Leal-Taixe et al. [2]	1,5,6	71.4	53.4	-	-	-
	Murray Evans et al. [4]	2 Cameras	63	55			
	Martin Hofmann et al. [3]	1,5	99.4	82.9	100	0	0
	Martin Hofmann et al. [3]	1,5,7	99.4	83.0	100	0	0

Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS2009 S2.L1	Our Approach	1,3	76.33	65.28	92.59	0.035	0.714

Figure 14. Result of our approach on PET2009 dataset.

## 7.10. Person Re-identification in Real-World Surveillance Systems

**Participants:** Furqan Mohammad Khan, François Brémond.

**Keywords:** re-identification, long term visual tracking, signature modeling

**Cost of supervised metric learning** Person re-identification problem has recently received a lot of attention and the recent focus is to use supervised model training to learn cross camera appearance transformation. In general,  $O(n^2)$  models are trained in a surveillance network with  $n$  cameras, one for each camera pair.  $2p$  tracks are required to train one model with  $p$  person identities. In a real-world surveillance network with non-overlapping fields of view, a person appears only in a subset of cameras (see figure 15, courtesy of [51]). This puts the requirement of number of tracks to train all models at  $O(n^2p)$ , or more precisely,  $n(n-1)p$ . That

is, to train each model with 100 people in a 10 camera network we need 9000 tracks. For supervised training, these tracks need to be given consistent identities, and worse, have their bounding boxes marked. This is a significant burden on human annotators for deployment in real-world. Further, the annotation cost has to be repaid at a significant fraction if only one new camera is added to the system (may be due to failure of an existing camera), or if the lighting changes significantly (in case of outdoor surveillance). In our opinion, this is a significant bottleneck for supervised metric learning based re-identification in real-world.

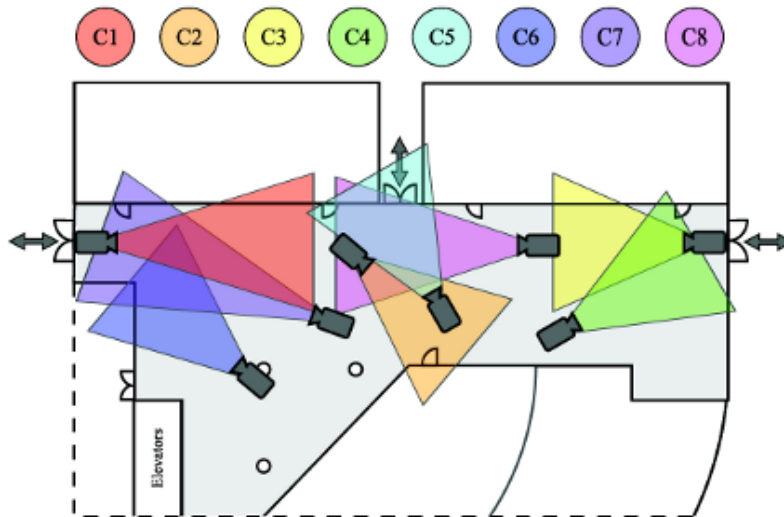


Figure 15. Camera arrangement in multi-camera surveillance scenario of SAIVT-SoftBio dataset [51]

**Improved re-identification through signature modeling** Re-identification is challenging because variance is intra-class appearance is often higher than inter-class appearance due to varying lighting conditions and viewpoints, and non-uniqueness of clothing. More importantly, in real-world when re-identification is fed by automated human detectors and trackers, significant mis-alignment or partial visibility of the person within proposed bounding box make it difficult to extract relevant features. Our work focuses on improving signature construction from low level features for multi-shot re-identification. We explicitly model multi-modality of person appearance using a feature mixture (corresponding publication is under review at this moment). This improves state-of-the-art re-identification performance on SAIVT-SoftBio [51] dataset and performs equally well as state-of-the-art metric learning methods on iLIDS-VID [88] and PRID2011 [64] datasets. The performance comparison of our method with state-of-the-art is presented using CMC in figure 16 (our results are denoted by MCAM).

## 7.11. Human Action Recognition in Videos

**Participants:** Piotr Bilinski, François Brémont.

**Keywords:** Action Recognition, Video Covariance Matrix Logarithm, VCML, Descriptor  
**Video Covariance Matrix Logarithm for Human Action Recognition in Videos**

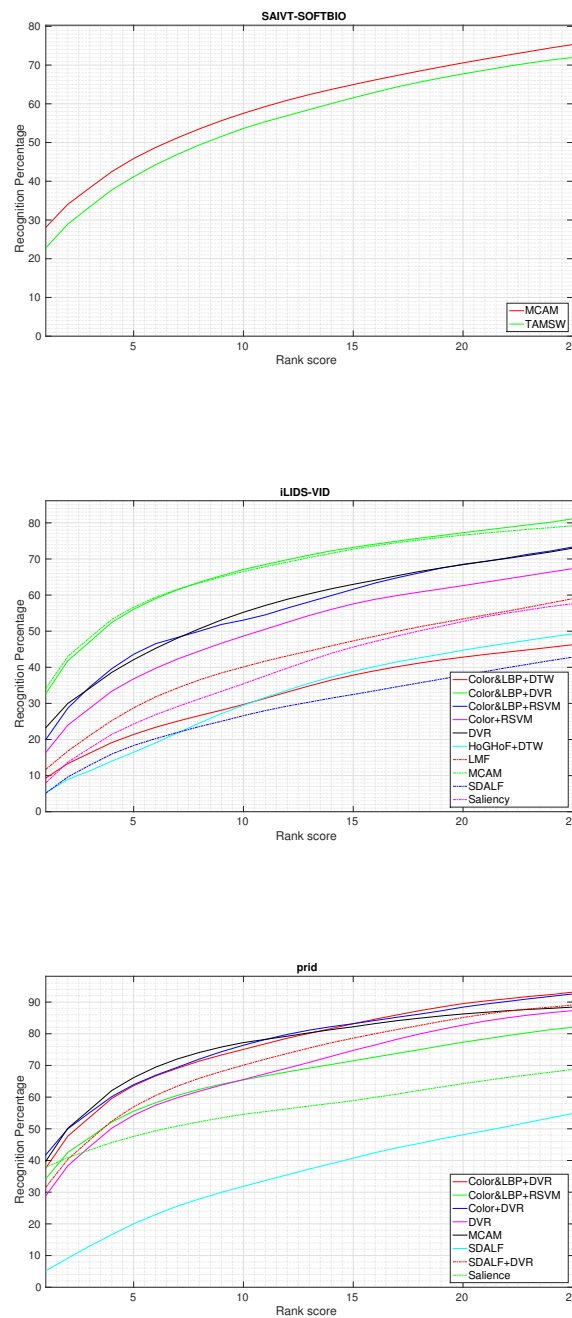


Figure 16. Performance comparison of our MCAM approach using CMC curves on different datasets. **Top:** Comparison with TAMSW [49] on SAIVT-SoftBio dataset; **middle:** Comparison with Color+RSVM [88], Color&LBP+DTW [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], HoG+HoF+DTW [88], LMF [97], Saliency [96], and SDALF [60] on iLIDS-VID dataset; **bottom:** Comparison with Color+DVR [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], Saliency [96], and SDALF [60] on PRID2011 dataset.

In this work, we propose a new local spatio-temporal descriptor for videos and we propose a new approach for action recognition in videos based on the introduced descriptor. Overview of the proposed action recognition approach based on the introduced descriptor is presented in Figure 17. The new descriptor is called the Video Covariance Matrix Logarithm (VCML). The VCML descriptor is based on a covariance matrix representation, and it models relationships between different low-level features, such as intensity and gradient. We apply the VCML descriptor to encode appearance information of local spatio-temporal video volumes, which are extracted by the (Improved) Dense Trajectories. Then, we present an extensive evaluation of the proposed VCML descriptor with the (Improved) Fisher Vector encoding and the Support Vector Machines on four challenging action recognition datasets (*i.e.* URADL, MSR Daily Activity 3D, UCF50, and HMDB51 datasets). We show that the VCML descriptor achieves better results than the state-of-the-art appearance descriptors. In comparison with the most popular visual appearance descriptor, *i.e.* the HOG descriptor, the VCML achieves superior results. Moreover, we present that the VCML descriptor carries complementary information to the HOG descriptor and their fusion gives a significant improvement in action recognition accuracy (*e.g.* the VCML improves the HOG by 15% on the HMDB51 dataset). Finally, we show that the VCML descriptor improves action recognition accuracy in comparison to the state-of-the-art (Improved) Dense Trajectories, and that the proposed approach achieves superior performance to the state-of-the-art methods. The proposed VCML based technique achieves 94.7% accuracy on the URADL dataset, 85.9% on the MSR Daily Activity 3D dataset, 92.1% on the UCF50 dataset, and 58.6% on the HMDB51 dataset. More results and comparisons with the state-of-the-art are presented in Table 9 and Table 10. To the best of our knowledge, this is the first time covariance based features are used to represent the trajectories. Moreover, this is the first time they encode the structural information and they are applied with the (Improved) Fisher Vector encoding for human action recognition in videos. This work has been published in [40].

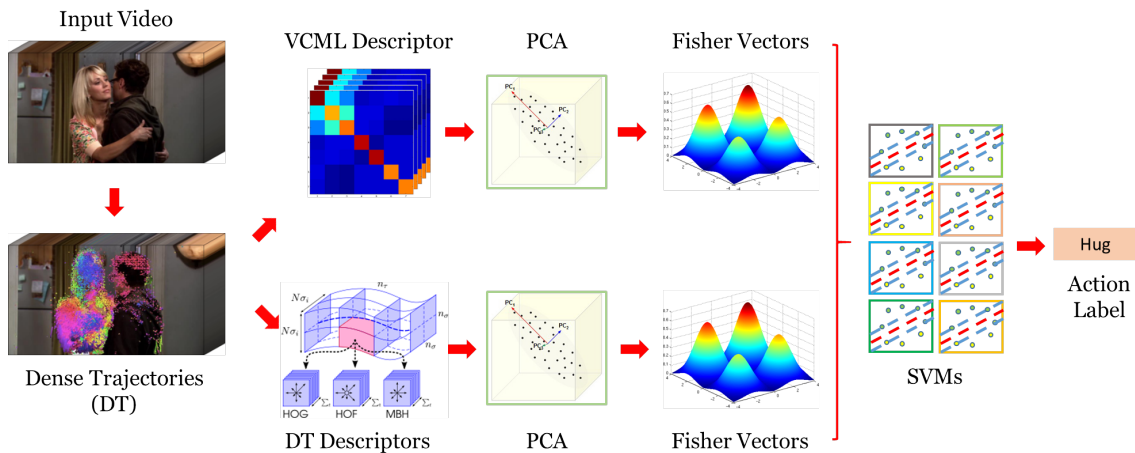


Figure 17. Overview of our action recognition approach based on the introduced VCML descriptor.

## 7.12. Evaluation of Event Recognition without Using Ground Truth

**Participants:** Ramiro Diaz, Carlos Fernando Crispim Junior, François Brémond.

**Keywords:** Computer Vision, Event Recognition, Video Summarization.

The main goal of the work is to improve the Event Recognition process and to improve the way we build the event models as well. The work concerns the Valrose Nursing Home, it consists in monitoring older people with health issues like Dementia and who are in need of care and stimulation.

