Activity Report 2016

Project-Team STARS

Spatio-Temporal Activity Recognition Systems

RESEARCH CENTER
Sophia Antipolis - Méditerranée

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team STARS

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

Keywords:

Computer Science and Digital Science:

2.1.8. - Synchronous languages
2.1.11. - Proof languages
2.3.3. - Real-time systems
2.4.2. - Model-checking
2.4.3. - Proofs
2.5. - Software engineering
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
3.4.6. - Neural networks
4.7. - Access control
5.1. - Human-Computer Interaction
5.3.2. - Sparse modeling and image representation
5.3.3. - Pattern recognition
5.4.1. - Object recognition
5.4.2. - Activity recognition
5.4.3. - Content retrieval
5.4.5. - Object tracking and motion analysis
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

Francois Bremond [Team leader, Inria, Research Scientist, HDR]
Sabine Moisan [Inria, Research Scientist, HDR]
Annie Ressouche [Inria, Research Scientist]

Faculty Member

Jean Paul Rigault [Univ. Nice, Professor]

Engineers

Manikandan Bakthavatchalam [Inria]
Vasanth Bathrinarayanan [Inria]
Anais Ducoffe [Inria]
Rachid Guerchouche [Inria]
Furqan Muhammad Khan [Inria]
Matias Marin [Inria]
Thanh Hung Nguyen [Inria]
Javier Ortiz [Inria]

PhD Students

Auriane Gros [CHU Nice]
Michal Koperski [Toyota, granted by CIFRE]
Farhood Negin [Inria]
Thi Lan Anh Nguyen [Inria]
Minh Khue Phan Tran [Genious, granted by CIFRE]
Ines Sarray [Inria]
Ujjwal Ujjwal [VEDCOM, granted by CIFRE]

Post-Doctoral Fellows

Julien Badie [Inria, granted by IRCA project]
Carlos Fernando Crispim Junior [Inria, granted by BPIFRANCE FINANCEMENT SA]
Remi Trichet [Inria, granted by MOVEMENT project]
Piotr Tadeusz Bilinski [Inria, granted by Toyota project, until Sep 2016]
Antitza Dantcheva [Inria, granted by Labex]

Visiting Scientists

Karel Krehnac [Centaur Project, from May 2016 to Jun 2016]
Jana Trojanova [Centaur project, from Mar 2016 to Sep 2016]
Salwa Baabou [Guest PhD, to Nov 2016]
Siyuan Chen [Guest PhD, to Feb 2016]
Adlen Kerboua [PhD]
Aimen Neffati [Inria, from Jul 2016 to Aug 2016]
Luis Emiliano Sanchez [University of Rosario (Argentina), from Sep 2016 to Dec 2016]

Administrative Assistant

Jane Desplanques [Inria]

Others

Kanishka Nithin Dhandapani [Inria, granted by Toyota project, until Oct 2016]
Seongro Yoon [Inria, from Apr 2016]
Yashas Annadani [Inria, from May 2016 to august 2016]
Ghada Bahloul [Inria, Engineers, granted by EIT project]
Robin Bermont [Inria, research support, from Oct 2016]
Ulysse Castet [Inria, from Apr 2016 to Aug 2016]
Etienne Corvee [Associate partner]
Chandraja Dharmana [Inria, from May 2016]
Margaux Failla [Inria, research support, from Oct 2016]
Loic Franchi [Inria, from Jun 2016 until Jul 2016]
Daniel Gaffe [External Collaborator, Univ. Nice, Associate Professor]
Renaud Heyrendt [Inria, from Apr 2016 to Jun 2016]
Guillaume Lombard [Inria, from Apr 2016 to Jul 2016]
Robinson Menetrey [Inria, from Apr 2016 to Jun 2016]
Nairouz Mrabah [Inria, from Apr 2016 to Sep 2016]

Isabel Rayas [Inria, from Jun 2016]
Philippe Robert [External Collaborator, CHU Nice and COBTECK]
Hugues Thomas [Inria,from Apr 2016 to Sep 2016]
Jean Yves Tigli [External Collaborator, Univ. Nice, Associate Professor]
Shanu Vashishtha [Inria, from May 2016 to Jul 2016]

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.

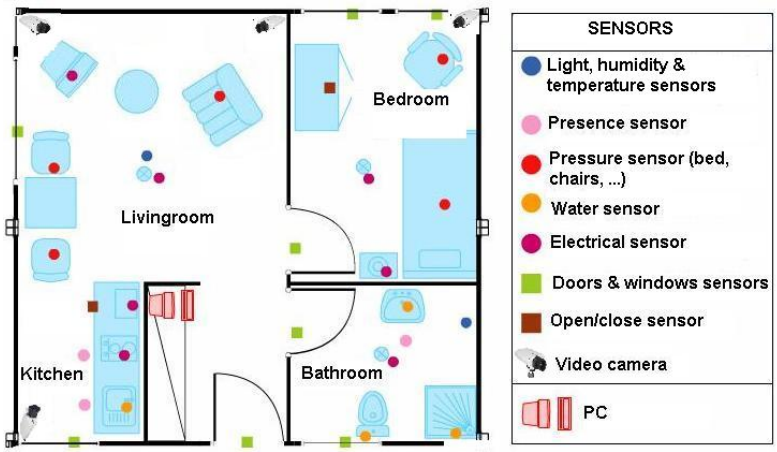

Figure 1. Homecare monitoring: the set of sensors embedded in an apartment

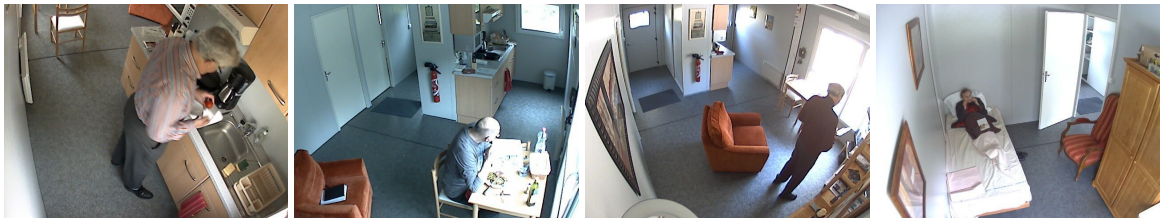

Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

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

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

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 detection 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

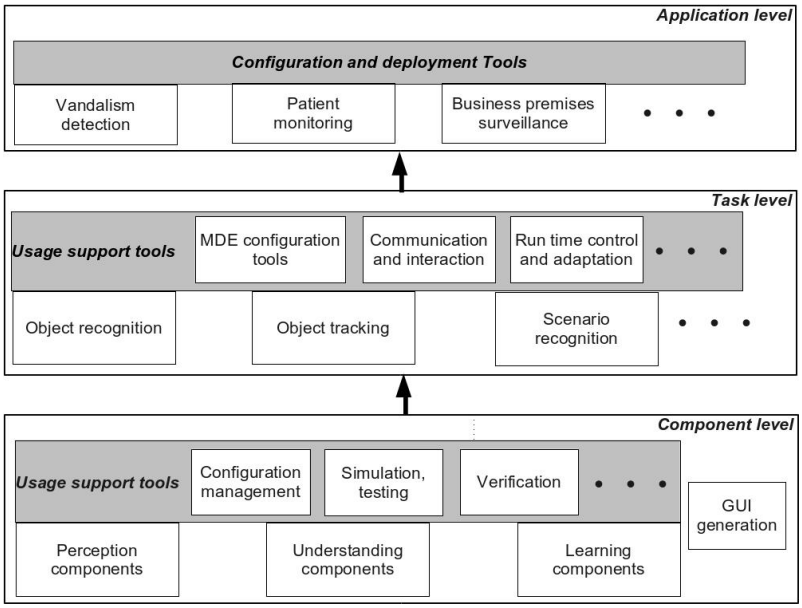

Figure 4. Global Architecture of an Activity Recognition The gray areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

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 [45]. 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 [77]. 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], [89], [82].

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 [63], [91] (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. New Software and Platforms

5.1. 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 Gaffe and Annie Ressouche
- Contact: Annie Ressouche
- URL: http://www-sop.inria.fr/teams/pulsar/projects/Clem/

5.2. 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, Francois Bremond and Vasanth Bathrinarayanan
- Contact: Francois Bremond

5.3. 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 Francois Bremond
- Contact: Francois Bremond

5.4. 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, Anais Ducoffe, Sofia Zaidenberg, Annunziato Polimeni and Julien Gueytat
- Partner: Neosensys
- Contact: Anais Ducoffe

5.5. 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 and Francois Bremond
- Contact: Francois Bremond

5.6. 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, Francois Bremond, Annunziato Polimeni and Yves Pichon
- Partner: Neosensys
- Contact: Francois Bremond

5.7. 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 Brémond, Carlos Fernando Crispim Junior and Etienne Corvée
- Partners: CEA - CHU Nice - I2R - Université de Hamburg - USC Californie
- Contact: Francois Bremond
- URL: https://team.inria.fr/stars/software

5.8. 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 Francois Bremond
- Contact: Francois Bremond
- URL: http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl_Description.html

5.9. py_ad

py action detection
FUNCTIONAL DESCRIPTION

Action Detection framework Allows user to detect action in video stream. It uses model trained in py_ar.

- Participants: Michal Koperski and Francois Bremond
- Contact: Michal Koperski

5.10. 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 Francois Bremond
- Contact: Michal Koperski

5.11. 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

5.12. 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 Francois Bremond
- Contact: Michal Koperski

5.13. 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 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 Francois Bremond
- Contact: Michal Koperski

6. New Results

6.1. 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.

6.1.1. Perception for Activity Recognition

Participants: Piotr Bilinski, François Brémond, Etienne Corvée, Antitza Dancheva, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Javier Ortiz, Remi Trichet, Jana Trojanova, Ujjwal Ujjval.

The new results for perception for activity recognition are:

- Exploring Depth Information for Head Detection with Depth Images (see 6.2)
- Modeling Spatial Layout of Features for Real World Scenario RGB-D Action Recognition (see 6.3)
- Multi-Object Tracking of Pedestrian Driven by Context (see 6.4)
- Pedestrian detection: Training set optimization (see 6.5)
- Pedestrian Detection on Crossroads (see 6.6)
- Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in Musical Mnemotherapy (see 6.7)
- Hybrid Approaches for Gender estimation (see 6.8)
- Unsupervised Metric Learning for Multi-shot Person Re-identification (see 6.9)

6.1.2. Semantic Activity Recognition

Participants: François Brémond, Carlos Fernando Crispim Junior, Michal Koperski, Farhood Negin, Thanh Hung Nguyen, Philippe Robert.

For this research axis, the contributions are :

- Semi-supervised Understanding of Complex Activities in Large-scale Datasets (see 6.10)
- On the Study of the Visual Behavioral Roots of Alzheimer's disease (see 6.11)
- Uncertainty Modeling with Ontological Models and Probabilistic Logic Programming (see 6.12)
- A Hybrid Framework for Online Recognition of Activities of Daily Living In Real-World Settings (see 6.13)
- Praxis and Gesture Recognition (see 6.14)

6.1.3. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé, Rachid Guerchouche, Matias Marin, Etienne Corvée, Julien Badie, Manikandan Bakthavatchalam, Vasanth Bathrinarayanan, Ghada Balhoul, Anais Ducoffe, Jean Yves Tigli, François Brémond.

The contributions for this research axis are:

- Scenario Recognition (see 6.15)
- The CLEM Workflow (see 6.16)
- Safe Composition in Middleware for Internet of Things (see 6.17)
- Verification of Temporal Properties of Neuronal Archetype (see 6.18)
- Dynamic Reconfiguration of Feature Models (see 6.19)
- Setup and management of SafEE devices (see 6.20)
- Brick & Mortar Cookies (see 6.21)

6.2. Exploring Depth Information for Head Detection with Depth Images

Participants: Thanh Hung Nguyen, Siyuan Chen.

Head detection may be more demanding than face recognition and pedestrian detection in the scenarios where a face turns away or body parts are occluded in the view of a sensor, but when locating people is needed. This year [29], we introduce an efficient head detection approach for single depth images at low computational expense. First, a novel head descriptor was developed and used to classify pixels as head or non-head. We used depth values to guide each window size, to eliminate false positives of head centers, and to cluster head pixels, which significantly reduce the computation costs of searching for appropriate parameters. High head detection performance was achieved in experiments with 90% accuracy for our dataset containing heads with different body postures, head poses, and distances to a Kinect2 sensor, and above 70% precision on a public dataset composed of a few daily activities, which is better than using a head-shoulder detector with HOG feature for depth images (see Figure 5)

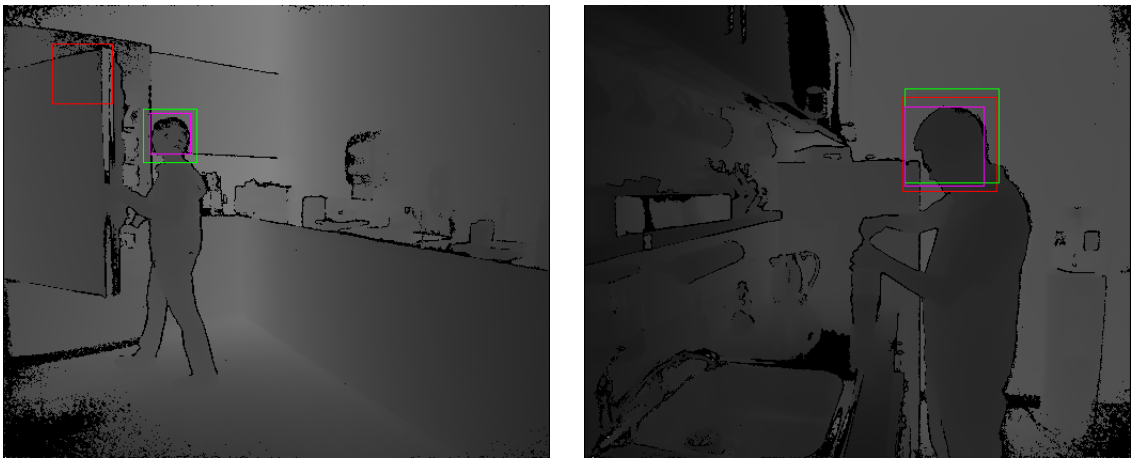

Figure 5. Examples of head detection where our algorithm successfully detects head. Pink square represents the ground truth, green rectangle represents our algorithm.

