Proposal for the INRIA Project-Team
STARS
Spatio-Temporal Activity Recognition Systems

May 2012

Theme: Vision, Perception and Multimedia Understanding

Domain: Perception, Cognition, Interaction

Abstract

The unprecedented capabilities of hardware devices as well as the progress made in computer science enable to address new challenging problems related to scene understanding which were seen as unrealistic one decade ago. Thus Stars aims at designing **Spatio-Temporal Activity Recognition Systems** for improving health and well-being as well as people safety, security and privacy. Our objective is to develop new methods for the real-time semantic interpretation of human activities performed in complex scenes from the real world. We will propose ambitious 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 dependant 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 persons and visual surveillance.

Keywords

Activity Recognition, Computer Vision, Perception, Semantics,
Machine Learning, Software Engineering, Health and Well Being.

Contents

1 Team Composition

The permanent scientific members are INRIA scientists with different background and skills[1]. François Brémond and Monique Thonnat are cognitive vision scientists with a large experience in artificial intelligence and computer vision. Guillaume Charpiat is specialized in computer vision and machine learning. Sabine Moisan and Annie Ressouche are computer scientists with respectively a large experience in artificial intelligence, software design and formal techniques for software verification.

Permanent members

- **Team Leader**: François Brémond DR2, HDR
- **Research scientists INRIA**:

  Guillaume Charpiat CR1

  Sabine Moisan CR1, HDR

  Annie Ressouche CR1

  Monique Thonnat DR1, HDR (40% of her time in Stars)
- **Team Assistant**: Christine Claux INRIA

Non permanent team members

- **External collaborators**:

  Prof Jean-Paul Rigault, UNSA, Nice Sophia Antipolis University

  Prof Philippe Robert, PUPH at CMRR CHU Nice Hospital

  Etienne Corvée, Research Engineer at Link Care Services

  Véronique Joumier, Research Engineer at CMRR CHU Nice Hospital

  Daniel Gaffé, associate professor, Nice Sophia-Antipolis University
- **Post-doctoral Researchers**:

  Carlos Crispim, Post-Doc from Federal University of Santa Catarina, Brazil
- **PhD Students**:

  Julien Badie, Nice Sophia-Antipolis University (1st year)

  Slawomir Bak, Nice Sophia-Antipolis University (3rd year)

  Piotr Bilinski, Nice Sophia-Antipolis University (2nd year)

  Carolina Garate, Nice Sophia-Antipolis University (1st year)

  Ratnesh Kumar, Nice Sophia-Antipolis University (1st year)

  Emmanuel Mulin, Nice Sophia-Antipolis University (2nd year)

  Guido-Tomas Pusiol, Nice Sophia-Antipolis University (3rd year)

  Rim Romdhane, Nice Sophia-Antipolis University (3rd year)

  Malik Souded, Nice Sophia-Antipolis University (2nd year)

[1]Their Web pages are accessible at http://www-sop.inria.fr/members/Firstname.Lastname/

- **INRIA Research Engineers**:

  Bernard Boulay, PhD (DGCIS National project QUASPER)
  Julien Gueytat, (ANR National project SWEETHOME)
  Duc Phu Chau, PhD (European project VANAHEIM)
  Leonardo Rocha, (DGCIS National project CIU-Santé)
  Sofia Zaidenberg, PhD (European project VANAHEIM)

HDR stands for "Habilité à Diriger des Recherches" which is the diploma enabling Ph.D. thesis supervision. CR1 stands for Chargé de Recherches first class and DR2 (respectively DR1) stands for "Directeur de Recherches" (senior scientist) second class (respectively first class).

2 Stars Overall Objective

The unprecedented capabilities of hardware devices as well as the progress made in computer science enable to address new challenging problems related to scene understanding which were seen as unrealistic one decade ago. Thus **STARS (Spatio-Temporal Activity Recognition Systems)** aims at designing Spatio-Temporal Activity Recognition Systems for improving health and well-being as well as people safety, security and privacy. In today's world, a huge amount of new sensors are currently available, adressing potentially new needs of the modern society. Smart homes have been well advertised but remain a dream due to the poor capability of automated systems to perceive and reason about their environment. A hard problem is also for an automated system to cope 24/7 with the variety and complexity of the real world.