Table 9. Comparison with the state-of-the-art on the URADL and MSR Daily Activity 3D datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

URADL		MSR Daily Activity 3D	
Benabbas <i>et Al.</i> , 2010	81.0	Koperski <i>et Al.</i> , 2014	72.0
Raptis and Soatto, 2010	82.7	JPF – Wang <i>et Al.</i> , 2012	78.0
Messing <i>et Al.</i> , 2009	89.0	Oreifej and Liu, 2013	80.0
Bilinski and Bremond, 2012	93.3	AE – Wang <i>et Al.</i> , 2012	85.7
Dense Trajectories	94.0	Dense Trajectories	76.2
<b>Our Approach (DT)</b>	<b>94.0</b>	<b>Our Approach (DT)</b>	<b>78.1</b>
<b>Our Approach (IDT)</b>	<b>94.7</b>	<b>Our Approach (IDT)</b>	<b>85.9</b>

Table 10. Comparison with the state-of-the-art on the UCF50 and HMDB51 datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

UCF50		HMDB51	
Kantorov and Laptev, 2014	82.2	Kantorov and Laptev, 2014	46.7
Shi <i>et Al.</i> , 2013	83.3	Jain <i>et Al.</i> , 2013	52.1
Oneata <i>et Al.</i> , 2013	90.0	Oneata <i>et Al.</i> , 2013	54.8
Wang and Schmid, 2013	91.2	Wang and Schmid, 2013	57.2
Dense Trajectories	84.2	Dense Trajectories	47.0
<b>Our Approach (DT)</b>	<b>88.1</b>	<b>Our Approach (DT)</b>	<b>52.9</b>
<b>Our Approach (IDT)</b>	<b>92.1</b>	<b>Our Approach (IDT)</b>	<b>58.6</b>

Since the video dataset contains data for about 8 months, a new evaluation method is required to properly analyze the whole dataset and gain a deeper understanding of it. Our approach consists in displaying the data in a way that can be useful either for doctors, as well for engineers to detect failures and to improve the event recognition process in an efficient way. Because of this need, a new evaluation tool has been developed and named Event Plotter.

This tool provides a new method for event evaluation. First of all, as we do not have ground truth information for the total duration of the 8 months, but just for one week, so another method is needed to check the event model efficiency. To address this issue, the tool displays all the events in the desired time period (as clusters on a timeline) and single events (or time intervals) can be selected to quickly check the video and visualize the results of the event recognition working on the fly -see Figure 18-. The goal of this work is to understand how event recognition works, change the models on the fly, import them, and see how the recognition changes in real time.

Also, to compare new event models with the old ones, video summarization is implemented as well. Event based video summarization is utilized here to check how the recognition of one particular event type changes globally on the whole video and to display the recognition results. Also video summarization can be useful for doctors to check the way patients behave, for example playing all the videos of event "Get-up from bed", trying to predict patterns.

The data processed to address this issue was 1 week, because it was the time corresponding to the ground truth data. With these processed data, we have tested the efficiency of the Event Plotter tool, and we are currently improving the event recognition process by changing event models, adding new zones, and testing them on the fly.

### 7.13. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition

**Participants:** Carlos Fernando Crispim-Junior, François Brémond.

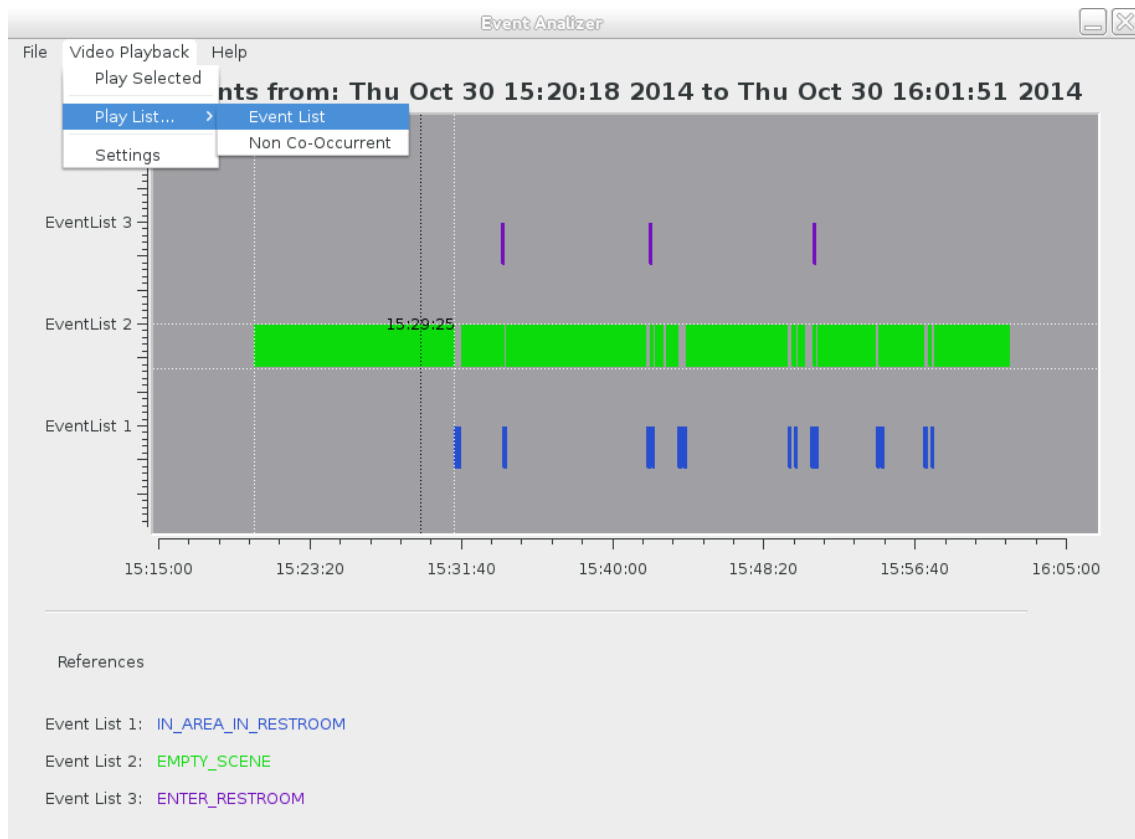


Figure 18. GUI of Event Plotter with 3 loaded Event Lists.



**Keywords:** Knowledge representation formalism and methods, Uncertainty and probabilistic reasoning, Concept synchronization, Activity recognition, Vision and scene understanding, Multimedia Perceptual System,

Combining multimodal concept streams from heterogeneous sensors is a problem superficially explored for activity recognition. Most studies explore simple sensors in nearly perfect conditions, where temporal synchronization is guaranteed. Sophisticated fusion schemes adopt problem-specific graphical representations of events that are generally deeply linked with their training data and focus on a single sensor. In this work we have proposed a hybrid framework between knowledge-driven and probabilistic-driven methods for event representation and recognition. It separates semantic modeling from raw sensor data by using an intermediate semantic representation, namely concepts. It introduces an algorithm for sensor alignment that uses concept similarity as a surrogate for the inaccurate temporal information of real life scenarios (Fig. 20). Finally, it proposes the combined use of an ontology language, to overcome the rigidity of previous approaches at model definition, and a probabilistic interpretation for ontological models, which equips the framework with a mechanism to handle noisy and ambiguous concept observations, an ability that most knowledge-driven methods lack (Fig. 19). We evaluate our contributions in multimodal recordings of elderly people carrying out instrumental activities of daily living (Table 11). Results demonstrated that the proposed framework outperforms baseline methods both in event recognition performance and in delimiting the temporal boundaries of event instances

This work has been developed as a collaboration between different teams in Dem@care consortium (Inria, University of Bordeaux, and CERTH). We thank the other co-authors for their contributions and support in the development of this work up to its submission for publication.

Table 11. Comparison to baseline methods in the test set

mean $F_1$ -score	Fusion approach		
	Baselines		Ours
IADL	SVM	OSF	
S. bus line	44.19	31.36	73.10
M.finances	43.99	0.00	43.73
P. pill box	45.86	49.11	65.02
P. drink	20.02	24.29	64.03
Read	90.18	91.82	95.22
T.telephone	72.12	0.00	75.58
W. TV	2.32	0.00	35.80
W. Plant	0.00	0.00	100.00
<b>Average</b>	39.83	24.57	69.06

OSF: Ontology-based Semantic Fusion

## 7.14. Semi-supervised Activity Recognition Using High-order Temporal-composite Patterns of Visual Concepts

**Participants:** Carlos Fernando Crispim-Junior, Michal Koperski, Serhan Cosar, François Brémond.

**Keywords:** visual concepts, semi-supervised activity recognition, complex activities, cooking composite activities

Methods for action recognition have evolved considerably over the past years and can now automatically learn and recognize short term actions with satisfactory accuracy. Nonetheless, the recognition of activities - a composition of actions and scene objects - is still an open problem due to the complex temporal, composite structure of this category of events. Existing methods either still focus on simple activities or oversimplify the modeling of complex activities by only targeting whole-part relations between activity components, like actions. In this work, we have investigated a hierarchical, semi-supervised approach that unsupervisedly learns actions from the composite patterns of atomic concepts (*e.g.*, slice, tomato), and complex activities from the

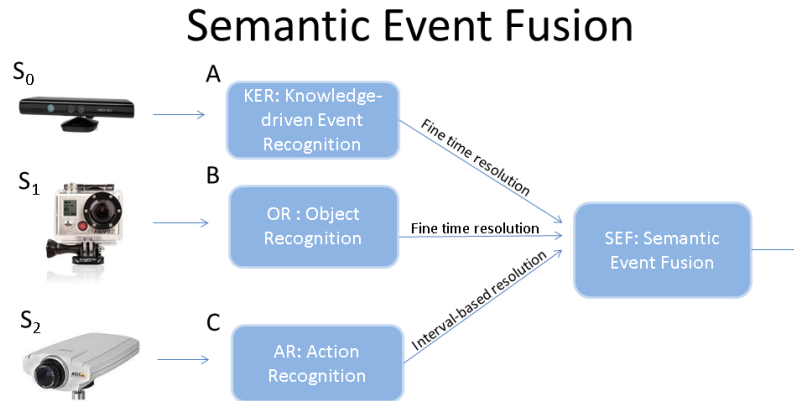


Figure 19. Semantic event fusion framework: detector modules (A-C) process data from their respective sensors ( $S_0$ - $S_2$ ) and output concepts (objects and low-level events). Semantic Event Fusion uses the ontological representation to initialize concepts to event models and then infer complex, composite activities. Concept fusion is performed on millisecond temporal resolution to cope with instantaneous errors of concept recognition.