6.3. Modeling Spatial Layout of Features for Real World Scenario RGB-D Action Recognition

Participants: Michal Koperski, François Brémond.

keywords: computer vision, action recognition

Challenges in action representation in real-world scenario using RGB-D sensor

With RGB-D sensor it is easy to take advantage of real-time skeleton detection. Using skeleton information we can model not only dynamics of action, but also static features like pose. Skeleton-based methods have been proposed by many authors, and have reported superior accuracy on various daily activity data-sets. But the main drawback of skeleton-based methods is that they cannot make the decision when skeleton is missing.

We claim that in real world scenario of daily living monitoring, skeleton is very often not available or is very noisy. This makes skeleton based methods unpractical. There are several reasons why skeleton detection fails in real-world scenario. Firstly, the sensor has to work outside of it's working range. Since daily living monitoring is quite an unconstrained environment, the monitored person is very often too far from sensor, or is captured from non-optimal viewpoint angle. In Figure 6 we show two examples where skeleton detection fails. In the first example, the person on the picture wears black jeans which interferes with sensor. In such a case depth information from lower body parts is missing, making skeleton detection inaccurate. In the second example (see Figure 7) the person is too far from sensor and basically disappears in the background. In this case depth information is too noisy, thus skeleton detection fails. All disadvantages mentioned above will affect skeleton-based action recognition methods, because they strictly require skeleton detection.

On the other hand, local points-of-interest methods do not require skeleton detection, nor segmentation. That is why they received great amount of interest in RGB based action recognition where segmentation is much more difficult than with RGB-D. Those methods rely mostly on detection of points-of-interest usually based on some motion features (eg optical flow). The features are either based on trajectory of points-of-interest or descriptors are computed around the points-of-interest. One of the main disadvantage of those methods is fact that they fail when they cannot "harvest" enough points-of-interest. It happens when action has low dynamics eg "reading a book" or "writing on paper". Such actions contain very low amount of motion coming from hand when writing or turning the pages. In addition local points-of-interest methods very often ignore the spatial layout of detected features.

Proposed method

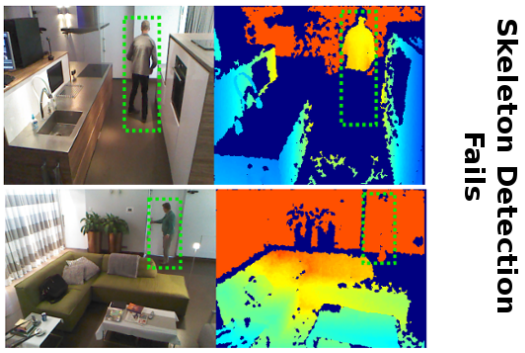

Figure 6. We show two examples where skeleton detection methods fail. Pictures on the left show RGB frame, pictures on the right show depth map (dark blue indicates missing depth information).

To address those problems we propose to replace skeleton detection by RGB-D based people detector. Note that person detection is much easier than skeleton detection. In addition we propose to use two people detectors: RGB and depth based - to take advantage of two information streams.

We propose to model spatial layout of motion features obtained from a local points-of-interest based method. We use Dense Trajectories [99] as a point of interest detector and MBH (Motion Boundary Histogram [62]) as a descriptor. To improve the discriminating power of MBH descriptor we propose to model spatial-layout of visual words computed based on MBH (Figure 7). We divide a person bounding box into Spatial Grid (SG) and we compute Fisher Vector representation in each cell. In addition, we show that other spatial-layout encoding methods also improve recognition accuracy. We propose 2 alternative spatial-layout encoding methods an we compare them with Spatial Grid.

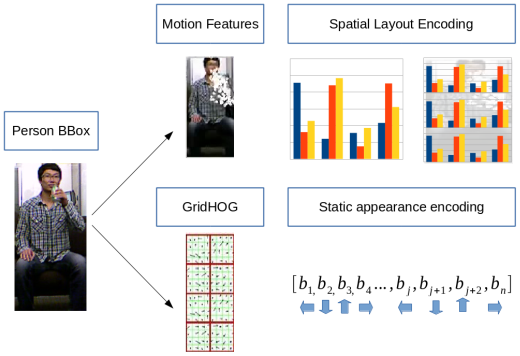

Figure 7. We show proposed method where we use people detection in place of skeleton. Next we propose to encode spatial-layout of visual words computed from motion features. In addition we propose GridHOG descriptor which encodes static appearance information.

To improve recognition of actions with low amount of motion we propose a descriptor which encodes rough static appearance (Figure 7). This can be interpreted as rough pose information. We divide the detected person bounding box into grid cells. Then we compute HOG [61] descriptor inside each cell to form the GHOG (GridHog) descriptor.

Further details can be find in the paper [37]. The contributions of this paper can be listed as follows:

- We propose to use two people detectors (RGB and depth based) to obtain person bounding box instead of skeleton.
- We propose to use Spatial Grid (SG) inside person bounding box. To model spatial-layout of MBH features.
- We propose to encode static information by using novel GHOG descriptor.
- We propose two other methods which model spatial-layout of MBH features and we compare them with Spatial Grid.

Experiments

We evaluate our approach on three daily activity data-sets: MSRDailyActivity3D, CAD-60 and CAD-120. The experiments show that we outperform most of the skeleton-based methods without requiring difficult in real-world scenario skeleton detection and thus being more robust (see Table1, Table2 and Table3).

6.4. Multi-Object Tracking of Pedestrian Driven by Context

Participants: Thi Lan Anh Nguyen, François Brémond, Jana Trojanova.

Keywords: Tracklet fusion, Multi-object tracking

Multi-object tracking (MOT) is essential to many applications in computer vision. As so many trackers have been proposed in the past, one would expect the tracking task as solved. It is true for scenarios containing solid background with a low number of objects and few interactions. However, scenarios with appearance changes due to pose variation, abrupt motion changes, and occlusion still represent a big challenge.

Table 1. Recognition Accuracy Comparison for MSRDailyActivity3D data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
NBNN [94]	*70.00*
HON4D [87]	*80.00*
STIP + skeleton [106]	*80.00*
SSFF [95]	*81.90*
DSCF [102]	*83.60*
Actionlet Ensemble [101]	*85.80*
RGGP + fusion [79]	*85.60*
Super Normal [80]	*86.26*
BHIM [74]	86.88
DCSF + joint [102]	*88.20*
Our Approach	**85.95**

Table 2. Recognition Accuracy Comparison for CAD-60 data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
STIP [106]	62.50
Order Sparse Coding [86]	*65.30*
Object Affordance [75]	*71.40*
HON4D [87]	*72.70*
Actionlet Ensemble [101]	*74.70*
JOULE-SVM [72]	*84.10*
Our Approach	**80.36**

Table 3. Recognition Accuracy Comparison for CAD-120 data-set. corresponds to methods which require skeleton detection.

Method	Accuracy [%]
Salient Proto-Objects [92]	78.20
Object Affordance [75]	*84.70*
STS [76]	*93.50*
Our Approach	**85.48**

In the state of the art, some sets of efficient methods are proposed to face this challenge: data association (local and global) and tracking parameter adaptation. A very popular method for local data association is the bipartite matching. The exact solution can be found via Hungarian algorithm [85]. These methods are computationally inexpensive, but can deal only with short term occlusion. An example of global method is the extension of the bipartite matching into network flow [104]. Given the objects detections at each frame, the direct acyclic graph is formed and the solution is found through minimum-cost flow algorithm. The algorithms reduce trajectory fragments and improve trajectory consistency but lack robustness to identity switches of close or intersecting trajectories.

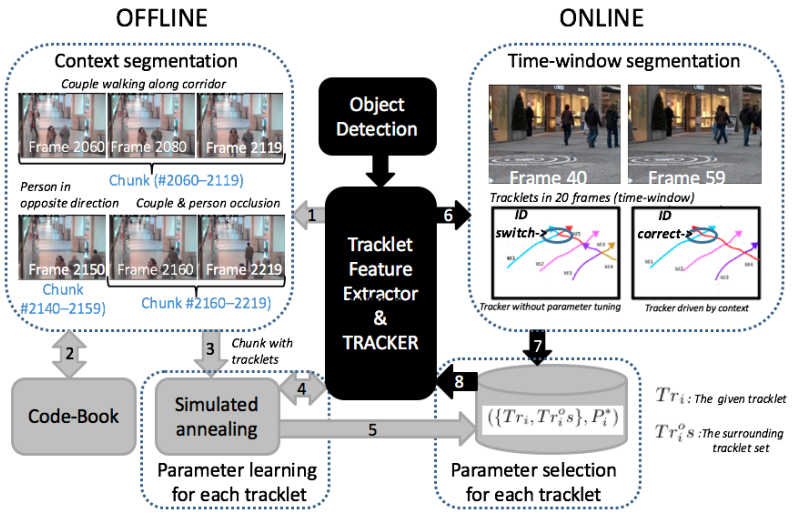

Figure 8. Our proposed framework.

Another set of methods for MOT is online parameter adaptation [56]. They tune automatically the tracking parameters based on the context information, while methods mentioned above use one appearance and/or one motion feature for the whole video. In [56], the authors learn the parameters for the scene context offline. In online phase the tracking parameters are selected from database based on the current context of the scene. These parameters are applied to all objects in the scene. Such a concept assumes discriminative appearance and trajectories among individuals, which is not always the case in real scenarios.

In order to overcome these limitations, we propose a new long term tracking framework. This framework has several dominant contributions:

- We introduce new long term tracking framework which combines short data association and the online parameter tuning for individual tracklets. In contrast to previous methods that used the same setting for all tracklets.
- We show that large number of parameters can be efficiently tuned via multiple simulated annealing, whereas previous method could tune only a limited number of parameters and fix the rest to be able to do exhaustive search.
- We define the surrounding context around each tracklet and similarity metric among tracklets allowing us to match learned context with unseen video set.

The proposed framework was trained on 9 public video sequences and tested on 3 unseen sets. It outperforms the state-of-art pedestrian trackers in scenarios of motion changes, appearance changes and occlusion of objects as shown in Table 4. The paper is accepted in conference AVSS-2016 [39].

Table 4. Tracking performance. The best values are printed in red.

Dataset	Method	MOTA	MOTP	GT	MT	PT	ML
PETS2009	Shitrit et al. [52]	0.81	0.58	21	–	–	–
	Bae et al.*-global association* [50]	0.73	0.69	23	*100*	0	*0.0*
	Chau et al. [57]	0.62	0.63	21	–	–	–
	Chau [58]([57] + parameter tuning for whole video context)	0.85	0.71	21	–	–	–
	Ours ([57] + Proposed approach)	*0.86*	*0.73*	21	76.2	14.3	9.5
TUD-Stadtmitte	Andriyenko et al. [47]	0.62	0.63	9	60.0	20.0	10.0
	Milan et al. [81]	*0.71*	*0.65*	9	*70.0*	20.0	*0.0*
	Chau et al. [57]	0.45	0.62	10	60.0	40.0	*0.0*
	Chau [58]([57] + parameter tuning for whole video context)	–	–	10	*70.0*	10.0	20.0
	Ours ([57] + Proposed approach)	0.47	*0.65*	10	*70.0*	30.0	*0.0*
TUD-Crossing	Tang et al. [96]	–	–	11	*53.8*	38.4	7.8
	Chau et al. [57]	0.69	0.65	11	46.2	53.8	*0.0*
	Ours ([57] + Proposed approach)	*0.72*	*0.67*	11	*53.8*	46.2	*0.0*

6.5. Pedestrian detection: Training set optimization

Participants: Remi Trichet, Javier Ortiz.

keywords: computer vision, pedestrian detection, classifier training, data selection, data generation, data weighting

The emphasis of our work is on data selection. Training for pedestrian detection is, indeed, a peculiar task. It aims to differentiate a few positive samples with relatively low intra-class variation and a swarm of negative samples picturing everything else present in the dataset. Consequently, the training set lacks discrimination and is highly imbalanced. Due to the possible creation of noisy data while oversampling, and the likely loss of information when undersampling, balancing positive and negative instances is a rarely addressed issue in the literature.

Bearing these data selection principles in mind, we introduce a new training methodology, grounded on a two-component contribution. First, it harnesses an expectation-maximization scheme to weight important training data for classification. Second, it improves the cascade-of-rejectors [105][54] classification by enforcing balanced train and validation sets every step of the way, and optimizing separately for recall and precision. A new data generation technique was developped for this purpose.

The training procedure unfolds as follows. After the initial data selection, we balance the negative and positive sample cardinalities. Then, a set of n negative data rejectors are trained and identified negative data are discarded. The validation set negative samples are iteratively oversampled after each training to ensure a balanced set. The final classifier is learned after careful data selection. Figure 9 illustrates the process.

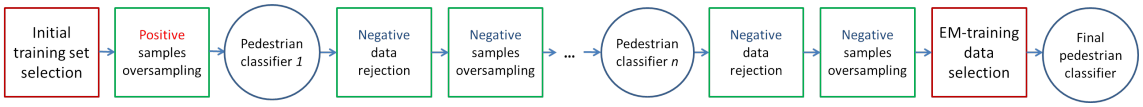

Figure 9. Training pipeline.