Therefore, the objective of Stars is to propose novel autonomous systems for the **real-time semantic interpretation of dynamic scenes** observed by sensors. We will study long-term spatio-temporal activities performed by several interacting agents such as human beings, animals and vehicles in the physical world. A typical example of a complex activity is shown in Figure 1 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 person and his/her apartment. The goal is to recognize everyday activities at home through formal activity models (as shown in Figure 2) 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 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 could be performed by an Activity Recognition System which transmits a textual report (containing only the necessary meta-data) to the General Practitioner or the caretaker. For instance, a challenge is for the system to understand the reasons of the person anxiety and to suggest palliative actions. Thanks to the recognized activities, the quality of life of the older people could thus be improved and their personal information can be preserved. More information is available on the Gerhome project website[2].

The ultimate goal for activity recognition systems is to infer **high level semantics**. These semantics are the essential elements to provide the appropriate services to a potential user. An important step is then to propose a computational representation of activities to adapt these services to the people. 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 only the pertinent raw data for adapting the services to the people while preserving their privacy. We plan to study ambitious solutions including of course the local processing of the data without transmission of images and the utilisation of new compact sensors developed for interaction (also called RGB-Depth sensors, an example being the Kinect in Figure 3) or networks of small non visual sensors.

2.1 Main Challenge

The major issue in semantic interpretation of dynamic scenes is the gap between the subjective interpretation of data and the objective measures provided by the sensors. The interpretation of observation data is not unique; it depends on the a priori knowledge of the observer and on his/her goal. For instance, a video showing airport apron scenes

[2] http://www-sop.inria.fr/members/Francois.Bremond/topicsText/gerhomeProject.html

Figure 1: Homecare monitoring: the set of sensors embedded in an apartment and 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))
Alarm (AText("prepare meal")
AType("not urgent")))

Figure 2: Homecare monitoring: example of an activity model describing a scenario related to the preparation of a meal with a high-level language

Figure 3: Privacy protection: activities (e.g. Person is standing, then he/she is walking) are correctly recognized even if the appearance of the older person is hidden by (a) a mask computed thanks to the depth map of the Kinect sensor or by (b) a 3D avatar.

will be interpreted differently by a security officer (looking for dangerous situations) or by an airport manager (looking at servicing operation scheduling).

2.2 Scientific Strategy

Our approach to address this problem is to establish a bridge between the application dependent subjective interpretation and the objective analysis of observation data (e.g. videos). First we define and use a set of objective measures which can be extracted in real-time from the data. Second, we propose formal models for users to express their activities of interest. Finally, we build matching techniques to bridge the gap between the objective measures and the subjective activity models. More precisely, Stars scientific strategy consists in designing an intermediate level (called *perceptual concepts* or *primitive events*) to bridge this semantic gap. The challenge is to propose new and efficient methods to infer perceptual concepts, which are generic enough to describe semantic activities and specific enough to be grounded into low level data.

Stars strategy is also to favor the easy generation of **activity recognition systems**. These systems correspond to a complex succession of pattern matching and clustering algorithms, combined with adequate knowledge representation (e.g. scene topology, temporal constraints) at different abstraction levels (from raw signal to semantics). As these algorithm chains are highly context dependent (e.g. parametrizable), Stars will propose optimization and supervision methods to obtain robust activity recognition systems given prior knowledge (such as an intuitive scene description). Therefore, Stars will propose new techniques in the field of **computer vision** and **machine learning** for *physical object detection*, *activity understanding and learning*, and *dependable activity recognition system design*. Thus our approach will be multidisciplinary and pragmatic.

However, to build concrete activity recognition systems, we will need results from several research fields where Stars plans only to use the state of the art techniques (e.g. Human Machine Interface, ontology technologies, audio processing). Whenever the performance of these results is not sufficient, we may have to extend or develop new techniques as it could be the case for uncertainty management and model-checking.

Within Stars we will have two major application domains : healthcare monitoring and visualsurveillance. We will strongly reinforce our implication in the **healthcare monitoring** domain to address new research challenges which are further elaborated in sections 2.3 and 4.1. We will continue studying activity recognition systems for the **visualsurveillance** application domain to deepen the research performed in Pulsar (e.g. detection of unexpected activities).

2.3 Strategic Partnership

Stars will develop ambitious research activities for healthcare monitoring; particularly the assistance of older people, the early detection of deteriorated health status and the early diagnosis of illness. The goal is to address obvious societal challenges (i.e. a growing older society) and to address new research challenges such as the analysis of complex spatio-temporal activities (e.g. fine gestures, complex interactions, long term activities), which are more detailed in section 4.