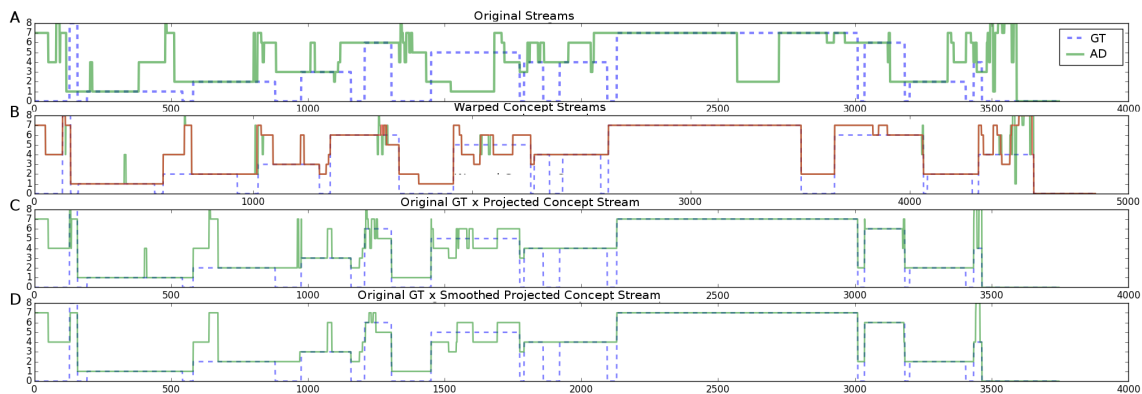


Figure 20. Semantic alignment between the concept stream of the action recognition detector (AR) and a concept stream (GT) generated from events manually annotated by domain experts using the time axis of the color-depth camera. X-axis denotes time in frames, and Y-axis denotes activity code (1-8), respectively, search bus line on the map, establish bank account balance, prepare pill box, prepare a drink, read, talk on the telephone, watch TV, and water the plant. From top to bottom, images denote: (A) original GT and AR streams, (B) GT and AR streams warped, AR stream warped and smoothed (in red), (C) original GT and AR stream warped and then backprojected onto GT temporal axis, (D) original GT and AR warped, backprojected, and then smoothed with median filtering.

temporal patterns of concept compositions (actions). On a first step, an unsupervised, inductive approach iteratively builds a multi-scale, temporal-composite model of the concept occurrences during the activity taking place (Fig. 22). Then, activity recognition is performed by comparing the similarity of the generated model of a given video and *a priori* learned and labeled unsupervised models. We have evaluated the proposed method in the MPII Cooking Composite Activities dataset (Fig. 21), a video collection where people perform a set of complex activities related to cooking recipes. To tackle this dataset it is necessary to recognize a large variety of visual concepts (*e.g.*, from actions, such as cutting and stirring, to objects, such as tomato and cutting board). Moreover, the detection of cooking activities is a very challenging problem since we observe a low inter-class variance between activity classes, and a high intra-class variance within an activity due to person to person differences in performing them. The proposed approach presents a mean average precision (mAP) of  $56.36\% \pm 5.1\%$ , and then outperforms previous methods ([81], mean AP 53.90%). This improvement is devoted to the modeling of deeper composite and temporal relations between visual concepts (from 2nd to 5th order compositions). The performance of the proposed method is mostly limited by the performance of low-level concept detectors. Future work will investigate ways to extend the current probabilistic model to handle more efficiently the differences in concept detector performance.



Figure 21. Illustration of one of the cooking recipes of Cooking Composite dataset [81]

## 7.15. From Activity Recognition to the Assessment of Seniors' Autonomy

**Participants:** Carlos Fernando Crispim-Junior, Carola Strumia, Alvaro Gomez Uria Covella, Alexandra Konig, François Brémond.

Activity recognition plays a fundamental role in several research fields as a way to extract semantic meaning from images and videos, to find more accurate matches for textual queries in video search engines, and to analyze long-term activity patterns in assisted living scenarios, such as seniors living at home. In this sense, we have continued our work on activity monitoring by proposing a novel knowledge-based event monitoring system that combines the observations of a vision system with expert knowledge and scene semantics, to recognize daily living activities in assisted living scenarios.

The approach's novelty lies in the combination of a flexible constraint-based ontology language for event modeling with efficient and robust algorithms to detect, track and re-identify people using color-depth sensing (low-level vision). The robust low-level vision promotes the modeling of longer and more complex events,

$$\begin{array}{l}
 \mathcal{Y}_{5,M}^* : \text{counter\_drawer\_hand\_knife\_take-out} \mathbf{after} \text{bowl\_cutting-board\_hand\_move\_plate} \\
 \vdots \\
 \mathcal{Y}_{5,1}^* : \text{cucumber\_cut\_cutting-board\_knife} \mathbf{before} \text{counter\_cutting-board\_drawer\_take-out} \\
 \vdots \\
 \mathcal{Y}_{3,1}^* : \text{cupboard\_hand\_plate} \mathbf{before} \text{hand\_plate\_take-out} \\
 \vdots \\
 \mathcal{Y}_{2,1}^* : \text{hand\_wash} \mathbf{before} \text{fridge\_hand}
 \end{array}$$

Figure 22. Example of temporal concept composite patterns of different arities extracted from an unlabeled activity.

while the ontology language provides a flexible way to describe event and incorporate domain knowledge, and ease knowledge transfer across different scenes. The proposed approach has been investigated for two assisted living scenarios: a) the monitoring of physical tasks and daily living activities in observation rooms of hospital and clinics, and b) daily and nightly activities of seniors living in nursing home apartments. To evaluate our approach performance compared to state of art methods, we have computed its results for GAADR dataset. This is public dataset, which is composed of videos of seniors performing physical tasks and activities of daily living. Evaluation results (Table 12) have demonstrated that our approach achieves an average  $F_1$  - score 20 % higher than the baseline method [89].

Table 12. Recognition of IADLs - GAADR data set - F 1 -score

Event	DT-HOG	DT-HOF	DT-MBHy	Proposed
Account Balance	44.96	34.71	42.98	66.67
Prepare Drink	81.66	44.87	52.00	100.00
Prepare Drug Box	14.19	0.00	0.00	57.14
Read Article	52.10	42.86	33.91	63.64
Talk on telephone	82.35	0.00	33.76	100.00
Turn on radio	85.71	42.52	58.16	94.74
Water Plant	0.00	0.00	0.00	52.63
<b>Average <math>\pm</math> SD</b>	<b>51.8 <math>\pm</math> 34.4</b>	<b>23.6 <math>\pm</math> 22.3</b>	<b>31.5 <math>\pm</math> 23.3</b>	<b>76.4 <math>\pm</math> 21.0</b>

Given the satisfactory performance of the proposed activity recognition framework we have also investigated it as a method to automatically measure a seniors' autonomy in quantitative and objective fashion. To do so, we have developed a probabilistic model that takes as input the recognized activities and gait-patterns from the period of time the person performs physical tasks. The proposed autonomy model has presented an average performance of 83.67 %, which suggests that the use of such technologies may provide clinicians with diagnostic relevant information, and decrease observer's biases when compared to clinical scales. The results of this investigation have been published in [33].

## 7.16. Serious Games Interfaces Using an RGB-D Camera : Results and Perspectives

**Participants:** Baptiste Fosty, François Brémond.

**Keywords:** RGB-D camera analysis, walking speed, serious games, startup project

Within the context of the development of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed algorithms to interact with the virtual environment through simple gesture recognition and walking speed computation. We have shown in previous work that the walking speed measured by our system is accurate enough within this context and reproducible. A paper has been submitted in Gait and Posture journal (now in reviewing process).

Concerning the gesture recognition algorithm, it consists in recognizing three basic gestures (right arm left on the side, left arm left on the side, right or left arm left on top). We performed a small experimentation to test the robustness of the system in detecting these gestures where participants (10 in total) had to perform 10 times each gesture while walking at 2.5km/h on the treadmill (see Figure 23). The results are shown in Table 13 .

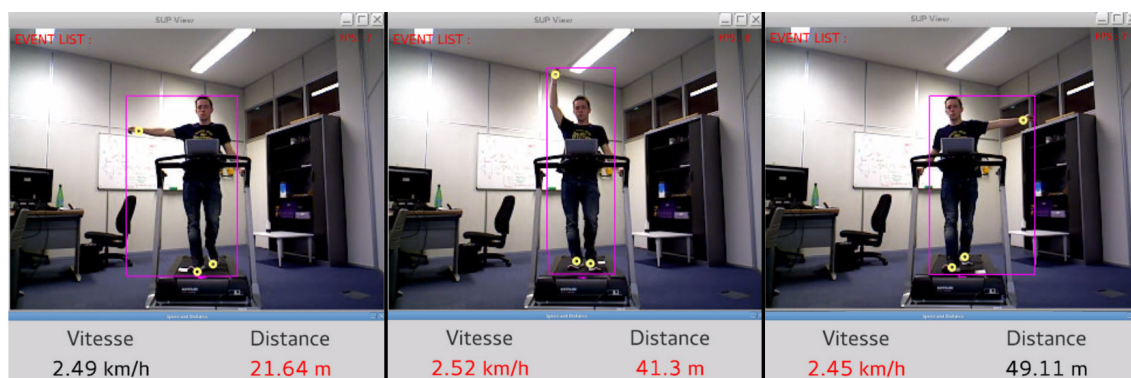


Figure 23. Display of gesture recognition and speed computation. Distance is red when right arm is left, speed is red when left arm is left, both are red when one arm is on top. Yellow dots correspond to feet and arm detection.

Table 13. Gesture recognition results.

	Right	Left	Top	TOTAL
Recall	97.0	95.0	100.0	97.3
Precision	100.0	97.9	94.3	97.3
<b>F-Score</b>	<b>98.5</b>	<b>96.4</b>	<b>97.1</b>	<b>97.3</b>

Following that, we decided to study whether this system would be useful in rehabilitation. Some experts of this domain in a rehabilitation center (Centre H elio Marin in Vallauris, France) have been interviewed and they were very enthusiasts about using this type of system to get objective gait parameters. To go further on the market opportunity and evaluate the feasibility of a technology transfer, we studied the concurrent products and contacted more than thirty other rehabilitation centers to have a deeper understanding of the needs and validate our idea. This investigation lead to the proposition of a startup project(BOMOTECH) to Inria which has been accepted and funded for the next 7 months, during what the goal is to get closer from a marketable tool.

## 7.17. Assistance for Older Adults in Serious Game Using an Interactive System

**Participants:** Minh Khue Phan Tran, Fran ois Br emond, Philippe Robert.

**Keywords:** Older Adults, Assistance, Serious Games

Serious Games offer a new way to older adults to improve their abilities such as vision, balance or memory. However, cognitive impairment causes a lot of difficulties to them when actively practising these games. Their engagement and motivation are reduced rapidly when encountering successive problems without any help. Our hypothesis is that this problematic situation can be handled if they are assisted regularly. We propose then an interactive system which can determine dynamically the situations and provide different helps in real-time. We focus on two main problems that the older players encounter regularly :

- they forget how to continue to play the game.
- they make a lot of errors.

The system determines the above problems by computing various characteristics of the player (skeleton, postures, gestures ...) along the game states. This process is presented in Figure 24. The characteristics of the player, which are collected by the Recognition Module thanks to Kinect Camera and the related SDK, are sent to the Interaction Module. This module associates these data with the game states provided by the game in order to recognize the problem and interacts with the player through a 3D-animated avatar.

The system is tested with 3 groups of patients described by 3 different cognitive states : mnesic plaits, MCI and Alzheimer. The patients are invited to play a concentration-based game with a Kinect camera. Each patient plays 3 phases : playing with therapist, playing alone and playing with the avatar. The playing time and the final score of each game phase are recorded. Here, the system takes into account the player's gestures and the game states for recognizing two situations :

- the player reacts too late and too slowly to the current game task.
- the player makes many mistakes.

The experimental results confirmed our hypothesis. Most of the patients have the best performance in phase "playing with the avatar". Their playing time is shorter and their final score is higher in phase "playing with assistance" than in phase "playing alone". The results are presented in the publication [36] accepted by the Games and Learning Alliance Conference in December 2015. Future work aims at improving the system and compare its efficiency with the one of "humans assistances".