Experiments carried out on the Inria [61] and PETS2009 [69] datasets, demonstrate the effectiveness of the approach, leading to a simple HoG-based detector to outperform most of its near real-time competitors.

Table 5. Comparison with the state-of-the-art on the Inria dataset. Our approach is in italic. Computation time are calculated according to 640×480 resolution frames. The used metric is the log-average miss rate (the lower the better).

Method	Inria	Speed
HoG [61]	46%	21fps
DPM-v1 [68]	44%	< 1fps
HoG-LBP [98]	39%	Not provided
MultiFeatures [100]	36%	< 1fps
FeatSynth [51]	31%	< 1fps
MultiFeatures+CSS [97]	25%	No
FairTrain - HoG + Luv	*25%*	*11fps*
FairTrain - HoG	*25%*	*16fps*
Channel Features [65]	21%	0.5fps
FPDW [64]	21%	2-5fps
DPM-v2 [67]	20%	< 1fps
VeryFast [53]	18%	8fps(CPU)
VeryFast [53]	18%	135fps(GPU)
WordChannels [60]	17%	8fps(GPU)

Table 6. Comparison with the state-of-the-art on the PETS2009 S2.L1 sequence. Our approach is in italic. The used metric is the MODA (the higher the better).

Method	PETS2009	Speed
Arsic [48]	44%	n.a.
Alahi [46]	73%	n.a.
Conte [59]	85%	n.a.
FairTrain - HoG	*85.38%*	*29fps*
FairTrain - HoG + Luv	*85.49%*	*18fps*
Breitenstein [55]	89%	n.a.
Yang [103]	96%	n.a.

6.6. Pedestrian Detection on Crossroads

Participants: Ujwal Ujwal, François Brémond.

Pedestrian detection has a specific relevance in the space of object detection problems in computer vision. Due to increasing role of automated surveillance systems in increasing areas, demands for a highly robust and accurate pedestrian detection system is increasing day after day. Recently, deep learning has emerged as an important paradigm to tackle complex object detection problems. This year, we performed our initial studies on pedestrian detection using deep learning techniques. These studies form an important basis for us to extend our work in the future.

Evaluation Metrics

The relative comparison of different pedestrian detection systems was done using evaluation metrics. In the area of pedestrian detection, the most widely used evaluation metric is that of *miss rate*(MR). *Miss rate* is related to the concept of *recall*, which is another very commonly used metric in computer vision, especially in problems related to retrieval of images and concepts. *Miss Rate* is defined as follows:

$$Miss\ Rate = \frac{False\ Negatives}{True\ Positives + False\ Negatives} \tag{1}$$

		Pedestrian Detector	
		Pedestrian	Other
Ground Truth	Pedestrian	True Positive (TP)	False Negative (FN)
	Other	False Positive (FP)	True Negative (TN)

Figure 10. True and False Positives in pedestrian detection

In equation 1, *True Positives*(TP) and *False Negatives*(FN) can be understood from figure 10. A good pedestrian detector should not miss many people in a scene and this aspect is reflected in the definition of equation 1. A good pedestrian detector is required to detect as few *False Positives*(FP) as possible. This is expressed in the literature usually in the form of False Positives Per Image(FPPI). FPPI is basically a per-image average of total number of FP detections.

Pedestrian detection systems usually work with a number of parameters. Different values of these parameters may tune a system to different MR and FPPI value. This is usually expressed in the form of a *Precision-recall*(PR) curve. This curve is created by varying a control parameter of a system and plotting MR and FPPI values. In literature it is customary to report MR value at 0.1 FPPI.

Experiments

We considered deep learning based models for our initial set of experiments. This is primarily owing to their popularity and the promise which they have demonstrated in the area of object detection over the past several years.

There are many deep learning based models which have been used for object detection. The purpose of these experiments was to gain a deeper insight into the performance of deep neural networks for pedestrian detection. We experimented with Faster-RCNN [88] and SSD detector [78]. These were chosen owing to the fact that they are recent models (2015 for Faster-RCNN and 2016 for SSD Detector), and have displayed state-of-art performance in terms of detection speeds and accuracy across many object categories.

The results shown in table7 were obtained by fine-tuning VGG-16 with imagenet and MS-COCO datasets which did not involve any public dataset specific to pedestrian detection. Hence, we took the fine-tuned model and further fine-tuned it with different pedestrian datasets to study the effectiveness of fine-tuning with pedestrian-specific datasets.

Each row in the first column of table8, reflects the dataset(s) which were used to fine-tune the model. For each row, the model was fine-tuned using the dataset indicated in its first column, as well as the datasets indicated in the first column of all rows preceding it. The model was then evaluated against the test set of each dataset and the miss-rates are indicated in the table.

Table 7. Performance of fine-tuned Faster RCNN on pedestrian detection datasets.Numbers indicate the miss-rate.

Performance of fine-tuned Faster RCNN		
Dataset	**Faster RCNN Performance**	**State of Art**
Inria	13.47%	13%
Daimler	37.7%	29%
ETH-Zurich	32.1%	
Caltech	26.7%	19%
TUD-Brussels	52.2%	45%

-

Table 8. Faster-RCNN performance after fine-tuning with pedestrian datasets. Numbers indicate the miss-rate.

	Image datasets				
Trained Model	**Inria**	**Daimler**	**TUD-Brussels**	**ETH-Zurich**	**Caltech**
+Inria	13.4%	36.9%	52%	32.1%	28.2%
+Daimler	13.6%	33.7%	51.1%	32.7%	29.1%
+ETH-Zurich	13.8%	34.6%	49.3%	32%	26%
+Caltech	16%	35.4%	48%	33.2%	25.2%

While the initial results as seen from table 7 are encouraging, they still need a lot of improvement especially with complex datasets such as TUD-Brussels and Caltech. We also see from table8, that fine-tuning with pedestrian datasets tends to improve the performance but the magnitude of improvement varies depending upon the dataset(s) being fine-tuned with and the dataset(s) being tested upon. These observations indicate some important research directions. Data in computer vision applications are highly varied and it is not very easy to capture its complexity and variations with sufficient ease. It is important to proceed to work on better

dataset usage by clustering the datasets together based on traits such as viewpoint, resolution etc. Resolution is another important element which significantly affects deep learning based approaches. This is because deep learning involves automated feature extractions from the pixel level and low resolution appearance often makes that problem difficult.
We intend to work upon and cover these issues in subsequent efforts towards solving the pedestrian detection problem.

6.7. Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in Musical Mnemotherapy

Participants: Antitza Dantcheva, Piotr Bilinski, Philippe Robert, François Brémond.

keywords: automated healthcare, healthcare monitoring, expression recognition

The elderly population has been growing dramatically and future predictions and estimations showcase that by 2050 the number of people over 65 years old will increase by 70%, the number of people over 80 years old will increase by 170%, outnumbering younger generations from 0-14 years. Other studies indicate that around half of the current population of over 75 year old suffer from physical and / or mental impairments and as a result are in need of high level of care. The loss of autonomy can be delayed by maintaining an active life style, which also would lead to reduced healthcare financial costs. With the expected increase of the world elderly population, and on the other hand limited available human resources for care a question arises as "How can we improve health care in an efficient and cost effective manner?".

Motivated by the above, we propose an approach for detecting facial expressions in Alzheimer's disease (AD) patients that can be a pertinent unit in an automated assisted living system for elderly subjects. Specifically, we have collected video-data of AD patients in musical therapy at the AD center Fondation G.S.F J. L. Noisiez in Biot, France from multiple therapy-sessions for validating our method. We note that in such sessions even AD patients suffering from apathy exhibit a number of emotions and expressions. We propose a spatio-temporal algorithm for facial expression recognition based on dense trajectories, Fisher Vectors and support vector machine classification. We compared the proposed algorithm to a facial-landmark-based algorithm concerning signal displacement of tracked points within the face.

Our algorithm differentiates between four different facial expressions: (i) neutral, (ii) smile, (iii) talking, and (iv) singing with an accuracy of 56%, outperforming the facial-landmark-based algorithm. Challenging for both algorithms has been the unconstrained setting involving different poses, changes in illumination and camera movement. One expected benefit for AD patients is that positive expressions and their cause could be determined and replicated in order to increase life standard for such patients, which also brings to the fore a delay in the development of AD (see figure 11).

This work is published in the Gerontolgy Journal.

6.8. Hybrid Approaches for Gender Estimation

Participants: Antitza Dantcheva, Piotr Bilinski.

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. Most commonly, the underlying algorithms analyze facial appearance for clues of gender.
Can a smile reveal your gender? [28], [35]

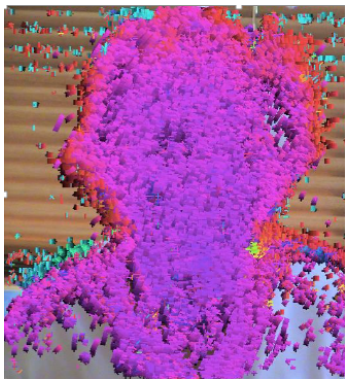

Figure 11. Expression recognition in AD patients based on dense trajectories and Fisher vectors. Dense trajectories visualization.

Deviating from such algorithms in [28] we proposed a novel method for gender estimation, exploiting dynamic features gleaned from smiles and show that (a) facial dynamics incorporate gender clues, and (b) that while for adults appearance features are more accurate than dynamic features, for subjects under 18, 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 obtained results (see Table 9) show that smile-dynamic include pertinent and complementary to appearance gender information. Such an approach is instrumental in cases of (a) omitted appearance-information (*e.g.* low resolution due to poor acquisition), (b) gender spoofing (*e.g.* makeup-based face alteration), as well as can be utilized to (c) improve the performance of appearance-based algorithms, since it provides complementary information.

Table 9. True gender classification rates. Age given in years.

Age	< 20	> 19
Subj. amount	143	214
OpenBR	52.45%	78.04%
Dynamics (SVM+PCA) [28]	60.1%	69.2%
Dynamics (AdaBoost) [28]	59.4%	61.7%
OpenBR + Dynamics (Bagged Trees) [28]	60.8%	80.8%
Motion-based descriptors [35]	77.7%	80.11%
Improved dynamics [35]	**86.3%**	**91.01%**

We improve upon the above work by proposing a spatio-temporal features based on dense trajectories, represented by a set of descriptors encoded by Fisher Vectors [35]. Our results suggest that smile-based features include significant gender-clues. The designed algorithm obtains true gender classification rates of 86.3% for adolescents, significantly outperforming two state-of-the-art appearance-based algorithms (*OpenBR*

and *how-old.net*), while for adults we obtain true gender classification rates of 91.01%, which is comparably discriminative to the better of these appearance-based algorithms (see Table 9).

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

In this work [36] we studied gender estimation based on information deduced jointly from face and body, extracted from single-shot images. The approach addressed challenging settings such as low-resolution-images, as well as settings when faces were occluded. Specifically the face-based features included local binary patterns (LBP) and scale-invariant feature transform (SIFT) features, projected into a PCA space. The features of the novel body-based algorithm proposed in this work included continuous shape information extracted from body silhouettes and texture information retained by HOG descriptors. Support Vector Machines (SVMs) were used for classification for body and face features. We conduct experiments on images extracted from video-sequences of the Multi-Biometric Tunnel database, emphasizing on three distance-settings: close, medium and far, ranging from full body exposure (far setting) to head and shoulders exposure (close setting). The experiments suggested that while face-based gender estimation performs best in the close-distance-setting, body-based gender estimation performs best when a large part of the body is visible. Finally we presented two score-level-fusion schemes of face and body-based features, outperforming the two individual modalities in most cases (see Table10 and Table 11).

Table 10. Performance (%) of the Face Gender Estimation algorithm (FGE) and the Body Gender Estimation algorithm (BGE).

Distance	FGE			BGE		
Scenario	Male TPR	Fem. TPR	Acc.	Male TPR	Fem. TPR	Acc.
Far	94.28	20	57.14	87.14	88.57	87.85
Medium	71.42	90	80.71	85.71	87.14	86.42
Close	88.57	90	89.28	78.57	80	79.28

Table 11. Performance (%) of the Sum fusion and Smarter Sum Fusion of FGE and BGE in terms of True Positive Rate (TPR) for Male and Female (Fem.), overall Accuracy (Acc.). Best performance (in terms of Acc.) of each distance-setting is bolded.

Distance	Sum Fusion			Prop. Sum Fusion		
Scenario	Male TPR	Fem TPR	Acc.	Male TPR	Fem TPR	Acc.
Far	87.14	88.57	**87.85**	87.14	88.57	**87.85**
Medium	88.57	90	**89.28**	88.57	90	**89.28**
Close	87.14	88.57	87.85	92.85	94.28	**93.57**

6.9. Unsupervised Metric Learning for Multi-shot Person Re-identification

Participants: Furqan Khan, François Brémond.

keywords: re-identification, long term visual tracking, metric learning, unsupervised labeling

Automatic label generation for metric learning

Appearance based person re-identification is a challenging task, specially due to difficulty in capturing high intra-person appearance variance across cameras when inter-person similarity is also high. Metric learning is often used to address deficiency of low-level features by learning view specific re-identification models. The models are often acquired using a supervised algorithm. This is not practical for real-world surveillance systems because annotation effort is view dependent. Therefore, everytime a camera is replaced or added, a significant amount of data has to be annotated again. We propose a strategy to automatically generate labels for person tracks to learn similarity metric for multi-shot person re-identification task. Specifically, we use the fact that non-matching (negative) pairs far out-number matching (positive) pairs in any training set. Therefore, the true class conditional probability of distance given negative class can be estimated using the empirical marginal

distribution of distance. This distribution can be used to sample non-matching person pairs for metric learning. A brief overview of the approach is presented below, please refer to [33] for details.