Therefore, we have initiated a strategic partnership (called CobTek[3]) with Nice hospital [34] (CHU Nice, Prof P. Robert) to study the impact of video understanding approaches on older people disorders. This partnership has been proposed for a duration of 4 years (starting from the first of January 2012) in the framework of the convention between INRIA and the Nice Sophia-Antipolis University. Stars is currently hosting one employee (Veronique Joumier) from Nice hospital. This Research Unit CobTek was positively evaluated in May 2011 by the national AERES evaluation committee of the Nice Sophia-Antipolis University.

To strengthen this research axis, we have also set-up an experimental laboratory (Ger'home) at Sophia Antipolis together with CSTB (the French scientific and technical center for building[4]) to test new sensors and new activity recognition techniques. In the same direction, we are participating to the large scale INRIA initiative (called Personally Assisted Living[5]) gathering 9 INRIA project-teams on providing services for older people. For instance, these services should allow older people to stay safely at home, to benefit from an automated medical supervision and should delay their entrance in nursing homes. In the healthcare domain, we are also involved in two European initiatives : EIT ICT Labs and FP7 ICT IP Dem@Care. The objective of all these studies is to improve the health of older people and to anticipate the deterioration of health status, such as dementia.

2.4 From Pulsar to Stars

The Stars team is in the continuation of the Pulsar INRIA project-team in the theme "Vision, Perception and Multimedia Understanding". Pulsar was created in January 2008 and stopped in 2011 because Monique Thonnat, the Pulsar leader, has taken responsibilities at INRIA national level as deputy scientific director. She will stay at 40% of her time in the new Stars team. All the Stars permanent research scientists were in the Pulsar INRIA project team. In the same way as Pulsar, Stars will be a multi-disciplinary team at the frontier of computer vision, machine learning, knowledge-based systems, and software engineering. Stars will focus on activity recognition.

In 2008, the two main Pulsar research axes were *scene understanding for activity recognition* and *software architecture for activity recognition*. Thus Stars can rely on several results obtained by Pulsar from 2008 to 2011. Among them we can cite the following ones:

[3]http://cmrr-nice.fr/?p=cobtek-presentation
[4]http://international.cstb.fr/
[5]http://pal.inria.fr/

- an original 4D semantic approach for video understanding including mobile object detection, classification, tracking and complex scenario recognition [239, 176, 9, 10, 278, 137, 138].
- a visual concept ontology [236] and a video event ontology [278].
- the SUP[6] Platform (Scene Understanding Platform) [137, 110], which was the Pulsar platform for real-time video understanding (see Annex A for more details).
- a model-driven approach to configure video surveillance systems, which has been validated on realistic applications [69, 119, 118].

The **major differences between Pulsar and Stars** are:

- A stronger emphasis on **healthcare**. The assistance to the older persons with our collaboration with Nice Hospital, will be deepened and will bring challenging new problems [181, 218]. To achieve our objectives on the early detection of deteriorated health status while preserving people privacy, four issues in particular are needed to be addressed. First, alarming events or critical situations (e.g. falls, depressions) need to be detected. Second, the degree of frailty of older people should be objectively assessed based on multi-sensor analysis and human activity recognition. This automated assessment should follow a well-defined geriatric protocol. Third, the behavior profile of a person - its usual and average activities - should be identified and then any deviation (missing activities, disorder, interruptions, repetitions, inactivity) from this profile should be detected. Moreover, to help to better understand the behavior disorders of older people, we propose to build a library of reference behaviors (containing raw signal and various meta-data) characterizing the different degrees of people frailty. Fourth, ambitious communication paradigms need to be conceived in order to preserve people privacy and in the same time to provide the appropriate information to the medical carers. Thanks to the proposed approach, new services will be provided to improve the engagement, safety, autonomy or empowerment of older persons as well as to deliver a more specialized assessment of their health status. This application domain will be a stimulating way to validate our work on activity recognition systems.
- A stronger emphasis on **machine learning**. Our objective is to propose complete cognitive systems mixing perception, understanding and learning techniques in order to obtain systems which can autonomously adapt themself to dynamic environment. This objective includes to learn in a more or less autonomous way the needed activity models and to automatically bridge the gap between these models and the perceptual level. A still open issue consists in determining whether these models should be given a priori or learned and how these models should be learned : in a supervised or unsupervised way or incrementally etc. A related challenge consists in understanding whether these learned models should be defined statistically or could have a semantic description.