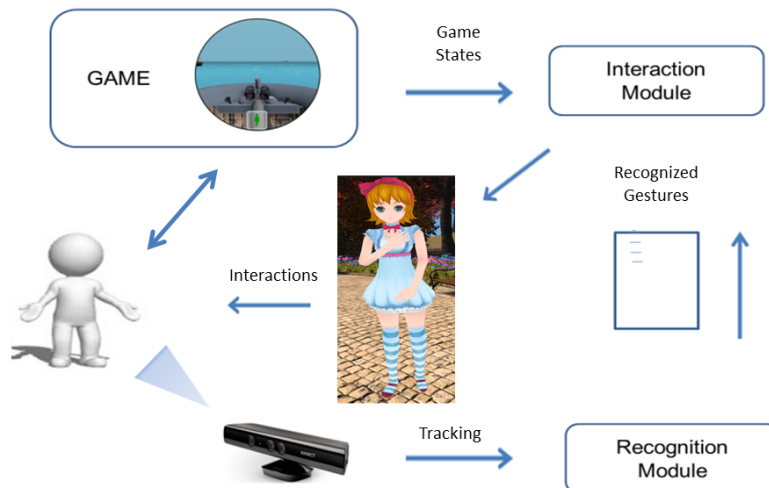


Figure 24. Assisting older adults in serious game playing

## 7.18. Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

**Participants:** Farhood Negin, Serhan Coşar, Michal Koperski, François Brémond.

**Keywords:** Unsupervised Activity Recognition

### Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

In this work, we propose an unsupervised approach that offers a comprehensive representation of activities by modeling both global and body motion of people. Compared to existing supervised approaches, our approach automatically learns and recognizes activities in videos without user interaction. First, the system learns important regions in the scene by clustering trajectory points. Then, a sequence of primitive events is constructed by checking whether people are inside a region or moving between regions. This enables to represent the global movement of people and automatically split the video into clips. After that, using action descriptors [90], we represent the actions occurring inside each region. Combining action descriptors with global motion statistics of primitive events, such as time duration, an activity model that represents both global and local action information is constructed. Since the video is automatically clipped, our approach performs online recognition of activities. The contributions of this work are twofolds: (i) generating unsupervised human activity models that obtains a comprehensive representation by combining global and body motion information, (ii) recognizing activities online and without requiring user interaction. Experimental results show that our approach increases the level of accuracy compared to existing approaches. Figure 25 illustrates the flow of the system.

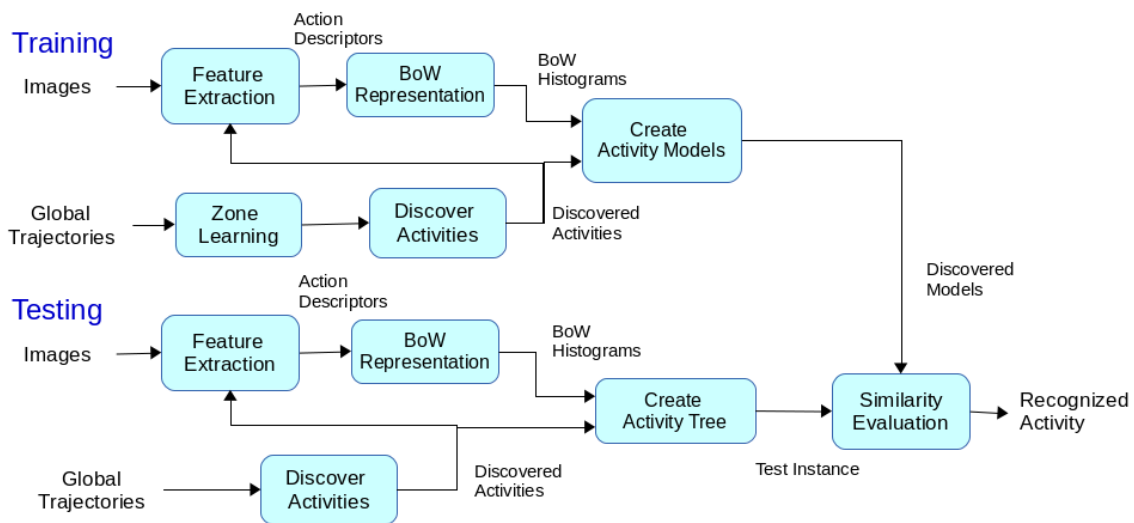


Figure 25. Architecture of the framework: Training and Testing phases

The performance of the proposed approach has been tested on the public GAARDR dataset [67] and CHU dataset (<http://www.demcare.eu/results/datasets>) that are recorded under EU FP7 Dem@Care Project1 in a clinic in Thessaloniki, Greece and in Nice, France, respectively. The datasets contain people performing everyday activities in a hospital room. The activities considered in the datasets are listed in Table 1 and Table 2. Each person is recorded using RGBD camera of 640x480 pixels of resolution. The GAARDR dataset contains 25 videos and the CHU dataset contains 27 videos. Each video lasts approximately 10-15 minutes.

We have compared our approach with the results of the supervised approach in [90]. We did also a comparison with an online supervised approach that follows [90]. For doing this, we train the classifier on clipped videos and perform the testing using sliding window. In the online approach, a SVM is trained using the action descriptors extracted from groundtruth intervals. We have also tested different versions of our approach that i) only uses global motion features and ii) which only uses body motion features. We have randomly selected 3/5 of the videos in both datasets for learning the activity models. The codebook size is set to 4000 visual words for all the methods.

The performance of the online supervised approach and our approach in GAARDR dataset are presented in Table 1. In all approaches that use body motion features, HoG descriptors are selected since they give the best results. It can be clearly seen that, using models that represent both global and body motion features, our unsupervised approach enables to obtain high sensitivity and precision rates. Compared to the online version of [90], thanks to the learned zones from positions and discovered activities, we obtain better activity localization, thereby better precision. However, since the online version of [90] utilizes only dense trajectories (not global motion), it fails to localize activities. Hence, it detects the intervals that does not include an activity (e.g. walking from radio desk to phone desk) and for "prepare drug box", "watering plant", and "reading" activities, it cannot detect the correct intervals of the activities. Compared to the unsupervised approach that either use global motion features or body motion features, we can see that, by combining both features, our approach achieves more discriminative and precise models, thereby improves both sensitivity and precision rates. By combining global and body motion features, our approach benefits from discriminative properties of both feature types. Table 1 also presents the results of the supervised approach in [90]. Although the supervised approach uses groundtruth intervals in test videos in an offline recognition scheme, it fails to achieve accurate recognition. As our approach learns the zones of activities, we discover the places where the activities occur, thereby we achieve precise and accurate recognition results. Since this information is missing in the supervised approach, it detects "turning on radio" while the person is inside drink zone preparing drink.

Table 2 shows the results of the online supervised approach and our approach in CHU dataset. MBH descriptor along y axis and HoG descriptor gives the best results for our approach and the online supervised approach, respectively. In this dataset, since people tend to perform activities in different places (e.g. preparing drink at phone desk), it is not easy to obtain high precision rates. However, compared to the online version of [90], our approach detects all activities and achieves a much better precision rate. The online version of [90] again fails to detect activities accurately, thereby misses some of the "preparing drink" and "reading" activities and gives many false positives for all activities.

Thanks to the activity models learned in unsupervised way, we accurately perform online recognition. In addition, the zones learned in an unsupervised way help to model activities accurately, thereby most of the times our approach achieves more accurate recognition compared to supervised approaches. This paper has been published in third Asian Conference on Pattern Recognition (ACPR 2015) [35].

## 7.19. Run-time Adaptation of Video Systems

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

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on run-time reconfiguration of such systems. The goal is to follow the "model at run-time" approach and to obtain context-aware self-adaptive video systems. In this approach models are kept and used at run-time. In our case, these models describe all the possible run-time configurations. They are specified using *Feature Models*.

Run-time reconfiguration means to react to context changes by tuning, adding, removing, or replacing components of the video chain, and possibly changing the chain itself.

So far, we have defined a run-time architecture consisting of three layers. The lower level describes the video analysis components and the context events. The upper one handles feature model adaptation. The middle layer is an adapter: it transforms lower level context event occurrences into upper level feature reconfiguration; in the other direction, it transforms the corresponding feature reconfigurations into video components reconfigurations.



Table 14. The activity recognition results for GAADR dataset. Bold values represent the best sensitivity and precision results for each class.

ADLs	Supervised Approach [90]		Online Version of [90]		Unsupervised (Only Global Motion)		Unsupervised (Only Body Motion)		Proposed Approach	
	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	<b>100</b>	88	<b>100</b>	70	<b>100</b>	<b>100</b>	57	<b>100</b>	<b>100</b>	<b>100</b>
Establish Acc. Bal.	67	<b>100</b>	<b>100</b>	29	<b>100</b>	86	50	<b>100</b>	<b>100</b>	86.67
Preparing Drink	<b>100</b>	69	<b>100</b>	69	78	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
Prepare Drug Box	<u>58.33</u>	<b>100</b>	11	20	33.34	<b>100</b>	33.34	<b>100</b>	33.34	<b>100</b>
Watering Plant	<u>54.54</u>	<b>100</b>	0	0	44.45	57	33	<b>100</b>	44.45	<b>100</b>
Reading	<b>100</b>	<b>100</b>	88	37	<b>100</b>	<b>100</b>	38	<b>100</b>	<b>100</b>	<b>100</b>
Turn On Radio	60	86	<u>100</u>	75	89	89	44	<b>100</b>	89	<b>100</b>
AVERAGE	77.12	91.85	71.29	42.86	77.71	90.29	50.71	<b>100</b>	<b>80.97</b>	98.10

Table 15. The activity recognition results for CHU dataset. Bold values represent the best sensitivity and precision results for each class.

ADLs	Supervised Approach [90]		Online Version of [90]		Proposed Approach	
	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	57	78	<b>100</b>	<b>86</b>	<b>100</b>	65
Preparing Drink	78	<b>73</b>	92	43	<b>100</b>	58
Prepare Drug Box	<b>100</b>	83	<b>100</b>	43	<b>100</b>	<b>100</b>
Reading	35	<b>100</b>	92	36	<b>100</b>	<b>78</b>
Using Bus Map	90	<b>90</b>	<b>100</b>	50	<b>100</b>	47
AVERAGE	72.0	<b>84.80</b>	90.95	48.76	<b>100</b>	70.00

This year we focused on the upper layer. We first formalized the run-time feature reconfiguration rules. First, any reconfiguration should respect the semantics of feature models and their attached constraints. Second, the reconfiguration should satisfy the requests from the middle layer, essentially selections and deselections of features. From these two requirements, we identified three possible outcomes: successful reconfiguration, impossible reconfiguration (selection/deselection conflicts), and “undefined” reconfiguration (not enough information to get through the process). We also determined the actions to take in these cases. In particular, in the last two cases, we decided to let the component configuration unchanged.

To implement this upper layer, we first attempted to rely on an existing feature model manipulation framework, namely FAMILIAR [45]. However, this approach suffers from a number of drawbacks. First, FAMILIAR is a standalone Java program, whereas the rest of the system is written in C++, for performance reasons and library availability. Hence, using FAMILIAR implies superfluous back and forth inter-module communications and data transformations. Second, we confirmed that FAMILIAR is more a system deployment tool than a run-time reconfiguration one. In particular it cannot fulfill all the reconfiguration rules that we have formalized. Therefore, we are completing a full re-implementation of the upper layer.

The programming language homogeneity permits a more efficient integration of the three layers. In particular, it becomes easier to incorporate our extensions to feature models such as quality metrics [34].

## 7.20. Scenario Description Language

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

Last year, we developed a scenario recognition engine based on the Synchronous Model of reactive systems. We now need a scenario description language friendly to our end users who are not computer scientists in general. In fact, Stars has already defined such a language. However, it is a declarative language based on (temporal) constraints. This is certainly not the most natural and the simplest way for end users to express their domain specific scenarios.

Consequently, we started this year a comparative study of different means to express scenarios in various domains (video understanding but also games, movies, music, criminology, military strategy...). We investigated 16 formalisms covering these domains. We defined a comparison grid based on criteria relevant for our video understanding goals. We retained 9 such criteria: application domain scope, ease of use, representation of scenario basic elements (background, scene, roles...), modularity (possibility of scenario hierarchy), time representation (absolute, logical, multi-clocks, no clocks...), expression of temporal constraints, representation of repetitive patterns, support for concurrency and parallelism, and finally formal foundations.