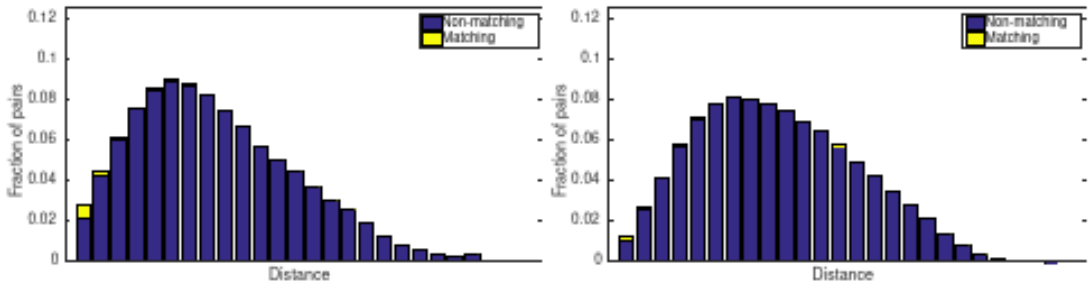

Figure 12. Distributions of distances between pairs of signature of randomly selected half of a) PRID, and b) iLIDS-VID datasets for MCM representation using Euclidean distance. The distributions are averaged for 10 trials.

In figure 12, empirical distribution of Euclidean distance (using MCM [43] representation) is plotted for two publicly available datasets. It can be noted that the positive samples lie on one side of distribution mode. Therefore, negative pairs can be sampled according to the probability proportional to the signed distance from the mode. Practically, we only select sample pairs that are farthest away in the distribution as negative pairs. For positive pairs, we use the fact that each track has more than one image for a person. Thus we generate positive pairs using the persons selected for negative pairs. We evaluated our approach on three publicly available datasets in multi-shot settings: iLIDS-VID, PRID and iLIDS-AA. Performance comparison of different representations using recognition rates at rank r are detailed in table 12, table 13 and table 14. Our results validate the effectiveness of our approach by considerably reducing the performance gap between fully-supervised models using KISSME algorithm and Euclidean distance.

Table 12. PRID

Method	r=1	r=5	r=10	r=20
MCM+MPD	53.6	83.1	91.0	96.9
MCM+UnKISSME	59.2	81.7	90.6	96.1
MCM+KISSME	64.3	86.1	94.5	98.0

Table 13. iLIDA-VID

Method	r=1	r=5	r=10	r=20
MCM+MPD	34.3	61.5	74.4	83.3
MCM+UnKISSME	38.2	65.7	75.9	84.1
MCM+KISSME	40.3	69.9	79.0	87.5

6.10. Semi-supervised Understanding of Complex Activities in Large-scale Datasets

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

keywords: Semi-supervised methods, activity understanding, probabilistic models, pairwise graphs

Table 14. iLIDS-AA

Method	r=1	r=5	r=10	r=20
MCM+MPD	56.5	79.7	90.9	95.2
MCM+UnKISSME	61.2	85.1	92.8	96.0
MCM+KISSME	62.9	84.7	93.4	97.0

Informations

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 complex activities - compositions of actions and scene objects - is still an open problem due to the complex temporal and composite structure of this category of events. Existing methods focus either on simple activities or oversimplify the modeling of complex activities by targeting only whole-part relations between its sub-parts (*e.g.*, actions). We study a semi-supervised approach (Fig. 13) that can learn complex activities from the temporal patterns of concept compositions in different arities (*e.g.*, "slicing-tomato" before "pouring_into-pan"). So far, our semi-supervised, probabilistic model using pairwise relations both in compositional and temporal axis outperforms prior work by 6 % (59% against 53%, mean Average precision, Fig. 14). Our method also stands out from the competition by its capability to handle relation learning in a setting with large number of video sequences (*e.g.*, 256) and distinct concept classes (Cooking Composite dataset, 218 classes, [90]), an ability that current state-of-the-art methods lack. Our initial achievements in this line of research has been published in [31]. Further work will focus on learning relations of higher arity.

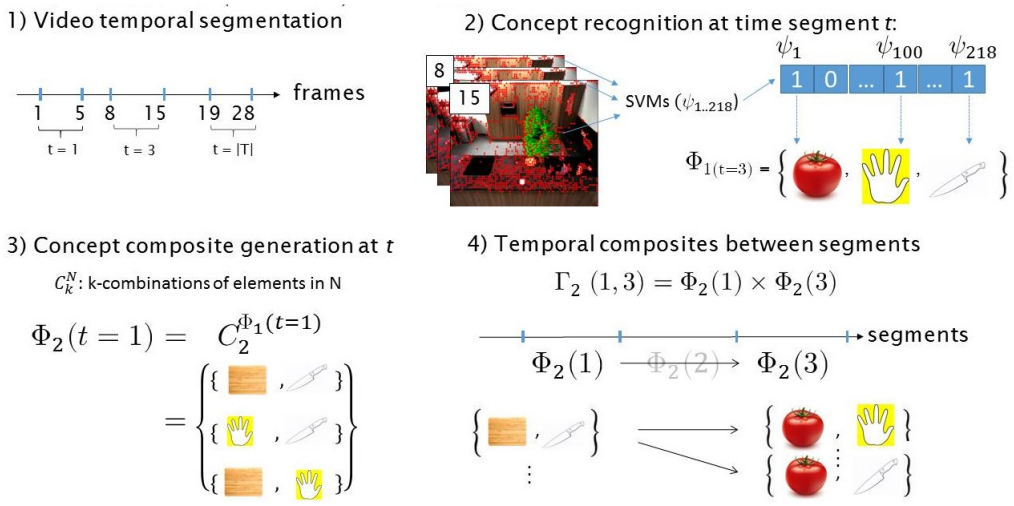

Figure 13. Semi-supervised learning of a video representation: 1) video temporal segmentation, 2) concept recognition 3) composite concept generation per time segment, 4) Temporal composite generation between segments.

6.11. On the Study of the Visual Behavioral Roots of Alzheimer's disease

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

Keywords: Activities of Daily Living, Dementia prediction, RGBD sensors, Activity Recognition, Cognitive Health

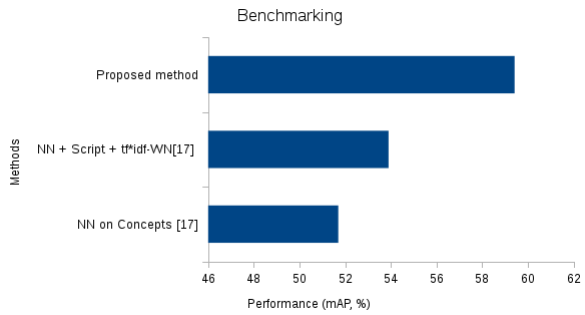

*Figure 14. Performance benchmarking of our approach against data set baselines: a) Nearest Neighbor classifier (NN) on concepts, script data, and tf*idf-WN, and b) NN only on concepts.*

Existing computer vision studies for the diagnosis of Dementia have focused on extracting discriminative patterns between healthy and people with dementia from neuroimagery exams, like functional MRI and PET scans. Nonetheless, the effects of dementia over human behaviors are a discriminative component that is barely explored by automatic vision-based methods. We studied a framework to automatically recognize the cognitive health of seniors from the visual observation of their activities of daily living (Fig.16). We employ a lightweight activity recognition system based on RGBD sensors to recognize the set of target activities (*e.g.*, prepare drink, prepare medication, make a payment transaction) performed by a person in a continuous video stream. Then, we summarize the absolute and relative activity patterns present in the video sequence using a novel probabilistic representation of activity patterns. Finally, this representation serves as input to Random Forest classifiers to predict the class of cognitive health that the person in question belongs to. We demonstrate that with the current framework can recognized the cognitive health status of seniors (*e.g.*, healthy, Mild Cognitive Impairment and Alzheimer's disease) with an average F_1-score of 69 % in real life scenarios.

6.12. Uncertainty Modeling with Ontological Models and Probabilistic Logic Programming

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

keywords: probabilistic logic programming, activities of daily living, senior monitoring, ontological models,

We have been investigating novel probabilistic, knowledge-driven formalisms that can join the representation expressiveness of an ontology-based language with the probabilistic reasoning of probabilistic graphical models, like probabilistic graphical models and probabilistic programming languages. The goal is to support the representation of events (entities, sub-events and constraints) and hierarchical structures (event, sub-events) and at the same time be capable of handling uncertainty related to both entity/sub-event detection and soft constraints. Prior work in probabilistic logic provides support to reasoning either about uncertainty related to entity recognition (probability of entity x in the scene defined in ProbLog2) or to soft-constraint (relevance of violation of constraint i to model y as defined in Markov Logic). In our current work in partnership with KU university of Leuven, we have extended the ontological models of our vision pipeline (Fig.17) with probabilistic logic formalism proposed by ProbLog (Fig.18), a probabilistic logic programming language. Current results on the recognition of daily activities of seniors are promising as they improved the precision of our prior method by 1%. Further work will focus on extending our uncertainty models to be robust to constraint violations.

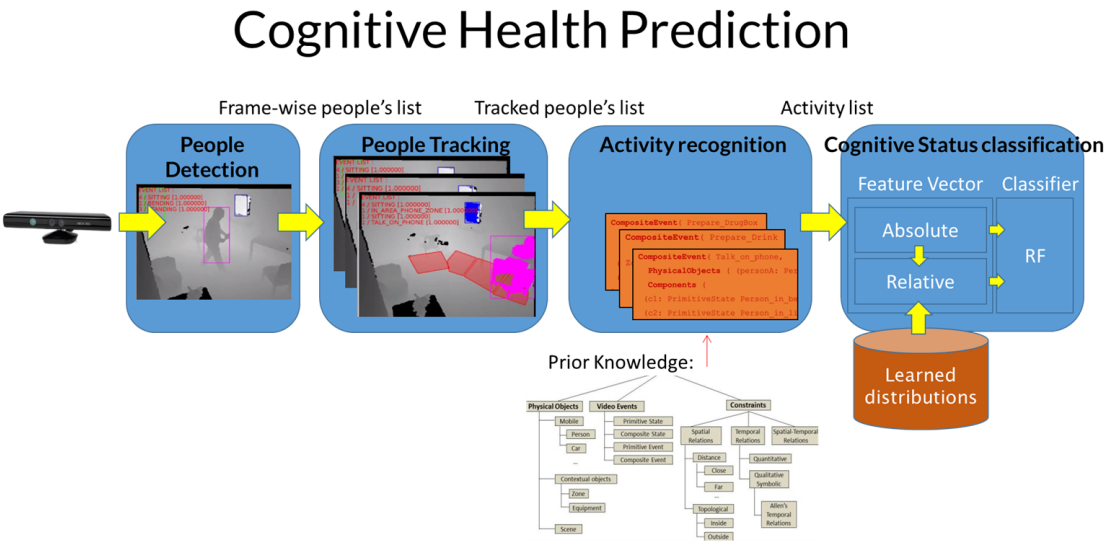

Figure 15. Automatic framework for visual recognition of cognitive health status: visual event recognition is responsible to detect and track people in the scene and recognize their events based on spatio-temporal relations with scene objects. Cognitive health classification represents absolute and relative information about the target classes.

Figure 16. Monitoring a senior performing at a gait-related event

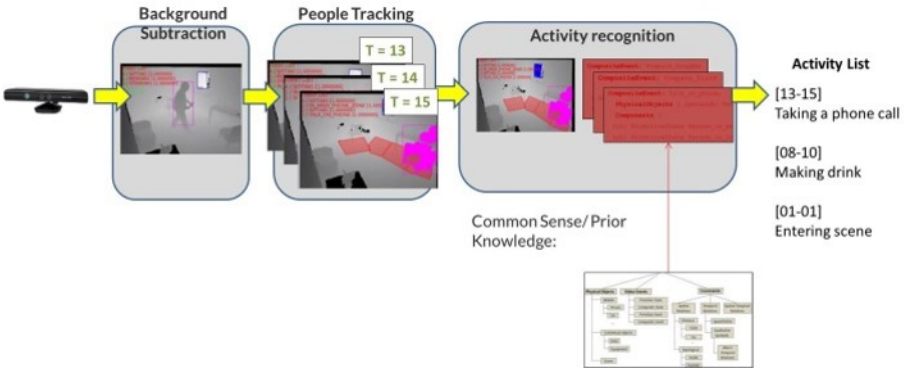

Figure 17. Pipeline for online activity recognition: given an acquisition camera (e.g. a Kinect), it firstly detects people using background subtraction algorithm, then it looks for appearance correspondence between people detected in the current frame with respect to past detections (past-present approach), and thirdly it recognizes the activities performed by each of the tracked people.

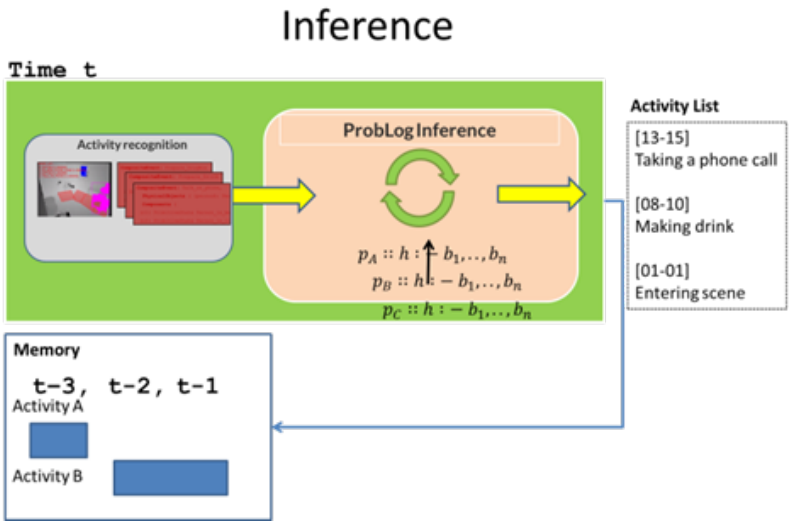

Figure 18. Temporal Inference using ProbLog engine. It takes as input deterministic observations and frame-wisely it recognizes the target events. Frame-events are aggregated into time intervals to create the time intervals of complex activities.