Therefore, we propose for Stars three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition.

[6] SUP is a generic software platform for building a large variety of real-time activity recognition systems dedicated to a particular application. This platform will enable Stars members to design, experiment and validate their new algorithms for activity recognition.

- The **perception** direction consists in extracting perceptual features characterizing the activities of objects moving in the observed scene. The hard problem is to extract meaningful features while coping 24/7 with the variety of a dynamic world. To face this challenge, new machine learning techniques are needed in particular for modeling the distribution of features and the associated noise.

- The **semantic activity recognition** direction consists in inferring high level semantics from the perceptual features. An ambitious objective is to model complex activities involving multiple interacting objects. Thus, there is an increasing importance of machine learning techniques to acquire a priori knowledge (i.e. the activity models), which is necessary to abstract the signal information (e.g. video, pixels) into semantics such as events or meta-data. There is an open issue which consists in determining how much knowledge should be given and how much should be learned. The whole challenge consists in managing a huge amount of information and in structuring all this knowledge in order to capitalize experiences, to share them with others and to update them along experimentation.

- The **software architecture** direction is an important research topic, as generating operational vision systems is a critical point for activity recognition. The challenge is to convert (semi-)formal requirements into an assembly of pre-existing software components. This axis has evolved from a knowledge-based approach to object-oriented frameworks and model-driven methods to support video system configuration and runtime adaptation.

3 Research Directions

As said previously, Stars will follow three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved**: the perception and the semantic activity recognition directions will provide original activity recognition algorithms, whereas the software architecture direction will provide new methodologies for building effective systems for visualsurveillance and healthcare applications.

Conversely these concrete systems will raise new software issues that will enrich the software engineering research direction.

3.1 Perception for Activity Recognition

Participants[7] : **François Brémond, Guillaume Charpiat, Monique Thonnat** and Sabine Moisan.

Despite few success stories, such as traffic monitoring (e.g. http://www.citilog.fr), swimming pool monitoring (e.g. http://www.poseidon-tech.com) and intrusion detection (e.g. http://www.objectvideo.com), video analytics systems remain brittle and can function only under restrictive conditions (e.g. daylight, diffuse lighting conditions, no shadows), with poor performance over time [148]. They are hardly modifiable, containing little a priori knowledge on their environment. Moreover, these systems are very specific and need to be developed from scratch for a new application. To address these issues, most researchers have tried to develop specific vision algorithms focusing only on particular situations. Up to now no vision algorithm has been able to address the

[7]The main permanent research scientists working in this research direction are in bold font, permanent research scientist working only partly in this research direction are in regular font

large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives.

3.1.1 Related Work

Many teams around the world (e.g. University of Leeds, University of Hamburg, University of Southern California, Georgia Tech, University of Central Florida, University of Kingston upon Thames, National University of Cheng Kung and Prima Team at Inria Grenoble) work on video processing for activity recognition, and a huge literature describes the large variety of video processing techniques. Among the classical approaches, we can cite motion estimation [287, 273, 183], object detection from a background image [232, 259], feature-based object detection [211, 237], template-based object detection [194], crowd detection [291, 146], human-body detection [274, 205, 142, 192], gait analysis [249, 221], vision systems [206, 182, 191] and video processing performance evaluation [156, 210, 293]. Several surveys [241, 143, 149] categorize these techniques in a more or less exhaustive way.

Object tracking and multi-sensor analysis We first highlight two classical issues related to perception for activity recognition: object tracking and multi-sensor analysis. Much work studied tracking algorithms (mostly based on feature-tracking [203, 238] or template-tracking [161, 172, 193] using for instance probabilistic techniques [189, 167, 285], 3D techniques [184, 271], active-vision techniques [198], long-term techniques [290, 153], graph-based techniques [267, 204]) . Concerning multi-sensor analysis, much work studied information fusion techniques (mostly with multiple-cameras [257, 188, 160] and, less frequently by combining video cameras with other sensors [186, 174, 200]) to improve results and robustness. In particular, several studies have been done to combine different types of media (video, audio and others) [42] to move from video understanding to scene understanding. However, a general framework is needed to perform seamlessly this information fusion. In comparison, the Stars team will design a global real-time framework for 4D spatio-temporal reasoning.