To complete the study, we conducted an experiment: describing a case study scenario using some of these formalisms to concretely estimate their advantages and drawbacks, especially their ease of use.

At this time, none of the studied languages fulfills completely our needs. Many languages are graphical ones. While this may appear as user friendly, scalability and automatic analysis become an issue. Some languages lack formal semantics, which is not acceptable in our case; others are merely extensions of computer languages, hence dedicated to specialists.

We plan to define our own version, which will rely on solid semantic foundations. (see section 7.21). To enforce user-friendliness, we started to collaborate with ergonomists.

## 7.21. Scenario Recognition

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

**Keywords:** Synchronous Modeling, Model checking, Mealy machine, Cognitive systems.

For a long time, Stars strategy has been to favor the easy generation of activity recognition systems. These systems correspond to a succession of pattern matching and clustering algorithms, combined with adequate knowledge representation (e.g. scene topology, temporal constraints) at different abstraction levels (from raw signal to semantics). Due to the large range of application domains (surveillance, safety, health care, ...), we propose a generic approach to design activity recognition engines. Moreover, such domains require high dependability due to possible safety issues. Thus, our approach should also rely on formal methods to describe, analyze, verify, and generate effective recognition engines. We consider activity recognition engines as reactive systems that react to input events from their environment and produce output events in the form of alarms or notifications. Such engines are intrinsically real time, reactive and they evolve in discrete time. As a consequence, to recognize scenarios, we adapt the usual techniques of synchronous modeling approach to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario.

Our previous developments, on top of existing synchronous languages as Lustre and LE (see section 7.22), were convenient for rapid prototyping. However, even if LE is not a closed environment, it appeared as difficult as Lustre to customize, for efficiency reasons. This year, in the framework of Ines Sarray PhD thesis, we began to define a synchronous semantics for the future scenario language (see section 7.20). The idea is to generate automatically recognition engines at compilation time. The compilation itself is totally based on the semantics. To complete this approach we will rely on both our experiment with the LE language last year and on the LE compilation process.

## 7.22. The Clem Workflow

**Participants:** Annie Ressousche, Daniel Gaffé, Imane Khalis.

**Keywords:** Synchronous languages, Synchronous Modeling, 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 26) 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 continued to focus on the improvement of both LE language and compiler concerning data handling and the generation of back-ends, required by other research axis of the team (see 7.21 and 7.23). We also improved the design of a new simulator for LE programs which integrates our new approach. 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. Last year, we begun to develop such a simulator in order to integrate the data part of the language. The simulator GUI has been designed in Qt and the simulator takes into account the values carried by signals. This year, during her internship, Imane Khalis has completed the simulator to allow an external computation of data values and a communication with the simulator through a socket mechanism. With this last development, the LEC simulator is complete and is integrated in the CLEM toolkit.

## 7.23. Safe Composition in WCOMP Middleware for Internet of Things

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

**Keywords:** Synchronous Modeling, Ubiquitous Computing, middleware, internet of things

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.7) with a service-oriented middleware approach dealing with dynamic evolutions of system infrastructure in ubiquitous computing, and particularly in the Internet of Things (IoT). The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [86], [85], [66]) to dynamically adapt and recompose assemblies of components.

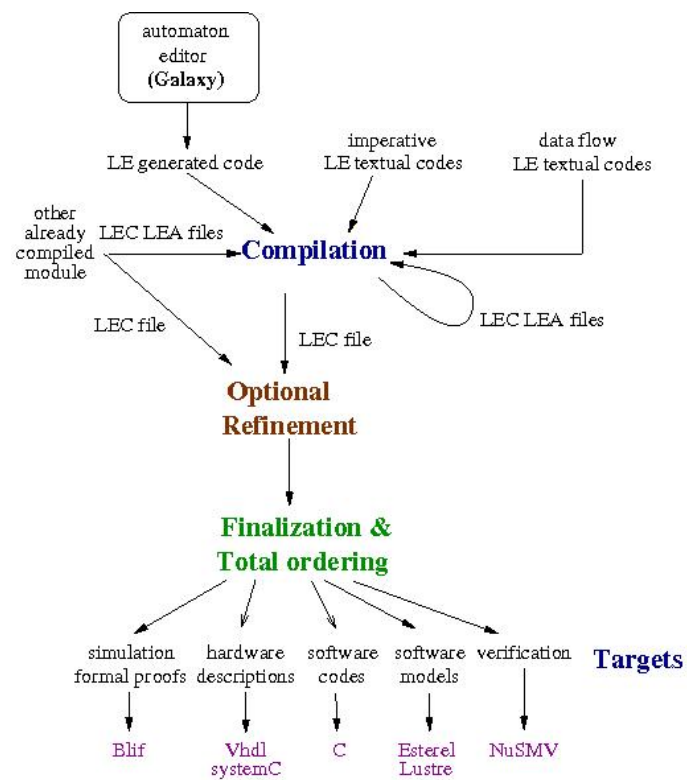


Figure 26. The Clem Toolkit

IoT is a way to combine computation and communication capabilities, sometimes in large scale information systems, with a huge number of complex devices connected to the physical world. Such infrastructures are often dedicated to the deployment of multiple applications, running concurrently. These applications are using shared devices from a common environment through different network middleware and numerous IoT protocols. Indeed, “Things”, also called the *Entities of Interest* [63], are the part of the real world in which devices are interacting and which must not be neglected. We aimed to model and validate concurrent accesses to shared devices without neglecting their associated *Entity of Interest*, their common physical context. One of the main challenge is then how to guarantee and validate some safety and integrity properties throughout the system’s evolution. In WComp middleware, we use synchronous models to facilitate the study and the validation of new composition mechanisms between applications at runtime. Then key problems to solve are: (1) how to specify and respect the "Thing" behavior? (2) how to ensure a safe combination of these multiple accesses when several services accesses a same entry of an *Entity of Interest* ? (3) how to manage multiple uses when applications simultaneously use a same service ?

This year, we addressed these problems by relying on formal method to model device behaviors as synchronous automata, taking into consideration their impact on the Entity Of Interest. Such an approach allows applying model-checking techniques to verify safety properties of applications. The main contribution is the definition of a sound way to compose models allowing context change adaptation. This composition relies on synchronous parallel composition paradigm. We proved that this operation preserves safety properties. However, it is not sufficient to obtain a global model of this composition because some devices may interact with the same Entity Of Interest. Moreover, several applications may use the same device services and then they can have concurrent accesses to their entries, so it can have an unexpected impact on our Entity Of Interest. Therefore, we added constraints to the device models composition and to applications level. We defined a generic way to express these constraints, independently of the knowledge about the devices and the applications, only their type is sufficient. We proposed the Description Constraint Language (DCL) to express these generic constraints and their compilation into LE Mealy machines. Thus we rely on CLEM model-checking facilities (see section 7.22) to validate the constraints. As a consequence, this approach ensures the adaptation to a context change and offers a means to formally perform validation.

These results have been published in [43]

## 7.24. Design of UHD Panoramic Video Camera

**Participants:** Carolina Da Silva Gomes Crispim, Rachid Guerchouche, Daniel Gaffé, François Brémond.

The goal of this work is to investigate the possibilities of designing a new camera-based system for retail. This work was carried in the context of a collaboration between STARS and Neosensys. The system is composed of several high definition cameras placed in a configuration such as it makes it possible to obtain a panoramic vision with 360° of field of view. The work was divided into 2 parts: theoretical part and practical part.

In the theoretical part, the different characteristics of the desired system were studied, such as: number of cameras, resolution of each camera, the different characteristics of the sensors (WDR, HDR, exposure) etc. Depending on these characteristics, data transmission through an IP network was addressed. In addition to the hardware characteristics, the possibility of embedding stitching capabilities was studied. After spending some time understanding the background behind such techniques, an existing implementation of the stitching was adopted. Simulations were then made in order to estimate the characteristics of an FPGA capable of handling 5 cameras with 12MP resolution each. An existing FPGA architecture extensively used in the industry was chosen and a mathematical model was provided in order to estimate the characteristics of such FPGA according to the different parameters of the camera-based system.

In the practical part, an implementation of the two first steps of the stitching algorithm (homograph estimation and warping) was performed on a FPGA using 2 cameras. The problems of code optimisation were addressed in order to achieve a functioning implementation with respect to the memory and computation capabilities of the FPGA.

## 7.25. Brick & Mortar Cookies

**Participant:** Anaïs Ducoffe.

The objective of the BMC project is to create a software that aims to present attendance and attractiveness of the customer in stores, based on automatic video analysis. This final system is designed to be used without changing current camera network of the customer store, dedicated to security purpose. Analysis should be given at different time and space resolutions. For instance, day attendance can be as interesting as year attendance. Moreover, shop owners want to be able to compare two given years or months, etc... As space resolution is concerned, the software should be able to give information about the global attractiveness of the store but should also analyze some specific zones.

### **IVA embedded on Bosch cameras**

Intelligence Video Analysis (IVA) is embedded in some models of Bosch cameras. The algorithms are composed of human detection and tracking. They can be configured directly on the camera interface via *tasks*. Following Bosch *tasks* were selected and studied:

- **Loitering and idle object in a field tasks** enable to detect stop actions in a zone, when they happened and the stop positions.
- **Entering and leaving field tasks** enable to know when a person enters or leaves a zone.
- **Detect people in a field task** enables counting people in a zone.
- **Crossing lines tasks** for counting people entering or leaving shop. We are able to know when the line was crossed and in what sense.

It is not possible to get people trajectory when metadata from Bosch cameras are acquired in offline mode. Then we studied live connection to get metadata directly from the camera stream using a RTSP connection. Metadata information is saved in XML format.

The previously enumerated tasks use algorithms to detect people and get their trajectories. STARS team has developed similar algorithms and has adapted their parameters values to the specific needs of this software. Moreover these algorithms can be run on any type of video cameras (live and offline modes) whereas Bosch IVA can only be run in live mode on compatible Bosch cameras. Stars algorithms can also combine several cameras at the same time in order to track people across the camera network. We need those algorithms to sell a system that doesn't need a new camera network but can be used with existing ones. They will be integrated in the final product.

### **Tests in real conditions**

A system for testing cameras and our software was installed in partner store (Super U). Cameras were installed and configured to process all our use cases and test our mechanism to extract the metadata. We used only Bosch camera with embedded IVA. We successfully acquire 2 hours of the desired metadata. The results of embedded algorithms are reliable on realistic data : we get good results in counting people and trajectories are accurate.

### **Metadata storage in database**

Metadata have to be stored in a hierarchical way as request of the metadata by the application should be easy and quick. We choose to store metadata in a database. This database design was constraint by data storage speed and a quick access for live computation. Different parts of database (store information, devices description etc...) were designed to be as much independent as possible.

### **Web interface (GUI)**

The graphic interface design is in progress. The interface will be a web based one to narrow compatibility problems: the application should be used as well with a computer as a tablet.

## 7.26. Monitoring Older People Experiments

**Participants:** Matias Marin, Etienne Corvée, François Brémond.

This year we have conducted many experiments, especially in Nice and partially in Thessaloniki, Lulea, Taiwan and Dublin, to validate our studies on monitoring older people suffering from various behavioral disorders in the framework of several projects.

### **DEM@CARE PROJECT**

For the project Dem@care (see section 9.2.1.3), we use PCs with ASUS cameras, for monitoring and collecting a video dataset associated with metadata<sup>1</sup>. The software CAR is installed to automatically annotate the videos<sup>2</sup>. Data is recorded locally, and backups are made automatically and remotely: one on the server (LAB at the nursery home) and another backup at Inria. These data can be reached locally at the nursery home, thanks to the server located in the lab; also, they are all accessible from Inria network by ssh.