6.13. A Hybrid Framework for Online Recognition of Activities of Daily Living In Real-World Settings

Participants: Farhood Negin, Serhan Cosar, Michal Koperski, Carlos Crispim, Konstantinos Avgerinakis, François Brémond.

keywords: Supervised and Unsupervised Learning, Activity Recognition

State-of-the-art and Current Challenges

Recognizing human actions from videos has been an active research area for the last two decades. With many application areas, such as surveillance, smart environments and video games, human activity recognition is an important task involving computer vision and machine learning. Not only the problems related to image acquisition, e.g., camera view, lighting conditions, but also the complex structure of human activities makes activity recognition a very challenging problem. Traditionally, there are two variants of approach to cope with these challenges: supervised and unsupervised methods. Supervised approaches are suitable for recognizing short-term actions. For training, these approaches require a huge amount of user interaction to obtain well-clipped videos that only include a single action. However, Activities of Daily Living (ADL) consists of many simple actions which form a complex activity. Therefore, the representation in supervised approaches are insufficient to model these activities and a training set of clipped videos for ADL cannot cover all the variations. In addition, since these methods require manually clipped videos, they can only follow an offline recognition scheme. On the other hand, unsupervised approaches are strong in finding spatio-temporal patterns of motion. However, the global motion patterns are not enough to obtain a precise classification of ADL. For long-term activities, there are many unsupervised approaches that model global motion patterns and detect abnormal events by finding the trajectories that do not fit in the pattern [70], [83]. Many methods have been applied on traffic surveillance videos to learn the regular traffic dynamics (e.g. cars passing a cross road) and detect abnormal patterns (e.g. a pedestrian crossing the road) [71].

Proposed Method

We propose a hybrid method to exploit the benefits of both approaches. With limited user interaction our framework recognizes more precise activities compared to available approaches. We use the term precise to indicate that, unlike most of trajectory-based approaches which cannot distinguish between activities under same region, our approach can be more sensitive in the detection of activities thanks to local motion patterns. We can summarize the contributions of this work as following: i) online recognition of activities by automatic clipping of long-term videos and ii) obtaining a comprehensive representation of human activities with high discriminative power and localization capability.

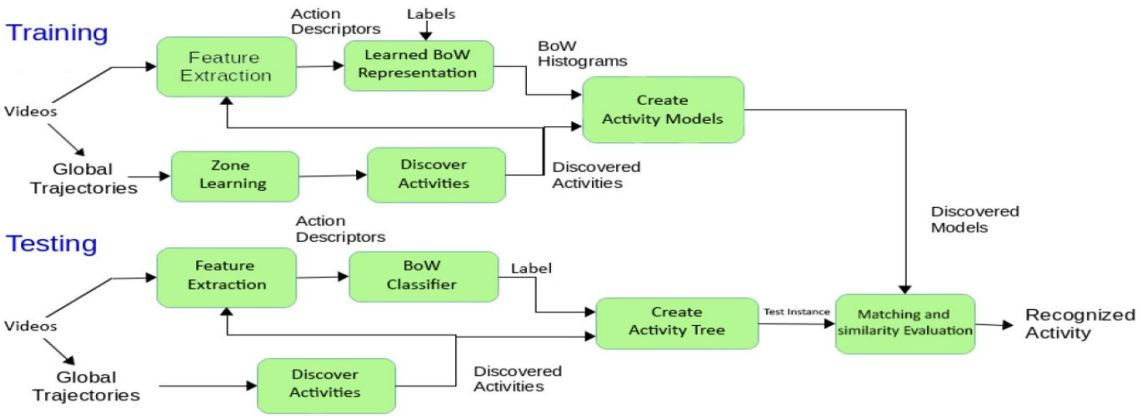

Figure 19. Architecture of the framework: Training and Testing phases

Figure 19 illustrates the flow of the training and testing phases in the proposed framework. For the training phase, the algorithm learns relevant zones in the scene and generates activity models for each zone by complementing the models with information such as duration distribution and BoW representations of discovered activities. At testing, the algorithm compares the test instances with the generated activity models and infers the most similar model.

The performance of the proposed approach has been tested on the public GAADRD dataset [73] and CHU dataset. Our approach always performs equally or better than online supervised approach in [99] (see Table15 and Table16). And even most of the time it outperforms totally supervised approach (manually clipped) of [99]. This reveals the effectiveness of our hybrid technique where combining information coming from both constituents could contribute to enhance recognition. The paper of this work was accepted in AVSS 2016 conference [30].

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

	Supervised (Manually Clipped) of [99]		Online Version of [99]		Unsupervised Using Global Motion [66]		Proposed Approach	
ADLs	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)
Answering Phone	57	78	**100**	**86**	**100**	60	**100**	81.82
P. Tea + W. Plant	89	**86.5**	76	38	84.21	80	**94.73**	81.81
Using Phar. Basket	**100**	83	**100**	43	90	**100**	**100**	**100**
Reading	35	**100**	92	36	81.82	**100**	**100**	91.67
Using Bus Map	90	**90**	**100**	50	**100**	54.54	**100**	83.34
AVERAGE	74.2	87.5	93.6	50.6	91.2	78.9	**98.94**	**87.72**

6.14. Praxis and Gesture Recognition

Participants: Farhood Negin, Jeremy Bourgeois, Emmanuelle Chapoulie, Philippe Robert, François Brémond.

keywords: Gesture Recognition, Dynamic and Static Gesture, Alzheimer Disease, Reaction Time, Motion Descriptors

Challenges and Proposed Method

Most of the developed societies are experiencing an aging trend of their population. Aging is correlated with cognitive impairment such as dementia and its most common type: Alzheimer's disease. So, there is an urgent need to develop technological tools to help doctors to do early and precise diagnoses of cognitive decline. Inability to correctly perform purposeful skilled movements with hands and other forelimbs most commonly is associated with Alzheimerâs disease [84]. These patients have difficulty to correctly imitate hand gestures and mime tool use, e.g. pretend to brush one's hair. They make spatial and temporal errors. We propose a gesture recognition and evaluation framework as a complementary tool to help doctors to spot symptoms of cognitive impairment at its early stages. It is also useful to assess one's cognitive status. First, the algorithm classifies the defined gestures in the gestures set and then it evaluates gestures of the same category to see how well they perform compared to correct gesture templates. Methods Shape and motion descriptors such as HOG (histogram of oriented gradient) [61] and HOF (histogram of optical flow) [62] are an efficient clue to characterize different gestures (Figure 20 Left). Extracted descriptors are utilized as input to train the

Table 16. The activity recognition results for GAADRD dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised (Manually Clipped) Approach [99]		Online Version of [99]		Classification by detection using SSBD [49]		Unsupervised Using Global Motion [66]		Proposed Approach	
ADLs	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)	Recall (%)	Prec. (%)
Answering Phone	**100**	88	**100**	70	96	34.29	**100**	**100**	**100**	88
Establish Acc. Bal.	67	**100**	**100**	29	41.67	41.67	**100**	86	67	**100**
Preparing Drink	**100**	69	**100**	69	96	80	78	**100**	**100**	82
Prepare Drug Box	58.33	**100**	11	20	**86.96**	51.28	33.34	**100**	22.0	**100**
Watering Plant	54.54	**100**	0	0	**86.36**	86.36	44.45	57	44.45	**80**
Reading	**100**	**100**	88	37	**100**	31.88	**100**	**100**	**100**	**100**
Turn On Radio	60	86	**100**	75	96.55	19.86	89	89	89	89
AVERAGE	77.12	**91.85**	71.29	42.86	**86.22**	49.33	77.71	90.29	74.57	91.29

classifiers. We use bag-of-visual-words approach to characterize gestures with descriptors. The classification happens in two steps: first we train a classifier to distinguish different gestures and after, we train another classifier with correct and incorrect samples of the same class. This way, we could recognize which gesture is performed and whether it is performed accurately or not.

Experiments and Results

The framework is fed by input data which come from a depth sensor (Kinect, Microsoft). At first phase, the correct samples of gestures performed by clinicians, are recorded. We train the framework using correct instances of each gesture class. In the second phase, participants were asked to perform the gestures. We use virtual reality as modality to interact with subjects to make the experiments more immersive and realistic experience. First an avatar performs a specific gesture and then she asks the subject to repeat the same gesture (Figure 20 Right). In this work, we analyze two categories of gestures. First category is dynamic gestures where the whole motion of the hands is considered as a complete gesture. Second category of gestures is static gestures where only a static pose of hands is the desired gesture. For static gestures, we also need to detect this key frame. Moreover, reaction time which starts after avatar asked the subject to do the gesture, until subject really starts to perform the gesture, could be an important diagnostic factor. Our algorithm uses motion descriptors to detect key frames and reaction time. In the preliminary tests, our framework successfully recognized more than 80% of the dynamic gestures. It also detects key frames and reaction time with a high precision. Thus the proposed gesture recognition framework helps doctors by providing a complete assessment of gestures performed by subject.

This work is published in [30] and will appear in the Gerontechnology Journal.

6.15. Scenario Recognition

Participants: Inès Sarray, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault.

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

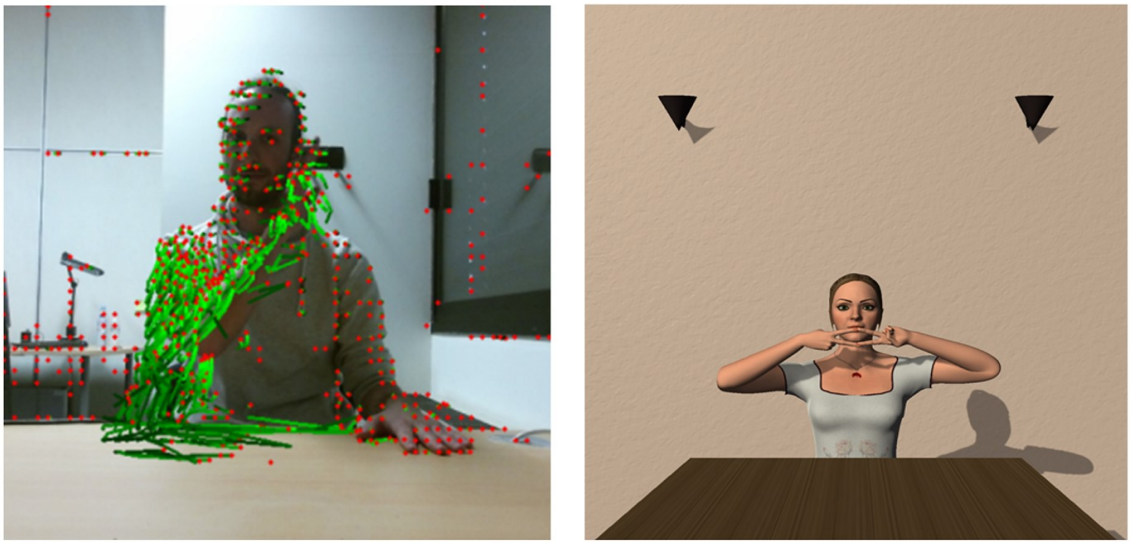

Figure 20. Left: Extracted motion descriptors while performing a gesture Right: virtual avatar guides patients in a virtual reality environment

Activity recognition systems aim at recognizing the intentions and activities of one or more persons in real life, by analyzing their actions and the evolution of the environment. This is done thanks to a 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).Stars has been working to ameliorate and facilitate the generation of these activity recognition systems. As we can use these systems in a big range of important fields, we propose a generic approach to design activity recognition engine. These engines should continuously and repeatedly interact with their environment and react to its stimuli. On the other hand, we should take into consideration the dependability of these engines which is very important to avoid possible safety issue, that's why we need also to rely on formal methods that allow us to verify these engines behavior. Synchronous modeling is a solution that allows us to create formal models that describe clearly the system behavior and its reactions when it detects different stimuli. Using these formal models, we can build effective recognition engines for each formal model and validate them easily using model checking. This year, we adapted this approach to create a new simple scenario language to express the scenario behaviors and to automatically generate its recognition automata at compile time. This automata will be embedded into the recognition engine at runtime.

Scenario description Language

As we work with non-computer-science end-users, we need a friendly description language that helps them to express easily their scenarios. To this aim, we collaborated with Ludotic ergonomists to define the easiest way for a simple user to deal with the new language. Using AxureRP tool, we defined two types of language:
1- Textual language:

For the textual language, we decided to use a simple language. Using 9 operators, and after the definition of the types, roles, and sub-scenarios, the user can describe a scenario in a simple way, such as in figure21.