Learning shape statistics and system control For Stars, there are also two new issues related to perception for activity recognition which combine machine learning and perception techniques: learning shape statistics and system control.

An interesting problem is not only to be able to recognize objects based on their coarse shape and appearance such as in [192], but also to classify finely their postures or gestures. The discriminative power of such classifications is crucial for further description and interpretation of events. However, the computation of shape statistics is still a matter of research, in particular when the sets of shapes to be studied are very heterogeneous, like human silhouettes. There exist currently mainly two approaches to shape statistics : the diffeomorphic approach, as in the work of Trouvé [272], and the kernel approach, as in the work of Cremers [169] or in spectral techniques to low-dimensional embedding [151]. In the first case, a mean shape is computed, as well as diffeomorphisms from the mean shape to all samples. Then PCA (principal component analysis) is performed in the tangent space of the mean shape over the diffeomorphisms. The problem of this approach is the existence and the meaning of the mean shape, the difficulty to compute shape deformations between samples, and the supposition that the cloud of shapes is somehow Gaussian. On the other hand, kernel methods applied to sets of shapes are able to handle any distribution of shapes, but these methods require that the density of these shape distributions is high enough, which in practice is never reached for datasets of human silhouettes (few hundreds of samples while the intrinsic dimension is more than 30). We proposed in [73] a manifold learning approach that can face this curse of dimensionality under a natural assumption.

Concerning statistical learning, the current main approaches rely on kernel methods or on clustering which have proved to be very efficient in many practical cases, but they do not really explain the data (in that they do not build models) and are thus far from bringing semantics. On the other side, work on automatic model construction, as the one by Zhu & Mumford [292] based on image grammars, seem promising, even if the statistical learning part remains to be improved. We plan to cope with the problem of automatic modeling by searching for suitable representations of data, based on compactness and time efficiency criteria.

Concerning system control, little work has been done to design learning techniques for task supervision (e.g. task planning) and for setting program parameters [209, 258, 240]. For instance, in the Prima Team at INRIA Grenoble, Hall [208] proposes an automatic parameter regulation scheme for a tracking system. An auto-critical function is able to detect a drop in system performances with respect to a scene reference model. In such a situation, an automatic regulation module is triggered to provide a new parameter setting with better performance. However the comparison with a manually tuned system shows that the controlled system is not able to reach the same performance. This is due to the fact that the controlled system does not use a priori knowledge of programs whereas the expert takes advantage of this kind of knowledge and is rapidly able to understand which parameters are relevant to which tracking problems. This work shows that more work is still needed. To tackle this problem, the Stars team proposes to use program supervision techniques to configure and control system execution, following the promising results we have obtained in [14, 134]. This implies adding a reasoning level which combines a priori knowledge of the program to tune and learned parameter values in function of the contextual class.

3.1.2 Proposed Approach

Perception for Activity Recognition Our contribution in perception will be twofold: first to propose robust techniques for physical object detection algorithms to adapt themself to dynamic environment, and second to deepen our work on the management of a library of video processing programs.

A first 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. For instance, some accurate techniques have been designed for people detection, but none of them are real-time and months of processing will be needed to analysis few weeks of videos even using dedicated hardware. We will not only rely on hardware solutions (as GPU Graphics Processing Unit or FPGA Field-Programmable Gate Array) for getting video frame rate performance but we will also design efficient software solutions either by adapting existing algorithms or by proposing new ones, such as [93, 105]. For instance, for people detection the dedicated state of the art algorithms struggle to provide reliable and useful results to the entire video understanding community. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios. In particular, reliability measurements could be provided to better analyse complex human appearances against all kinds of background scenes. There exists many features (e.g. HOG for high resolution images, extended HAAR, LPB for low resolution images) in the literature that should be better combined to increase feature robustness. The performance of training algorithms depends strongly on the content and quality of large non-sorted training databases (e.g. with or without occlusions, colors, variety of view-points, variety of human poses). The people classification scheme should be trained with various efficient training algorithms (e.g. Adaboost, Hierarchical Trees). In addition, too few studies have been achieved in studying the temporal behaviour of the features of the overlapping human candidates. This task could be achieved by clustering potiental candidates depending on their position and their reliability. This task would also 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 (e.g. covariance matrice descriptor) [105], the future approaches should focus more on aligning two people detections and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. This problem can be overcame by learning a distance metric to maximize matching accuracy. As the results of the re-identification are always given as the ranking list, re-identificaiton should focus 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.