### **SafEE PROJECT**

SafEE project (see section 9.1.1.2) experiments in the nursing home have started at the end of 2015, and are made up with different technologies (wifi, wired network, smart phones, Kinect, RFID, tablet...). In the nursing room, 2 PCs with KINECT2 are connected for monitoring the residents and are stored in a database. Another PC with Windows software is configured for SafEE serious game (cognitive games and music box, 1.7 version). Moreover, a Wifi access point will be used by medical staff at the nursery home to connect to a Graphical User Interface through a website designed by INDES team<sup>3</sup>, to consult patient data (their daily activities), from the activity history stored in the database or in real-time. Another device called AromaCare is installed in the rooms, which is a connected aroma diffuser by Wi-Fi. With the app Aroma Therapeutics (smartphone or tablet) we can manage several diffusers, by scheduling different programs each day and change the intensity of the diffusion.

In patients' home the same devices than Nursing Home have to be configured. Today, only recordings are done and stored at Inria.

### **OTHER PROJECTS at ICP (Institut Claude Pompidou)**

ICP has now a remote access by using rdesktop, which is safer than team-viewer screen sharing session. The installation of new experimentations (e.g. praxis, relaxation, serious games) is now in progress: the configuration expected includes PCs with KINECT2 connected at ICP network and accessible from Inria. Some experimentations will use wireless sensors (e.g. accelerometer, pressure), controlled by the app wireless tag (on smartphone or tablet) to measure fine patient activities: motion, kettle utilization, etc.

## **8. Bilateral Contracts and Grants with Industry**

### **8.1. 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 older adults.
- **LinkCareServices:** this project with Link Care Services runs from 2010 upto 2015. It aims at designing a novel system for Fall Detection. This study consists in evaluating the performance of video-based systems for Fall Detection in a large variety of situations. Another goal is to design a novel approach based on RGBD sensors with very low rate of false alarms.

<sup>1</sup><https://team.inria.fr/stars/demcare-chu-dataset/>

<sup>2</sup><https://team.inria.fr/stars/software/car-complex-activity-recognition-component-installation/>

<sup>3</sup><http://webrobotics.inria.fr:8081/hop/events>

## 9. Partnerships and Cooperations

### 9.1. National Initiatives

#### 9.1.1. ANR

##### 9.1.1.1. MOVEMENT

Program: ANR CSOSG

Project acronym: MOVEMENT

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

Duration: January 2014-June 2017

Coordinator: MORPHO (FR)

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

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

##### 9.1.1.2. SafEE

Program: ANR TESCAN

Project acronym: SafEE

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

Duration: December 2013-May 2017

Coordinator: CHU Nice

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

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

#### 9.1.2. Investment of Future

##### 9.1.2.1. Az@GAME

Program: DGCIS

Project acronym: Az@GAME

Project title: Medical diagnosis aid tool for Alzheimer disease and similar pathologies (un outil d'aide au diagnostic médical sur l'évolution de la maladie d'Alzheimer et les pathologies assimilées).

Duration: January 2012- December 2015



Coordinator: Groupe Genious

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

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

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

### **9.1.3. FUI**

#### *9.1.3.1. Visionum*

Program: FUI

Project acronym: Visionum

Project title: Visonium.

Duration: January 2015- December 2018

Coordinator: Groupe Genious

Other partners: Inria(Stars), StreetLab, Fondation Ophtalmologique Rothschild, Fondation Hospitaliere Sainte-Marie.

Abstract: This French project from Industry Minister aims at designing a platform to re-educate at home people with visual impairment.

## **9.2. European Initiatives**

### **9.2.1. FP7 & H2020 Projects**

#### *9.2.1.1. CENTAUR*

Title: Crowded ENvironments moniToring for Activity Understanding and Recognition

Programm: FP7

Duration: January 2013 - December 2016

Coordinator: Honeywell

Partners:

Computer Vision Laboratory, Ecole Polytechnique Federale de Lausanne (Switzerland)

honeywell, Spol. S.R.O (Czech Republic)

Data Centric Technologies Group, Neovision Sro (Czech Republic)

Centre for Intelligent Sensing, Queen Mary University of London (United Kingdom)

Inria contact: François Brémond

We aim to develop 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. Research activities will be driven by specific sets of scenarios, requirements and datasets that reflect security operators' needs for guaranteeing the safety of EU citizens. CENTAUR gives a unique opportunity to academia to be exposed to real life dataset, while enabling the validation of state-of-the-art video surveillance methodology developed at academia on data that illustrate real operational scenarios. The research agenda is motivated by ongoing advanced research activities in the participating entities. With Honeywell as a multi-industry partner, with security technologies developed and deployed in both its Automation and Control Solutions and Aerospace businesses, we have multiple global channels to exploit the developed technologies. With Neovision as a SME, we address small fast paced local markets, where the quick assimilation of new technologies is crucial. Three thrusts identified will enable the monitoring of crowded scenes, each led by an academic partner in collaboration with

scientists from Honeywell: (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. We expect a long term impact on the field of video surveillance by: contributions to the state-of-the-art in the field, dissemination of results within the scientific and practitioners community, and establishing long term scientific exchanges between academia and industry, for a forum of scientific and industrial partners to collaborate on addressing technical challenges faced by scientists and the industry.

#### 9.2.1.2. PANORAMA

Title: Ultra Wide Context Aware Imaging

Programm: FP7

Duration: April 2012 - March 2015

Coordinator: Philips

Inria contact: François Brémond

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 setup 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.

#### 9.2.1.3. DEM@CARE

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

Type: FP7

Defi: Cognitive Systems and Robotics

Instrument: Industry-Academia Partnerships and Pathway

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

Duration: November 2011-November 2015

Coordinator: Centre for Research and Technology Hellas (G)

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

Inria contact: François Brémond

Abstract: The objective of Dem@Care is the development of a complete system providing personal health services to persons with dementia, as well as medical professionals, by using a multitude of sensors, for context-aware, multiparametric monitoring of lifestyle, ambient environment, and health parameters. Multisensor data analysis, combined with intelligent decision making mechanisms, will allow an accurate representation of the person's current status and will provide the appropriate feedback, both to the person and the associated medical professionals. Multi-parametric monitoring of daily activities, lifestyle, behavior, 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.

## 9.3. International Initiatives

### 9.3.1. Inria International Labs

#### 9.3.1.1. Informal International Partners

##### 9.3.1.1.1. Collaborations with Asia:

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

##### 9.3.1.1.2. Collaboration with U.S.A.:

Stars collaborates with the University of Southern California.

##### 9.3.1.1.3. Collaboration with Europe:

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

#### 9.3.1.2. Other IIL projects

The ANR SafEE (see section 9.1.1.2) collaborates with international partners such as Taichung Veterans General Hospital TVGH (TW), NCKU Hospital(TW), SMILE Lab at National Cheng Kung University NCKU (TW) and BDE (TW).

## 9.4. International Research Visitors

### 9.4.1. Visits of International Scientists

This year, Stars has been visited by the following international scientists:

- Salwa Baabou, Ecole Nationale d'Ingénieurs de Gabès, Tunisia;
- Siyuan Chen, University of New South Wales, Australia;
- Jesse Hoey, University of Waterloo, Canada;
- Adlen Kerboua, University of Skikda, Algeria;
- Caroala Strumia, University of Genova, Italy.

#### 9.4.1.1. Internships

Ujjwal Ujjwal

Date: June 2015 - Nov 2015

Institution: International Institute of Information, Hyderabad (India)

Supervisor: François Brémond

Ghada Bahloul

Date: Jul 2015 - Sept 2015

Institution: Ecole Polytechnique de Sousse (Tunisia)

Supervisor: Rachid Guerchouche

Kanishka Nithin Dhandapani

Date: June 2015 - Nov 2015

Institution: IIT Madras (India)

Supervisor: Carlos Fernando Crispim Junior

Ramiro Leandro Diaz

Date: Jul 2015

Institution: UNICEN, Buenos Aires, Argentina

Supervisor: Carlos Fernando Crispim Junior

Alvaro Gomez Uria Covella

Date: Jan 2015 - Apr 2015

Institution: Universidad Nacionale de Rosario, Argentina

Supervisor: Carlos Fernando Crispim Junior

### **9.4.2. Visits to International Teams**

#### *9.4.2.1. Research stays abroad*

Piotr Bilinski

- Date: Apr 2015- Aug 2015
- Institution: Honeywell, Spol. S.R.O (Czech Republic)

## **10. Dissemination**

### **10.1. Promoting Scientific Activities**

#### **10.1.1. Scientific events organisation**

##### *10.1.1.1. General chair, scientific chair*

François Brémond was organizer of a PANORAMA (see section 9.2.1.2) workshop, part of SPIE Electronic Imaging (EI) event (8-12 February 2015, San Francisco).

François Brémond was editor of a PANORAMA (see section 9.2.1.2) special issue of Journal of Electronic Imaging Letters (JEI): Ultra Wide Context and Content Aware Imaging.

##### *10.1.1.2. Member of the organizing committee*

François Brémond was a member of the Management Committee and COST Action IC1307 in 2015.

#### **10.1.2. Scientific events selection**

##### *10.1.2.1. Member of the conference program committees*

François Brémond was program committee member of the conferences and workshops: AAI-15, GROW2015, ICDP2015, KSE 2015, PETS2015, PR4MCA and ICVS15.

François Brémond was publication chair and session chair of AVSS-15.

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

##### *10.1.2.2. Reviewer*

François Brémond was reviewer for the conferences : CVPR2015-6, ECCV2015, VOT2015, Grets2015, ICIP2015, MSF2015, ChaLearn2015, ICCV2015, WACV 2016.

#### **10.1.3. Journal**

##### *10.1.3.1. Member of the editorial boards*

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

##### *10.1.3.2. Reviewer - Reviewing activities*

François Brémond was reviewer for the journal *Frontiers in Human Neuroscience*.

#### **10.1.4. Invited talks**

François Brémond was invited to give a talk at Bristol University, SPHERE Seminar, 25 June 2015.

François Brémond was invited to give a tutorial at Leuven University at the 1st iV&L Net Training School, 3 June 2015.

François Brémond has participated to the Workshop Innovation Alzheimer, 12 November, 2015.

Antitza Dantcheva gave an invited talk at SBA, Vienna, Austria, July 2015.

Antitza Dantcheva gave an invited talk at the University of Cyprus, Nicosia, Cyprus, April 2015.

Antitza Dantcheva presented a demo at the Workshop Innovation Alzheimer (IA 2015), November 2015.

### **10.1.5. Scientific expertise**

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 Brussels in 15 September 2015.

François Brémond was expert for the Foundation Médéric Alzheimer, for the doctoral fellowship selection, October 2015.

François Brémond was expert for the Sophia Labex Network, for the doctoral fellowship selection, October 2015.

François Brémond was expert for Inria Sophia Working Group, for Inria Team Proposals.

François Brémond was expert for reviewing Computer Science Discovery Grant application for the Natural Sciences and Engineering Research Council of Canada (NSERC).

François Brémond was expert for reviewing Consolidator Grant proposal for European Research Council (ERC).

François Brémond was expert for the Foundation Médéric Alzheimer, for the Alzheimer's Outlook, Paris, 11 December 2015.

## **10.2. Teaching - Supervision - Juries**

### **10.2.1. Teaching**

Master : Annie Ressouche, Safety in Middleware for Internet of Things, 10h, niveau (M2), Polytech Nice School of Nice University.

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.

### **10.2.2. Supervision**

PhD : Julien Badie, Optimizing process for tracking people in video-camera network, 17-11-2015, François Brémond.