This year, we implemented this textual language and it is under testing.
2)- Graphical language:

```
Type Personne, Equipement,Zone;

Scenario coupTel :

role
        Patient:Personne;
        Tel: Equipement;
        table: Equipement;
        sejour: Zone;

Subscenarios
        entend(Personne, Equipement);
        decroche(Personne, Equipement);
        commence_a_parler(Personne);
        finit_de_parler(Personne);
        raccroche(Personne,Equipement);

EtatInitial : dans_Zone(Patient, sejour);

debut

   pres_de(Patient, table) parallele entend(Patient, Tel)
puis
   decroche(Patient, Tel)
puis
   commence_a_parler(Patient)
puis
   finit_de_parler(Patient)
puis
   raccroche(Patient, Tel)
puis
      Alert (fin_de_scenario)

fin
```

Figure 21. Example of the textual language

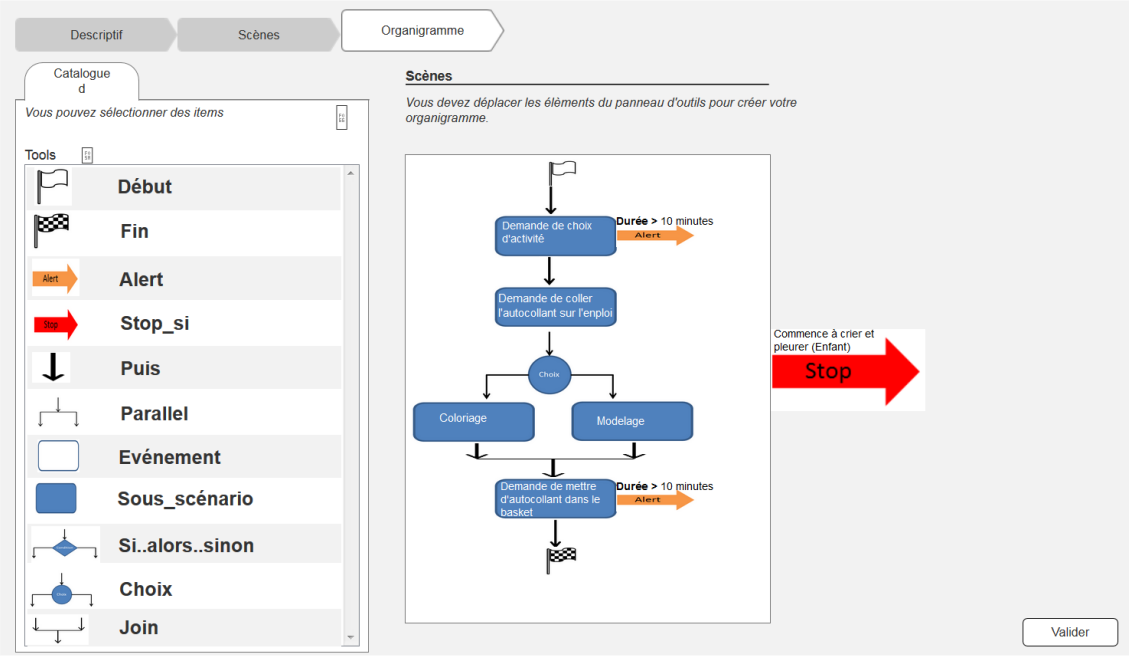

Figure 22. Generic flowchart

The graphical language model has 3 basic interfaces: The first interface allows the user to define the types, roles, and the initial state of the scenario. The second one is dedicated to describe the sub-scenarios and to express simple scenarios using a timeline. In case of complicated scenarios, the third interface offers users a tool panel that allows them to describe their scenarios in a hierarchical way using a flowchart-like representation (see figure 22).

Recognition Automata

This year, we worked also on recognition automata generation. We used the synchronous modeling and semantics to define these engines. The semantics consists in a set of formal rules that describe the behavior of a program. We specified first the language operators: we rely on a 4-valued algebra with a bilattice structure to define two semantics for the recognition engine: a behavioral and equational one. A behavioral semantics defines the behavior of a program and its operators and gives it a clear interpretation. Equational semantics allows us to make a modular compilation of our programs using rules that translate each program into an equation system. After defining these two semantics, we verified their equivalence for all operators, by proving that these semantics agree on both the set of emitted signals and the termination value for a program P. We implemented these semantics and we are now working on the automatic generation of the recognition automata.

6.16. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé.

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 around the language (see Figure 23) to design, simulate, verify, and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality.

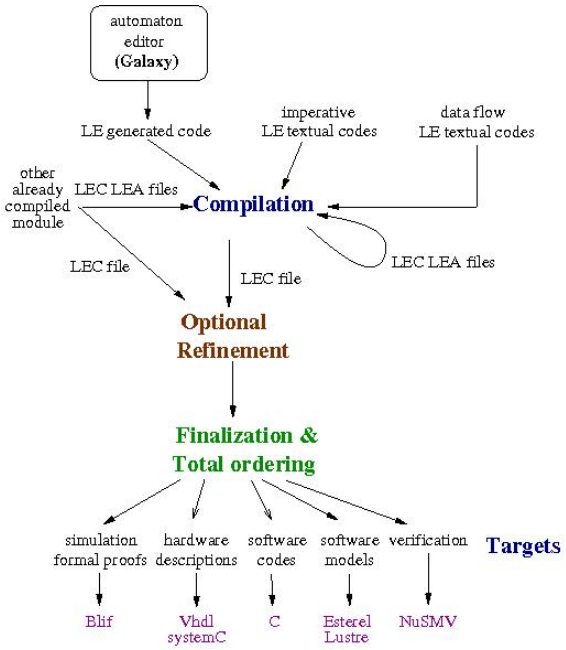

Figure 23. The Clem Toolkit

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. We also designed a large application: a mechatronics system in CLEM and we have proved that its main safety properties hold in our modeling. Now, to complete the improvement done these two last years concerning data handling, we want to extend the verification side of CLEM. To this aim, this year we began to replace the fundamental representation of Boolean values as BDD (Binary Decision Diagrams) with LDD (Logical Decision Diagrams), which allow to encode integer values in a very efficient way. It turns out that the validation mechanism of CLEM could take into account properties over integer data. However, this is a first test and the integration of a model checking technique in CLEM remains a challenge.

6.17. Safe Composition in Middleware for Internet of Things

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

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

The main concern of this research axis is the dependability of a component-based adaptive middleware which dynamically adapt and recompose assemblies of web components. Such a middleware plays an important role in the generation of event recognition engines we are currently building in Stars team (see section 6.15). One of the main challenge is how to guarantee and validate some safety and integrity properties throughout the system's evolution. These two last years, we have proposed to rely on synchronous models to represent component behavior and their composition and to verify that these compositions verify some constraints during the dynamic adaptation to appearance and disappearance of components. We defined a generic way to express these constraints and we proposed the Description Constraint Language (DCL) to express these constraints. Hence, we compile them into LE programs (see 6.16) and we benefit from CLEM model checking facilities to ensure that they are respected [93]. This year, we improved the DCL language in order to take into account both the dynamic variation of components and also applications which use these components and we are currently

testing the efficiency of our method to add and remove components. Moreover, genericity is expressed by the notion of type and we aim at extending this notion to a thinner representation of knowledge about components.

6.18. Verification of Temporal Properties of Neuronal Archetypes

Participants: Annie Ressouche, Daniel Gaffé.

Keywords: Synchronous Modeling, model-checking, lustre, temporal logic, biologic archetypes

This year, we began a collaboration with with the I3S CNRS laboratory and Jean Dieudonné CNRS laboratory to verify temporal properties of neuronal archetypes. There exist many ways to connect two, three or more neurons together to form different graphs. We call archetypes only the graphs whose properties can be associated to specific classes of biologically relevant structures and behaviors. These archetypes are supposed to be the basis of typical instances of neuronal information processing. To model different representative archetypes and express their temporal properties, we use a synchronous programming language dedicated to reactive systems (Lustre). Then, we generate several back ends to interface different model checkers supporting data types and automatically validate these properties. We compare the respective results. They mainly depend on the underlying abstraction methods used in model checkers.

These results are published in [32]

6.19. Dynamic Reconfiguration of Feature Models

Participants: Sabine Moisan, Jean-Paul Rigault.

Keywords: feature models, model at run time, self-adaptive systems

In video understanding systems, context changes (detected by system sensors) are often unexpected and can combine in unpredictable ways, making it difficult to determine in advance (off line) the running configuration suitable for each context combination. To address this issue, we keep, at run time, a model of the system and its context together with its current running configuration. We adopted an enriched Feature Model approach to express the variability of the architecture as well as of the context. A context change is transformed into a set of feature modifications (selection/deselection of features) to be processed on the fly. This year we proposed a fully automatic mechanism to compute at run time the impact of the current selection/deselection requests. First, the modifications are checked for consistency; second, they are applied as a single atomic "transaction" to the current configuration to obtain a new configuration compliant with the model; finally, the running system architecture is updated accordingly. This year we implemented the reconfiguration step and its algorithms and heuristics and we evaluated its run time efficiency.

Our ultimate goal is to control the system through a feed back loop from video components and sensor events to feature model manipulation and back to video components modifications.

The fully automatic adaptation that we propose is similar to a Feature Model editor. That is the reason why our previous attempt was to embed a general purpose feature model editor at run time. This revealed two major differences between our mechanism and an editor. First, in a fully automatic process there is no human being to drive a series of edits, hence heuristics are required. Second, the editor operations are often elementary while we need a global "transaction-like" application of all the selections/deselections to avoid temporary unconsistencies.

In order to evaluate our algorithm performance, we randomly generated feature models (from 60 to 1400 features). We also randomly generated context changes. The results are shown on figure 24: no processing time explosion is noticeable; in fact the time seems to grow rather linearly. Moreover, the computation time of a new initial partial configuration does not exceed 3ms for a rather big model. The algorithm and its evaluation are detailed in [41].

6.20. Setup and management of SafEE devices

Participants: Matias Marin, Etienne Corvée, François Brémond.

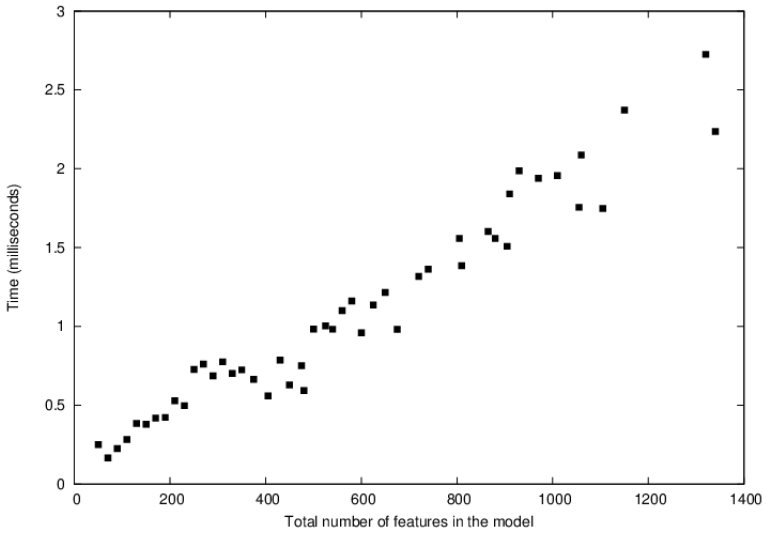

Figure 24. Computation time of initial models

The aim of the SafEE project (see section 8.1.1.2) is to provide assistance for the safety, autonomy and quality of life of elderly people at risk or already presenting Alzheimer's disease or related pathology.

Within EHPAD building (in Nice), 4 patients participated to our experiment and we plan to include more patients in the project throughout next years. Besides, 2 other patients have participated in the project at their own home.

More precisely, the SafEE project focuses on specific clinical targets: behavior, motricity, cognitive capabilities For this, the SafEE project includes:

- srvsafee(web server): a behavior analysis platform has been created to allow identification of certain daytime behavior disturbances (agitation, for example) and nocturnal disturbances (sleep disorders), locomotor capacities (walking and posture). It centralizes data saved in each local PC with Kinect2 sensor on the one hand, and postgresql database, on the other hand. About 30 Gb data are recorded for each patient in a day, which represents a huge amount to manage in the long run.
- Aroma diffuser (AromaCare): for sleep disturbances, using in particular an automated device for diffusing fragrances (aromatherapy) adapted to the perturbations detected by the analysis platform.
- Tablet (Serious game, MusicCare): for disturbances in spatial orientation, improved procedural memory and a sense of control and confidence in technological tools, using multimedia interfaces using an application for Android OS.
- Kinect2: motion detection for analysis linked to a PC, with a database to store recorded events.
- Bed sensor: able to track the sleep by analyzing the movements of the body, the breathing, and the beating of the heart.

Fig. 25 shows the Safee project environment.

6.21. Brick & Mortar Cookies

Participants: Julien Badie, Manikandan Bakthavatchalam, Vasanth Bathrinarayanan, Ghada Balhoul, Anais Ducoffe.

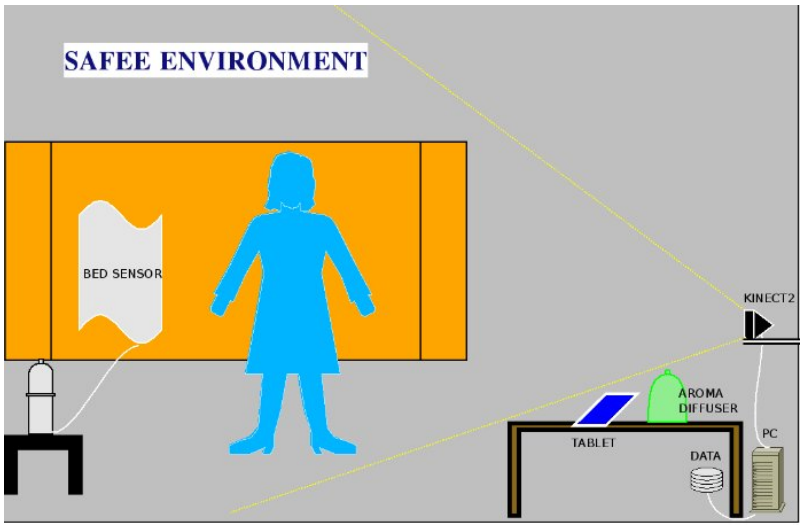

Figure 25. The Safee environment

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. The final system should be designed to be used without changing the current camera network of the customer store, dedicated to security purpose. Analysis should be given at different time and space resolutions. For instance, attendance of one particular day can be as interesting as attendance of the entire year. 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*. We are using a live connection to get metadata directly from the camera stream using a RTSP connection. Thie year we improved the results of last year using calibration tool embedded in the camera : shape of people detected was better, feet were followed with more precision as bounding boxes were more stable. We also tested the new IVA developed by BOSCH which was built to better manage changes in scene brightness and crossing of people. In the former version people close to each other were often detected as one person. Our first tests in shop revealed that it reduces the number of false detection but people were detected later than in the previous version. The case of people crossing doesn't seem to be better managed than before.