PhD in progress : Auriane Gros, Evaluation and Specific Management of Emotionnal Disturbances with Activity Recognition Systems for Alzheimer patient, Sept 2014, François Brémond.

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

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

PhD in progress : Thi Lan Anh Nguyen, Complex Activity Recognition from 3D sensors, Dec 2014, François Brémond.

PhD in progress : Ines Sarray, Activity Recognition System Design, Oct 2014, Sabine Moisan.

PhD in progress : Farood Negin, People detection for activity recognition using RGB-Depth sensors, Jan 2015, François Brémond.

### 10.2.3. Juries

François Brémont was jury member of the following PhD theses:

PhD, Dana Codreanu, Toulouse University, 21 May 2015.

PhD, Matthieu Rogez, Lyon University, LIRIS, Foxstream, 9 June 2015.

PhD, Cédric Le Barz, Pierre and Marie Curie University, Thales, 30 June 2015.

PhD, Owais Mehmood, School of centrale de Lille, IFSTTAR, 28th September 2015.

PhD, Geoffroy Cormier, Rennes University, ECAM Rennes Louis de Broglie, NeoTec-Vision, 10 Novembre 2014.

PhD Thi Khanh Hong Nguyen, Nice University - LEAT, 18 November 2015.

PhD Cyrille Beaudry, La Rochelle University, MIA, 26 November 2015.

PhD Arsène Fansi Tchango, Lorraine University, 4 December 2015.

PhD, Andrei Stoian, CNAM, Paris, 15 January 2016.

## 10.3. Popularization

François Brémont was invited to give a talk at Conférence des métiers at International Lycée (CIV) in Sophia 28 January 2015.

François Brémont was invited to give a talk at Nice University Conférence, La sSnté du Futur, 5 November 2015.

François Brémont was invited to give a talk at Aditel, Bank Forum, La Baule, Palais des congrès 1 and 2 October 2015.

François Brémont was interviewed by Industry Magazine, 1 October 2015.

## 11. 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<sup>o</sup> 15, pp. 1785-1796
- [4] F. BRÉMOND, M. THONNAT. *Issues of Representing Context Illustrated by Video-surveillance Applications*, in "International Journal of Human-Computer Studies, Special Issue on Context", 1998, vol. 48, pp. 375-391
- [5] G. CHARPIAT. *Learning Shape Metrics based on Deformations and Transport*, in "Proceedings of ICCV 2009 and its Workshops, Second Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA)", Kyoto, Japan, September 2009
- [6] N. CHLEQ, F. BRÉMOND, M. THONNAT. *Advanced Video-based Surveillance Systems*, Kluwer A.P. , Hingham, MA, USA, November 1998, pp. 108-118

- 
- [7] F. CUPILLARD, F. BRÉMOND, M. THONNAT. *Tracking Group of People for Video Surveillance*, Video-Based Surveillance Systems, Kluwer Academic Publishers, 2002, vol. The Kluwer International Series in Computer Vision and Distributed Processing, pp. 89-100
- [8] F. FUSIER, V. VALENTIN, F. BREMOND, M. THONNAT, M. BORG, D. THIRDE, J. FERRYMAN. *Video Understanding for Complex Activity Recognition*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 167-188
- [9] B. GEORIS, F. BREMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 189-205
- [10] C. LIU, P. CHUNG, Y. CHUNG, M. THONNAT. *Understanding of Human Behaviors from Videos in Nursing Care Monitoring Systems*, in "Journal of High Speed Networks", 2007, vol. 16, pp. 91-103
- [11] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", December 2004, vol. 16, n<sup>o</sup> 1, pp. 33-40
- [12] V. MARTIN, J.-M. TRAVERE, F. BREMOND, V. MONCADA, G. DUNAND. *Thermal Event Recognition Applied to Protection of Tokamak Plasma-Facing Components*, in "IEEE Transactions on Instrumentation and Measurement", Apr 2010, vol. 59, n<sup>o</sup> 5, pp. 1182-1191
- [13] S. MOISAN. *Knowledge Representation for Program Reuse*, in "European Conference on Artificial Intelligence (ECAI)", Lyon, France, July 2002, pp. 240-244
- [14] S. MOISAN. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*, Université de Nice-Sophia Antipolis, April 1998, Habilitation à diriger les recherches
- [15] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Blocks, a Component Framework with Checking Facilities for Knowledge-Based Systems*, in "Informatica, Special Issue on Component Based Software Development", November 2001, vol. 25, n<sup>o</sup> 4, pp. 501-507
- [16] J. PATINO, H. BENHADDA, E. CORVEE, F. BREMOND, M. THONNAT. *Video-Data Modelling and Discovery*, in "4th IET International Conference on Visual Information Engineering VIE 2007", London, UK, 25th - 27th July 2007
- [17] J. PATINO, E. CORVEE, F. BREMOND, M. THONNAT. *Management of Large Video Recordings*, in "2nd International Conference on Ambient Intelligence Developments AmI.d 2007", Sophia Antipolis, France, 17th - 19th September 2007
- [18] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, in "Software Engineering Research, Management and Applications", R. LEE (editor), Studies in Computational Intelligence, Springer, 2008, vol. 150, pp. 157-171, selected as one of the 17 best papers of SERA'08 conference
- [19] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Revue des sciences et technologies de l'information, série Théorie et Science Informatique", June 2011, vol. 4, n<sup>o</sup> 30, pp. 441-471, <http://hal.inria.fr/inria-00524499/en>

- [20] M. THONNAT, S. MOISAN. *What Can Program Supervision Do for Software Re-use?*, in "IEE Proceedings - Software Special Issue on Knowledge Modelling for Software Components Reuse", 2000, vol. 147, n<sup>o</sup> 5
- [21] M. THONNAT. *Vers une vision cognitive: mise en oeuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images*, Université de Nice-Sophia Antipolis, October 2003, Habilitation à diriger les recherches
- [22] M. THONNAT. *Special issue on Intelligent Vision Systems*, in "Computer Vision and Image Understanding", May 2010, vol. 114, n<sup>o</sup> 5, pp. 501-502
- [23] A. TOSHEV, F. BRÉMOND, M. THONNAT. *An A priori-based Method for Frequent Composite Event Discovery in Videos*, in "Proceedings of 2006 IEEE International Conference on Computer Vision Systems", New York USA, January 2006
- [24] V. VU, F. BRÉMOND, M. THONNAT. *Temporal Constraints for Video Interpretation*, in "Proc of the 15th European Conference on Artificial Intelligence", Lyon, France, 2002
- [25] V. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)", 9-15 September 2003
- [26] N. ZOUBA, F. BREMOND, A. ANFOSSO, M. THONNAT, E. PASCUAL, O. GUERIN. *Monitoring elderly activities at home*, in "Gerontechnology", May 2010, vol. 9, n<sup>o</sup> 2

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

- [27] J. BADIE. *Optimizing Process for Tracking People in video-camera network*, Université Nice Sophia Antipolis, November 2015, <https://hal.inria.fr/tel-01254613>

### Articles in International Peer-Reviewed Journals

- [28] S. BAK, F. BREMOND. *Person re-identification employing 3D scene information*, in "Journal of Electronic Imaging", September 2015, <https://hal.inria.fr/hal-01213036>
- [29] G. CHARPIAT, G. NARDI, G. PEYRÉ, F.-X. VIALARD. *Piecewise rigid curve deformation via a Finsler steepest descent*, in "Interfaces and Free Boundaries", December 2015, <https://hal.archives-ouvertes.fr/hal-00849885>
- [30] C. CUNJIAN, A. DANTCHEVA, R. ARUN. *An ensemble of patch-based subspaces for makeup-robust face recognition*, in "Information Fusion", October 2015 [DOI : 10.1016/J.INFFUS.2015.09.005], <https://hal.inria.fr/hal-01247886>
- [31] A. DANTCHEVA, P. ELIA, R. ARUN. *What else does your biometric data reveal? A survey on soft biometrics*, in "IEEE Transactions on Information Forensics and Security", September 2015, vol. 11, n<sup>o</sup> 3, pp. 441 - 467 [DOI : 10.1109/TIFS.2015.2480381], <https://hal.inria.fr/hal-01247885>



- [32] S. ELLOUMI, S. COSAR, G. PUSIOL, F. BREMOND, M. THONNAT. *Unsupervised discovery of human activities from long-time videos*, in "IET Computer Vision", March 2015, 1 p. , <https://hal.inria.fr/hal-01123895>
- [33] A. KÖNIG, C. F. CRISPIM-JUNIOR, A. G. U. COVELLA, F. BREMOND, A. DERREUMAUX, G. BENSADOUN, R. DAVID, F. VERHEY, P. AALTEN, P. ROBERT. *Ecological Assessment of Autonomy in Instrumental Activities of Daily Living in Dementia Patients by the Means of an Automatic Video Monitoring System*, in "Frontiers in Aging Neuroscience", June 2015, vol. 7, 30 p. [DOI : 10.3389/FNAGI.2015.00098], <https://hal.inria.fr/hal-01160022>
- [34] L.-E. SANCHEZ, J. ANDRÉS DIAZ-PACE, A. ZUNINO, S. MOISAN, J.-P. RIGAULT. *An approach based on feature models and quality criteria for adapting component-based systems*, in "Journal of Software Engineering Research and Development", June 2015, vol. 3, n<sup>o</sup> 10 [DOI : 10.1186/s40411-015-0022-1], <https://hal.inria.fr/hal-01238074>

### Invited Conferences

- [35] F. NEGIN, S. COSAR, M. KOPERSKI, F. BREMOND. *Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition*, in "asian conference on pattern recognition (ACPR 2015)", kuala lumpur, Malaysia, November 2015, <https://hal.inria.fr/hal-01233494>
- [36] M. K. PHAN TRAN, F. BREMOND, P. ROBERT. *Assistance for Older Adults in Serious Game using an Interactive System*, in "Games and Learning Alliance conference", Rome, Italy, December 2015, <https://hal.archives-ouvertes.fr/hal-01092329>

### International Conferences with Proceedings

- [37] D. P. CHAU, K. SUBRAMANIAN, F. BREMOND. *Adaptive Neuro-Fuzzy Controller for Multi-Object Tracker*, in "10th International Conference on Computer Vision Systems", Copenhagen, Denmark, July 2015, <https://hal.inria.fr/hal-01164734>
- [38] E. MAGGIORI, Y. TARABALKA, G. CHARPIAT. *Improved Partition Trees for Multi-Class Segmentation of Remote Sensing Images*, in "2015 IEEE International Geoscience and Remote Sensing Symposium - IGARSS 2015", Milan, Italy, IEEE, July 2015, <https://hal.inria.fr/hal-01182772>
- [39] E. MAGGIORI, Y. TARABALKA, G. CHARPIAT. *Optimizing Partition Trees for Multi-Object Segmentation with Shape Prior*, in "26th British Machine Vision Conference", Swansea, United Kingdom, September 2015, <https://hal.inria.fr/hal-01182776>

### Conferences without Proceedings

- [40] P. BILINSKI, F. BREMOND. *Video Covariance Matrix Logarithm for Human Action Recognition in Videos*, in "IJCAI 2015 - 24th International Joint Conference on Artificial Intelligence (IJCAI)", Buenos Aires, Argentina, July 2015, <https://hal.inria.fr/hal-01216849>
- [41] T. L. A. NGUYEN, D. P. CHAU, F. BREMOND. *Robust Global Tracker based on an Online Estimation of Tracklet Descriptor Reliability*, in "Advanced Video and Signal-based Surveillance", Karlsruhe, Germany, August 2015, <https://hal.inria.fr/hal-01185874>

- [42] J. ORTIZ, S. BAK, M. KOPERSKI, F. BRÉMOND. *Minimizing hallucination in Histogram of Oriented Gradients*, in "The 12th IEEE International Conference on Advanced Video and Signal-based", Karlsruhe, Germany, August 2015, <https://hal.inria.fr/hal-01199386>
- [43] I. SARRAY, A. RESSOUCHE, D. GAFFÉ, J.-Y. TIGLI, S. LAVIROTTE. *Safe Composition in Middleware for the Internet of Things*, in "Middleware for Context-aware Applications for Internet of thing (M4IoT)", Vancouver, Canada, December 2015 [DOI : 10.1145/2836127.2836131], <https://hal.inria.fr/hal-01236976>

## References in notes

- [44] 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>
- [45] M. ACHER, P. COLLET, P. LAHIRE, R. FRANCE. *Managing Feature Models with FAMILIAR: a Demonstration of the Language and its Tool Support*, in "Fifth International Workshop on Variability Modelling of Software-intensive Systems(VaMoS'11)", Namur, Belgium, VaMoS, ACM, January 2011
- [46] 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>
- [47] J. BADIE, F. BREMOND. *Global tracker: an online evaluation framework to improve tracking quality*, in "AVSS", 2014, <https://hal.inria.fr/hal-01062766>
- [48] S.-H. BAE, K.-J. YOON. *Robust Online Multi-Object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning*, in "CVPR", Columbus, IEEE, June 2014
- [49] S. BAK, S. ZAIDENBERG, B. BOULAY, F. BREMOND. *Improving Person Re-identification by Viewpoint Cues*, in "Advanced Video and Signal-based Surveillance", Seoul, South Korea, August 2014, pp. 1-6, <https://hal.archives-ouvertes.fr/hal-01004390>
- [50] J. BERCLAZ, F. FLEURET, P. FUA. *Multiple Object Tracking using Flow Linear Programming*, 2009
- [51] A. BIALKOWSKI, S. DENMAN, S. SRIDHARAN, C. FOOKES, P. LUCEY. *A database for person re-identification in Multi-Camera Surveillance Networks*, in "DICTA", 2012
- [52] A. BORJI, S. FRINTROP, D. SIHITE, L. ITTIL. *Adaptive object tracking by learning background context*, in "CVPR", 2012, [http://ilab.usc.edu/publications/doc/Borji\\_et al12ego.pdf](http://ilab.usc.edu/publications/doc/Borji_et al12ego.pdf)
- [53] D. P. CHAU, F. BREMOND, M. THONNAT. *Online evaluation of tracking algorithm performance*, in "The 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP)", London, UK, , December 2009, <https://hal.inria.fr/inria-00486479>
- [54] 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/>

- [55] N. DALAL, B. TRIGGS. *Histograms of oriented gradients for human detection*, in "Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on", IEEE, 2005, vol. 1, pp. 886–893
- [56] 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
- [57] P. DOLLAR, C. WOJEK, B. SCHIELE, P. PERONA. *Pedestrian detection: An evaluation of the state of the art*, in "Pattern Analysis and Machine Intelligence, IEEE Transactions on", 2012, vol. 34, n<sup>o</sup> 4, pp. 743–761
- [58] M. ENZWEILER, D. M. GAVRILA. *Monocular pedestrian detection: Survey and experiments*, in "Pattern Analysis and Machine Intelligence, IEEE Transactions on", 2009, vol. 31, n<sup>o</sup> 12, pp. 2179–2195
- [59] M. EVANS, C. J. OSBORNE, J. FERRYMAN. *Multicamera object detection and tracking with object size estimation*, in "Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on", 2013, pp. 177–182
- [60] M. FARENZENA, L. BAZZANI, A. PERINA, V. MURINO, M. CRISTANI. *Person re-identification by symmetry-driven accumulation of local features*, in "CVPR", 2010
- [61] R. GIRSHICK, J. DONAHUE, T. DARRELL, J. MALIK. *Rich feature hierarchies for accurate object detection and semantic segmentation*, in "Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on", IEEE, 2014, pp. 580–587
- [62] D. GRAY, H. TAO. *Viewpoint invariant pedestrian recognition with an ensemble of localized features*, in "Computer Vision–ECCV 2008", Springer, 2008, pp. 262–275
- [63] S. HALLER. *The Things in the Internet of Things*, in "Poster at the (IoT 2010). Tokyo, Japan, November", 2010, vol. 5, 26 p. , [http://www.iot-a.eu/public/news/resources/TheThingsintheInternetofThings\\_SH.pdf](http://www.iot-a.eu/public/news/resources/TheThingsintheInternetofThings_SH.pdf)
- [64] M. HIRZER, C. BELEZNAI, P. M. ROTH, H. BISCHOF. *Person Re-identification by Descriptive and Discriminative Classification*, in "Image Analysis", Springer, 2011, pp. 91–102
- [65] M. HOFMANN, D. WOLF, G. RIGOLL. *Hypergraphs for Joint Multi-View Reconstruction and Multi-Object Tracking*, 2013
- [66] 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
- [67] A. KARAKOSTAS, A. BRIASSOULI, K. AVGERINAKIS, I. KOMPATSIARIS, M. TSOLAKI. *The Dem@Care Experiments and Datasets: a Technical Report*, 2014
- [68] A. KRIZHEVSKY, G. HINTON. *Learning multiple layers of features from tiny images*, Citeseer, 2009
- [69] A. KRIZHEVSKY, I. SUTSKEVER, G. E. HINTON. *Imagenet classification with deep convolutional neural networks*, in "Advances in neural information processing systems", 2012, pp. 1097–1105

- [70] C.-H. KUO, C. HUANG, R. NEVATIA. *Multi-target tracking by on-line learned discriminative appearance models*, in "CVPR", 2010
- [71] 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
- [72] Y. LECUN, L. BOTTOU, Y. BENGIO, P. HAFFNER. *Gradient-based learning applied to document recognition*, in "Proceedings of the IEEE", 1998, vol. 86, n<sup>o</sup> 11, pp. 2278–2324
- [73] 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)
- [74] A. MILAN, S. ROTH, K. SCHINDLE. *Continuous energy minimization for multi-target tracking*, in "CVPR", 2014, [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6520846](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6520846)
- [75] A. MILAN, K. SCHINDLER, S. ROTH. *Detection-and Trajectory-Level Exclusion in Multiple Object Tracking*, in "CVPR '13", 2013, [http://www.cv-foundation.org/openaccess/content\\_cvpr\\_2013/html/Milan\\_Detection-and\\_Trajectory-Level\\_2013\\_CVPR\\_paper.html](http://www.cv-foundation.org/openaccess/content_cvpr_2013/html/Milan_Detection-and_Trajectory-Level_2013_CVPR_paper.html)
- [76] S. MOISAN, J.-P. RIGAULT, M. ACHER, P. COLLET, P. LAHIRE. *Run Time Adaptation of Video-Surveillance Systems: A software Modeling Approach*, in "ICVS, 8th International Conference on Computer Vision Systems", Sophia Antipolis, France, September 2011, <http://hal.inria.fr/inria-00617279/en>
- [77] A. T. NGHIEM, E. AUVINET, J. MEUNIER. *Head detection using Kinect camera and its application to fall detection*, in "Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on", 2012, pp. 164-169, <http://doi.org/10.1109/ISSPA.2012.6310538>
- [78] M. OREN, C. PAPAGEORGIOU, P. SINHA, E. OSUNA, T. POGGIO. *Pedestrian Detection Using Wavelet Templates*, in "cvpr", 1997, pp. 193-99
- [79] G. PONS-MOLL, B. ROSENHAHN. *Branch-and-price global optimization for multi-view multi-target tracking*, in, in "Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012", 2012
- [80] L. M. ROCHA, S. MOISAN, J.-P. RIGAULT, S. SAGAR. *Girgit: A Dynamically Adaptive Vision System for Scene Understanding*, in "ICVS", Sophia Antipolis, France, September 2011, <http://hal.inria.fr/inria-00616642/en>
- [81] M. ROHRBACH, M. REGNERI, M. ANDRILUKA, S. AMIN, M. PINKAL, B. SCHIELE. *Script Data for Attribute-Based Recognition of Composite Activities*, in "Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I", 2012, pp. 144–157
- [82] 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>

- [83] P. SERMANET, D. EIGEN, X. ZHANG, M. MATHIEU, R. FERGUS, Y. LECUN. *Overfeat: Integrated recognition, localization and detection using convolutional networks*, in "arXiv preprint arXiv:1312.6229", 2013
- [84] S. TANG, M. ANDRILUKA, A. MILAN, K. SCHINDLER, S. ROTH, B. SCHIELE. *Learning People Detectors for Tracking in Crowded Scenes*, in "IEEE International Conference on Computer Vision (ICCV)", December 2013, [http://www.cv-foundation.org/openaccess/content\\_iccv\\_2013/html/Tang\\_Learning\\_People\\_Detectors\\_2013\\_ICCV\\_paper.html](http://www.cv-foundation.org/openaccess/content_iccv_2013/html/Tang_Learning_People_Detectors_2013_ICCV_paper.html)
- [85] 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<sup>o</sup> 3-4, ISSN 0003-4347 (Print) ISSN 1958-9395 (Online)
- [86] 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<sup>o</sup> 1, ISSN (Online): 1694-0784 ISSN (Print): 1694-0814
- [87] J. R. UIJLINGS, K. E. VAN DE SANDE, T. GEVERS, A. W. SMEULDERS. *Selective search for object recognition*, in "International journal of computer vision", 2013, vol. 104, n<sup>o</sup> 2, pp. 154–171
- [88] T. WANG, S. GONG, X. ZHU, S. WANG. *Person Re-identification by Video Ranking*, in "ECCV", 2014
- [89] H. WANG, A. KLASER, C. SCHMID, C.-L. LIU. *Action Recognition by Dense Trajectories*, in "Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition", Washington, DC, USA, CVPR '11, IEEE Computer Society, 2011, pp. 3169–3176, <http://dx.doi.org/10.1109/CVPR.2011.5995407>
- [90] H. WANG, A. KLÄSER, C. SCHMID, C.-L. LIU. *Action Recognition by Dense Trajectories*, in "CVPR", June 2011, pp. 3169-3176
- [91] K. WEINBERGER, J. BLITZER, L. SAUL. *Distance metric learning for large margin nearest neighbor classification*, in "NIPS", 2006, [www.jmlr.org/papers/volume10/weinberger09a/weinberger09a.pdf](http://www.jmlr.org/papers/volume10/weinberger09a/weinberger09a.pdf)
- [92] C. WOJEK, S. WALK, B. SCHIELE. *Multi-cue onboard pedestrian detection*, in "Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on", IEEE, 2009, pp. 794–801
- [93] B. WU, R. NEVATIA. *Cluster boosted tree classifier for multi-view, multi-pose object detection*, in "Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on", IEEE, 2007, pp. 1–8
- [94] I. A. K. XU YAN, S. K. SHAH. *What do I see? Modeling Human Visual Perception for Multi-person Tracking*, in "ECCV", Zurich, Germany, Sept 2014, <http://www2.cs.uh.edu/~xuyan/docs/eccv2014.pdf>
- [95] A. R. ZAMIR, A. DEGHAN, M. SHAH. *GMCP-Tracker: Global Multi-object Tracking Using Generalized Minimum Clique Graphs*, in "ECCV", 2012
- [96] R. ZHAO, W. OUYANG, X. WANG. *Unsupervised salience learning for person re-identification*, in "CVPR", 2013
- [97] R. ZHAO, W. OUYANG, X. WANG. *Learning Mid-level Filters for Person Re-identification*, in "CVPR", 2014

- [98] W. ZHENG, S. GONG, T. XIANG. *Person re-identification by probabilistic relative distance comparison*, in "CVPR", 2011, <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5995598>