Inria algorithms : people detection and tracking

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. To improve results after some tests made during summer, the people detection is now using a deep learning method. People are detected earlier than before with this new algorithm and people crossing and occlusions are far better managed. The performances and the reliability of those algorithms were tested using an annotation tool developed in Stars Team.

Annotation tool

Manual annotation of videos requires major human effort. It can take hours and hours of fastidious work to annotate a tiny set of data. That's why we propose a semi-automatic tool which reduces the time of the annotation. This new semi automatic annotation tool uses a simple input data format, XML file or XGTF file to describe the video contents and algorithms output. Users only have to correct false or missing detection and to fix some wrong object id of the algorithms results using the annotation tool interface.

Tests in real conditions

We tested our video acquisition tool and our people detection and people tracking algorithms during summer in a partner supermarket in Nice. We successfully acquire 2 weeks of the desired metadata. By the end of summer, our results were highly improved by using a deep learning method to detect people. Moreover we can get results in quasi real-time. Except for the video stream acquisition tool, which needs to be connected to the camera network, our system is now running on an independent and local network. In case there is a crash of our system, the supermarket network will not be affected. Moreover, sensitive data are protected. A test is starting soon in SuperU to run and evaluate this new prototype.

Metadata storage in database

Last year metadata outputs of our analysis were first stored in XML files. Now to manage the quasi real-time solution, metadata are stored directly in the database we designed last year. We improve architecture of this database to manage simultaneously several connections as the final solution is supposed to be composed of several servers which will manage several video streams at the same time.

Web interface (HIM)

The web graphic interface is in progress. User interactions were added and improved so that the interface should be more user-friendly. We also changed some charts and tables so that statistical results should be better understood by users.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

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

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR

8.1.1.1. MOVEMENT

Program: ANR CSOSG

Project acronym: MOVEMENT

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

Duration: January 2014-June 2017

Coordinator: MORPHO (FR)

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

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

8.1.1.2. SafEE

Program: ANR TESCAN

Project acronym: SafEE

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

Duration: December 2013-May 2017

Coordinator: CHU Nice

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

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

8.1.2. FUI

8.1.2.1. 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.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. CENTAUR

Title: Crowded ENvironments moniToring for Activity Understanding and Recognition

Programm: FP7

Duration: January 2013 - December 2016

Coordinator: Honeywell

Partners:

Ecole Polytechnique Federale de Lausanne (Switzerland)

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

Neovision Sro (Czech Republic)

Queen Mary University of London (United Kingdom)

Inria contact: François Bremond

'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 Neovison 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.'

8.3. International Initiatives

8.3.1. Inria International Labs

8.3.1.1. Informal International Partners

- **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.
- **Collaboration with U.S.A.:** Stars collaborates with the University of Southern California.
- **Collaboration with Europe:** Stars collaborates with Multitel in Belgium, the University of Kingston upon Thames UK, and the University of Bergen in Norway.

8.3.1.2. Other IIL projects

The ANR SafEE (see section 8.1.1.2) 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).

8.4. International Research Visitors

8.4.1. 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;
- Adlen Kerboua, University of Skikda, Algeria;
- Karel Krehnac, Neovision, Praha, Czech Republic;
- Jana Trojnova, Honeywell, Praha, Czech Republic;
- Luis Emiliano Sanchez, Rosario University, Argentina.

8.4.1.1. Internships

Seongro Yoon

Date: Apr 2016-Dec 2016

Institution: Korea Advanced Institute of Science and Technology, Daejeon, Korea

Supervisor: François Brémond

Yashas Annadani

Date: May2016-June 2016

Institution: National Institute Of Technology Karnataka, India

Supervisor: Carlos Fernando Crispim Junior

Chandraja Dharmana

Date: May 2016-June 2016

Institution: Birla Institute of Technology and Science, Pilani, Hyderbrad

Supervisor: Carlos Fernando Crispim Junior

Shanu Vashistha

Date: May 2016-June 2016

Institution: Indian Institute of Technology, Kanpur, India

Supervisor: Carlos Fernando Crispim Junior

Nairouz Mrabah

Date: Apr 2016-Sep 2016

Institution: National School of Computer Science (ENSI), Tunisia

Supervisor: Inès Sarray

Isabel Rayas

Date: June 2016-Dec 2016

Institution: Massachusetts Institute of Technology, USA

Supervisor: Farhood Negin

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

François Brémond was organizer of the ISG 2016, 10th World Conference of Gerontechnology, Nice,28th to 30th September 2016.

François Brémond was editor of the Crowd Understanding workshop, part of ECCV, Amsterdam, October 2016.

9.1.1.2. Member of the organizing committee

François Brémond was a member of the Management Committee and COST Action IC1307 in 2016.

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees

François Brémond was program committee member of the conferences and workshops: KSE 2016, PETS2016, MMM2017.

François Brémond was ACM Multimedia Area Chair for Multimedia and Vision, Amsterdam, 2016.

François Brémond was session chair of AVSS-16, Colorado Springs, USA, 2016.

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

Antitza Dantcheva was program committee member of the conference International Conference on Biometrics (ICB 2016), the CVPR Workshop ChaLearn Looking at People 2016 and the Healthcare Conference Workshop within the EAIInternational Conference on Pervasive Computing Technologies.

9.1.2.2. Reviewer

François Brémond was reviewer for the conferences : CVPR2016-7, ECCV2016, VOT2016, MARMI2016, WACV 2017.

Carlos Fernando Crispim Junior was reviewer for the conferences: International Conference on Intelligent Robot and Systems, IEEE International Conference on Robotics and Automaton, Brazilian Conference in Biomadical Engineering, AMBIANT Conference, Computer on the Beach.

9.1.3. Journal

9.1.3.1. Member of the editorial boards

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

9.1.3.2. Reviewer - Reviewing activities

François Brémond was reviewer for the journal revue Retraite et sociétéänd Medical Engineering & Physics.

Carlos Fernando Crispim Junior was reviewer for the journals: Pattern Recognition, Neurocomputing, Computer Vision and Image Understanding Journal, Computers in Biology and Medecine Journal, PLOS One Journal, Frontiers in Neuroscience, Sensors.

Antitza Dantcheva reviewed for the journals: IEEE Transactions on Information Forensics and Security (TIFS), Information Processing Letters, The Computer Journal, IET Biometrics, Multimedia Systems, International Journal for Information Management, Information Fusion (INFFUS), Sensors, Pattern Recognition.

9.1.4. Invited talks

François Brémond was invited by Prof. Ram Nevatia to give a talk on research initiatives and new directions in Video Understanding, USC, LA, USA 17 August 2016.

François Brémond was invited by Prof. Jonathan Ventura to give a talk on People detection, at the SLDP 2016 workshop of AVSS, Colorado Springs, USA,23 August 2016.

François Brémond was invited by Prof. William Robson Schwartz,Department of Computer Science, Federal University of Minas Gerais to give a talk on Video Analytic, at Video Surveillance workshop in Belo Horizonte-Brazil, 03 October 2016.

François Brémond was invited by Prof. William Robson Schwartz to give a talk on People Tracking, at SIBGRAPI 2016, Sao Paulo-Brazil, 05 October 2016.

François Brémond was invited by Prof. Cosimo Distante, Consiglio Nazionale delle Ricerche to give a talk on Activity Recognition, at ACIVS 2016, Lecce, Italy, 26 October 2016.

François Brémond was invited by Sebastien Ambellouis (IFSTTAR) to give a talk on Activity Monitoring, at IEEE IPAS 2016,Hammamet, Tunisia,5-7 November 2016.

Carlos Fernando Crispim Junior was invited to give a talk at the 1st Inter-lalex seminar "Smart Systems", Besançon, France, November 23rd 2016.

Carlos Fernando Crispim Junior was invited to make a presentation at PSI-VISICS seminar at ESAT department in KU Leuven University,October 24th 2016.

Carlos Fernando Crispim Junior was invited to make a presentation at Machine Learning seminar at Computer Science department of KU Leuven University, October 17th 2016.

Carlos Fernando Crispim Junior was invited to give a talk at ISG 2016 - Seminar European FP7 project Dem@care: Automatic Video Analysis for Diagnosis and Care, Nice, France, October 29th 2016.

Carlos Fernando Crispim Junior was invited speaker at the internal seminar of LAAS-CNRS, Toulouse, France, July3rd-4th.

Carlos Fernando Crispim Junior was invited speaker at CPUEX seminar, LABRI-CNRS, Bordeaux, France, February 2016.

9.1.5. Scientific expertise

François Brémond was expert for EU European Reference Network for Critical Infrastructure Protection (ERNCIP) - Video Anlaytics and surveillance Group, at European Commission's Joint Research Centre in Brussels in July 2016.

François Brémond was expert for the Foundation Médéric Alzheimer, for the doctoral fellowship selection, September 2016.

9.2. Teaching - Supervision - Juries

9.2.1. 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.

9.2.2. Supervision

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 2015, Sabine Moisan.

PhD in progress : Farhood Negin, People Detection for Activity Recognition using RGB-Depth Sensors, Jan 2015, François Brémond.

PhD in progress : Ujjwal Ujjwal, Pedestrian Detection to Dynamically Populate the Map of a Crossroad, Sep 2016, François Brémond.

9.2.3. Juries

François Brémond was jury member of the following PhD theses:

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

PhD, Romain Endelin, University of Montpellier, 2nd June 2016.

PhD, Jean-Charles Bricola, CMM, Mines ParisTech, Fontainebleau, 19 October 2016.

PhD, Salma Moujtahid, LIRIS-Equipe IMAGINE-INSA Lyon, 3 November 2016.

PhD, Marion Chevalier, Laboratoire d'Informatique de Paris 6, Thales Optronique S.A.S., Paris,2 December 2016.

9.3. Popularization

François Brémond was invited to give a talk at Conférence des métiers at International Lycée (CIV) in Sophia, January 2016.

François Brémond was interviewed by the agence Citizen Press February 2016.

François Brémond was invited to give a talk at the Artificial Intelligence Workshop Meeting Amadeus - Inria, June 2016.

François Brémond was invited to give a talk at la rencontre Inria Industrie Ed-Tech, 1 December 2016.

François Brémond has published an article in ERCIM news, December 2016.

10. Bibliography

Major publications by the team in recent years

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

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

[3] B. BOULAY, F. BREMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System*, in "Pattern Recognition Letter", 2006, vol. 27, n^{o} 15, pp. 1785-1796

[4] F. BRÉMOND, M. THONNAT. *Issues of Representing Context Illustrated by Video-surveillance Applications*, in "International Journal of Human-Computer Studies, Special Issue on Context", 1998, vol. 48, pp. 375-391

[5] G. CHARPIAT. *Learning Shape Metrics based on Deformations and Transport*, in "Proceedings of ICCV 2009 and its Workshops, Second Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA)", Kyoto, Japan, September 2009

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

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

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

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

[10] C. LIU, P. CHUNG, Y. CHUNG, M. THONNAT. *Understanding of Human Behaviors from Videos in Nursing Care Monitoring Systems*, in "Journal of High Speed Networks", 2007, vol. 16, pp. 91-103

[11] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", December 2004, vol. 16, n^{o} 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^{o} 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^{o} 4, pp. 501-507

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

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

[18] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, in "Software Engineering Research, Management and Applications", R. LEE (editor), Studies in Computational Intelligence, Springer, 2008, vol. 150, pp. 157-171, selected as one of the 17 best papers of SERA'08 conference

[19] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Revue des sciences et technologies de l'information, série Théorie et Science Informat ique", June 2011, vol. 4, n^{o} 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^{o} 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^{o} 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^{o} 2

Publications of the year

Articles in International Peer-Reviewed Journals

[27] C. F. CRISPIM-JUNIOR, V. BUSO, K. AVGERINAKIS, G. MEDITSKOS, A. BRIASSOULI, J. BENOIS-PINEAU, Y. KOMPATSIARIS, F. BREMOND. *Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2016, vol. 38, pp. 1598 - 1611 [*DOI :* 10.1109/TPAMI.2016.2537323], https://hal.inria.fr/hal-01399025

[28] A. DANTCHEVA, F. BRÉMOND. *Gender estimation based on smile-dynamics*, in "IEEE Transactions on Information Forensics and Security", 2016, 11 p. [*DOI :* 10.1109/TIFS.2016.2632070], https://hal.archives-ouvertes.fr/hal-01412408

Invited Conferences

[29] S. CHEN, F. BREMOND, H. NGUYEN, H. THOMAS. *Exploring Depth Information for Head Detection with Depth Images*, in "AVSS 2016 - 13th International Conference on Advanced Video and Signal-Based Surveillance", Colorado Springs, United States, August 2016, https://hal.inria.fr/hal-01414757

[30] F. F. NEGIN, S. COSAR, M. F. KOPERSKI, C. F. CRISPIM-JUNIOR, K. AVGERINAKIS, F. F. BREMOND. *A hybrid framework for online recognition of activities of daily living in real-world settings*, in "13th IEEE

International Conference on Advanced Video and Signal Based Surveillance - AVSS 2016", Colorado springs, United States, IEEE, August 2016 [*DOI :* 10.1109/AVSS.2016.7738021], https://hal.inria.fr/hal-01384710

International Conferences with Proceedings

[31] C. F. CRISPIM-JUNIOR, M. KOPERSKI, S. COSAR, F. BREMOND. *Semi-supervised understanding of complex activities from temporal concepts*, in "13th International Conference on Advanced Video and Signal-Based Surveillance", Colorado Springs, United States, August 2016, https://hal.inria.fr/hal-01398958

[32] E. DE MARIA, A. MUZY, D. GAFFÉ, A. RESSOUCHE, F. GRAMMONT. *Verification of Temporal Properties of Neuronal Archetypes Modeled as Synchronous Reactive Systems*, in "HSB 2016 - 5th International Workshop Hybrid Systems Biology", Grenoble, France, Lecture Notes in Bioinformatics series, October 2016, 15 p. [*DOI :* 10.1007/978-3-319-47151-8_7], https://hal.inria.fr/hal-01377288

[33] F. M. KHAN, F. BREMOND. *Unsupervised data association for metric learning in the context of multi-shot person re-identification*, in "Advance Video and Signal based Surveillance", Colorado Springs, United States, August 2016 [*DOI :* 10.1109/AVSS.2016.7738058], https://hal.inria.fr/hal-01400147

[34] F. NEGIN, J. BOURGEOIS, E. CHAPOULIE, P. ROBERT, F. BREMOND. *Praxis and Gesture Recognition*, in "The 10th World Conference of Gerontechnology (ISG 2016)", Nice, France, September 2016, https://hal.inria.fr/hal-01416372

Conferences without Proceedings

[35] P. BILINSKI, A. DANTCHEVA, F. BRÉMOND. *Can a smile reveal your gender?*, in "15th International Conference of the Biometrics Special Interest Group (BIOSIG 2016)", Darmstadt, Germany, September 2016, https://hal.archives-ouvertes.fr/hal-01387134

[36] E. GONZALEZ-SOSA, A. DANTCHEVA, R. VERA-RODRIGUEZ, J.-L. DUGELAY, F. BRÉMOND, J. FIERREZ. *Image-based Gender Estimation from Body and Face across Distances*, in "23rd International Conference on Pattern Recognition (ICPR 2016): "Image analysis and machine learning for scene understanding"", Cancun, Mexico, December 2016, https://hal.archives-ouvertes.fr/hal-01384324

[37] M. KOPERSKI, F. BREMOND. *Modeling Spatial Layout of Features for Real World Scenario RGB-D Action Recognition*, in "AVSS 2016", Colorado Springs, United States, August 2016, pp. 44 - 50 [*DOI :* 10.1109/AVSS.2016.7738023], https://hal.inria.fr/hal-01399037

[38] M. K. PHAN TRAN, P. ROBERT, F. BREMOND. *A Virtual Agent for enhancing performance and engagement of older people with dementia in Serious Games*, in "Workshop Artificial Compagnon-Affect-Interaction 2016", Brest, France, June 2016, https://hal.archives-ouvertes.fr/hal-01369878

[39] N. THI LAN ANH, F. BREMOND, J. TROJANOVA. *Multi-Object Tracking of Pedestrian Driven by Context*, in "Advance Video and Signal-based Surveillance", Colorado Springs, United States, IEEE, August 2016, https://hal.inria.fr/hal-01383186

Research Reports

[40] E. DE MARIA, A. MUZY, D. GAFFÉ, A. RESSOUCHE, F. GRAMMONT. *Verification of Temporal Properties of Neuronal Archetypes Using Synchronous Models*, UCA, Inria ; UCA, I3S ; UCA, LEAT ; UCA, LJAD, July 2016, n^{o} RR-8937, 21 p. , https://hal.inria.fr/hal-01349019

[41] S. MOISAN, J.-P. RIGAULT. *Dynamic Reconfiguration of Feature Models: an Algorithm and its Evaluation*, Inria Sophia Antipolis, November 2016, n^{o} RR-8972, 16 p. , https://hal.inria.fr/hal-01392796

Other Publications

[42] C. F. CRISPIM-JUNIOR, A. KONIG, R. DAVID, P. ROBERT, F. BREMOND. *Automatic prediction of autonomy in activities of daily living of older adults*, November 2016, 74s p. , Short-paper, https://hal.inria.fr/hal-01399259

[43] F. M. KHAN, F. M. BRÉMOND. *Person Re-identification for Real-world Surveillance Systems*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01399939

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. 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

[46] A. ALAHI, L. JACQUES, Y. BOURSIER, P. VANDERGHEYNST. *Sparsity-driven people localization algorithm: Evaluation in crowded scenes environments*, in "PETS workshop", 2009

[47] A. ANDRIYENKO, K. SCHINDLER. *Multi-target tracking by continuous energy minimization*, in "Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on", June 2011, pp. 1265-1272 [*DOI :* 10.1109/CVPR.2011.5995311]

[48] D. ARSIC, A. LYUTSKANOV, G. RIGOLL, B. KWOLEK. *Multi-camera person tracking applying a graph-cuts based foreground segmentation in a homography framework*, in "PETS workshop", 2009

[49] K. AVGERINAKIS, A. BRIASSOULI, I. KOMPATSIARIS. *Activity detection using sequential statistical boundary detection (ssbd)*, in "to appear in Computer Vision and Image Understanding", CVIU, 2015

[50] 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

[51] A. BAR-HILLEL, D. LEVI, E. KRUPKA, C. GOLDBERG. *Part-based feature synthesis for human detection*, in "ECCV", 2010

[52] H. BEN SHITRIT, J. BERCLAZ, F. FLEURET, P. FUA. *Tracking multiple people under global appearance constraints*, in "IEEE International Conference on Computer Vision (ICCV)", 2011, pp. 137-144

[53] R. BENENSON, M. MATHIAS, R. TIMOFTE, L. V. GOOL. *Pedestrian detection at 100 frames per second*, in "CVPR", 2013

[54] B. BERKIN, B. K. HORN, I. MASAKI. *Fast Human Detection With Cascaded Ensembles On The GPU*, in "IEEE Intelligent Vehicles Symposium", 2010

[55] M. D. BREITENSTEIN, F. REICHLIN, B. LEIBE, E. KOLLER-MEIER, L. VAN GOOL. *Markovian tracking-by-detection from a single, uncalibrated camera*, in "PETS workshop", 2009, pp. 71-78

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

[57] 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

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

[59] D. CONTE, P. FOGGIA, G. PERCANNELLA, M. VENTO. *Performance evaluation of a people tracking system on the pets video database*, in "PETS workshop", 2009

[60] A. D. COSTEA, S. NEDEVSCHI. *Word Channel Based Multiscale Pedestrian Detection without Image Resizing and Using Only One Classifier*, in "CVPR", 2014

[61] 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

[62] N. DALAL, B. TRIGGS, C. SCHMID. *Human detection using oriented histograms of flow and appearance*, in "European conference on computer vision", Springer, 2006, pp. 428–441

[63] 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

[64] P. DOLLAR, S. BELONGIE, P. PERONA. *The Fastest Pedestrian Detector in the West*, in "BMVC", 2010

[65] P. DOLLAR, Z. TU, P. PERONA, S. BELONGIE. *Integral channel features*, in "BMVC", 2009

[66] 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

[67] P. FELZENSZWALB, R. GIRSHICK, D. MCALLESTER, D. RAMANAN. *Object Detection with Discriminatively Trained Part-Based Models*, in "PAMI", 2009, vol. 32, n^{o} 9, pp. 1627–1645

[68] P. FELZENSZWALB, D. MCALLESTER, D. RAMANAN. *A discriminatively trained, multiscale, deformable part model*, in "CVPR", 2008

[69] J. FERRYMAN, A. SHAHROKNI. *An overview of the pets2009 challenge*, in "PETS", 2009

[70] Q. GAO, S. SUN. *Trajectory-based human activity recognition with hierarchical Dirichlet process hidden Markov models*, in "Proceedings of the 1st IEEE China Summit and International Conference on Signal and Information Processing", 2013

[71] W. HU, X. XIAO, Z. FU, D. XIE, T. TAN, S. MAYBANK. *A system for learning statistical motion patterns*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2006, vol. 28, n° 9, pp. 1450–1464

[72] J.-F. HU, W.-S. ZHENG, J. LAI, J. ZHANG. *Jointly learning heterogeneous features for RGB-D activity recognition*, in "CVPR", 2015 [*DOI :* 10.1109/CVPR.2015.7299172]

[73] A. KARAKOSTAS, A. BRIASSOULI, K. AVGERINAKIS, I. KOMPATSIARIS, M. TSOLAKI. *The Dem@Care Experiments and Datasets: a Technical Report*, 2014

[74] Y. KONG, Y. FU. *Bilinear heterogeneous information machine for RGB-D action recognition*, in "CVPR", 2015

[75] H. S. KOPPULA, R. GUPTA, A. SAXENA. *Learning Human Activities and Object Affordances from RGB-D Videos*, in "Int. J. Rob. Res.", July 2013, vol. 32, n° 8, pp. 951–970, http://dx.doi.org/10.1177/0278364913478446

[76] H. KOPPULA, A. SAXENA. *Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation*, in "ICML", 2013

[77] 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

[78] W. LIU, D. ANGUELOV, D. ERHAN, C. SZEGEDY, S. REED. *SSD: Single Shot MultiBox Detector*, in "arXiv preprint", 2015, pp. 1–15 [*DOI :* 10.1016/J.NIMA.2015.05.028], http://arxiv.org/abs/1512.02325

[79] L. LIU, L. SHAO. *Learning Discriminative Representations from RGB-D Video Data*, in "IJCAI", 2013

[80] C. LU, J. JIA, C.-K. TANG. *Range-Sample Depth Feature for Action Recognition*, in "CVPR", 2014

[81] A. MILAN, K. SCHINDLER, S. ROTH. *Multi-Target Tracking by Discrete-Continuous Energy Minimization*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2015, vol. PP, n° 99, pp. 1-1 [*DOI :* 10.1109/TPAMI.2015.2505309]

[82] 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

[83] B. MORRIS, M. TRIVEDI. *Trajectory Learning for Activity Understanding: Unsupervised, Multilevel, and Long-Term Adaptive Approach*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", Nov 2011, vol. 33, n° 11, pp. 2287-2301 [*DOI :* 10.1109/TPAMI.2011.64]

[84] M. MOZAZ, M. GARAIGORDOBIL, L. J. G. ROTHI, J. ANDERSON, G. P. CRUCIAN, K. M. HEILMAN. *Posture recognition in Alzheimer's disease*, in "Brain and cognition", 2006, vol. 62, n° 3, pp. 241–245

[85] T. L. A. NGUYEN, D. P. CHAU, F. BREMOND. *Robust Global Tracker based on an Online Estimation of Tracklet Descriptor Reliability*, in "Advanded Video and Signal-based Surveillance", Karlsruhe, Germany, August 2015, https://hal.inria.fr/hal-01185874

[86] B. NI, P. MOULIN, S. YAN. *Order-Preserving Sparse Coding for Sequence Classification*, in "ECCV", 2012

[87] O. OREIFEJ, Z. LIU. *HON4D: Histogram of oriented 4D normals for activity recognition from depth sequences*, in "CVPR", 2013

[88] S. REN, K. HE, R. GIRSHICK, J. SUN. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*, 2015, https://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf

[89] 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

[90] 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, http://dx.doi.org/10.1007/978-3-642-33718-5_11

[91] 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

[92] L. RYBOK, B. SCHAUERTE, Z. AL-HALAH, R. STIEFELHAGEN. *Important stuff, everywhere! Activity recognition with salient proto-objects as context*, in "WACV", 2014

[93] 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

[94] L. SEIDENARI, V. VARANO, S. BERRETTI, A. DEL BIMBO, P. PALA. *Recognizing Actions from Depth Cameras as Weakly Aligned Multi-part Bag-of-Poses*, in "CVPRW", 2013

[95] A. SHAHROUDY, G. WANG, T.-T. NG. *Multi-modal feature fusion for action recognition in RGB-D sequences*, in "ISCCSP", 2014

[96] 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

[97] S. WALK, N. MAJER, K. SCHINDLER, B. SCHIELE. *New features and insights for pedestrian detection*, in "CVPR", 2010

[98] X. WANG, T. X. HAN, S. YAN. *An hog-lbp human detector with partial occlusion handling*, in "ICCV", 2009

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

[100] C. WOJEK, B. SCHIELE. *A performance evaluation of single and multi-feature people detection*, in "DAGM Symposium Pattern Recognition", 2008

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

[102] L. XIA, J. AGGARWAL. *Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera*, in "CVPR", 2013

[103] J. YANG, Z. SHI, P. VELA, J. TEIZER. *Probabilistic multiple people tracking through complex situations*, in "PETS workshop", 2009

[104] L. ZHANG, Y. LI, R. NEVATIA. *Global data association for multi-object tracking using network flows*, in "Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on", June 2008, pp. 1-8 [*DOI :* 10.1109/CVPR.2008.4587584]

[105] Q. ZHU, S. AVIDAN, M. YEH, K. CHENG. *Fast Human Detection using a Cascade of Histograms of Oriented Gradients*, in "CVPR", 2006

[106] Y. ZHU, W. CHEN, G. GUO. *Evaluating spatiotemporal interest point features for depth-based action recognition*, in "Image and Vision Computing", 2014, vol. 32, n^{o} 8, pp. 453 - 464 [*DOI :* 10.1016/J.IMAVIS.2014.04.005], http://www.sciencedirect.com/science/article/pii/S0262885614000651