We plan also to 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). For instance, meta-data provided by different cameras can be combined based on the meta-data reliability as proposed in [175]. Currently we are working with CSTB (Centre Scientifique et Technique du Batiment) to improve Ger'home (http://gerhome.cstb.fr/), an experimental laboratory to study different environmental sensor configurations. We are also involved in a strong partnership with Nice hospital (CHU Nice, Prof P. Robert) to combine accelerometer with video analysis.

For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in visualsurveillance and several days in healthcare). Despite all the work done in object detection and tracking, state of the art algorithms remain brittle. 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 probabilist (e.g. Bayesian) framework modeling the uncertainty of the tracked objects [110]. This reasoning layer will take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

Another possible approach to improve jointly segmentation and tracking is to consider videos as 3D volumic data and to search for trajectories of points that are statistically coherent both spatially and temporally [260, 231, 166]. This point of view enables new kinds of statistical segmentation criteria and of ways to learn them. More generally, to improve segmentation quality and speed, different optimization tools such as graph-cuts can be used, extended or improved.

Moreover a lot of research work is needed to provide fine and accurate models of human shape and gesture. Currently the SUP library is limited to coarse motion detection and global 3D features. We plan to deepen the work we have done on human posture recognition with B. Boulay [9, 157, 158] to match 3D models and 2D silhouettes. For instance, we plan to work, with P. Bilinski, one of our PhD students, on gesture recognition based on feature tracking [71]. These kinds of detailed spatio-temporal features are required for patient monitoring in healthcare.

We also plan to use the shape statistics developed in [73] for the segmentation of images or videos with shape prior, by learning local segmentation criteria that are suitable for parts of shapes. This would unify patch-based detection methods and active-contour-based segmentation methods in a single framework. These shape statistics will be used also for a fine classification of postures and gestures, in order to extract more precise information from videos for further activity recognition. In particular, the notion of shape dynamics has to be studied.

Another way to extract meaningful information from videos with little supervision is to let a statistical learning algorithm find redundancies or correlations in the video flow and model automatically these data accordingly. Fully automatic modeling is a very hard subject in machine learning but is a good track to fill the gap between perception and understanding.

A second type of issues is to control automatically a library of video processing programs. One objective in this domain will be to build an effective perception library to be integrated in SUP that can be dynamically controlled by an activity recognition system. More precisely the activity recognition system will not only select robust algorithms for feature extraction but also ensure they work efficiently with real-time constraints. To reach this objective, we will formalize the perception algorithm conditions of use during run-time [137]. In the case of video processing, at least two problems are still open: robust image segmentation and meaningful feature extraction. Since it is still very difficult to build a video processing system for a new scene, we propose to study how machine learning techniques can solve these two issues. For example, the work we have done with B. Georis [14, 134] has shown in a specific case of illumination change how clustering techniques applied to intensity histograms could help in learning the best parameters for object segmentation. The core idea consists for a given class in exploring the parameter space using the simplex algorithm to optimize an evaluation function for finding the segmentation parameters. This research work can be generalized to many video analysis programs, for example, to learn parameter initialization criteria in the program supervision formalism. More generally there is a need for learning techniques for program supervision knowledge base refinement (e.g. to select automatically the best program for a given task). Also program supervision formalism enables a top-down approach to process more efficiently video data. For instance, as described in A. Ghiem PhD [137], object detection algorithms can take advantage of tracking results - using output data evaluation criteria - to better classify moving regions into noise whenever these regions are not correctly tracked throughout the video.

3.2 Semantic Activity Recognition

Participants[8] : **François Brémond, Guillaume Charpiat, Monique Thonnat** and Sabine Moisan

[8]The main permanent research scientists working in this research direction are in bold font, permanent research scientist working only partly in this research direction are in regular font

3.2.1 Related Work

The state-of-the-art in semantic activity recognition embraces several decades of productive work. It recently led to a new research theme: *Cognitive Vision*.

Semantic activity recognition is an ambitious research topic which aims at solving the complete interpretation problem ranging from low level signal analysis to semantic description of what is happening in the scene viewed by video cameras and possibly other sensors. This issue implies solving several problems grouped in two major categories: high level understanding and learning event patterns.

Activity recognition should not be confused with *action recognition*. *Action recognition* seeks to detect short-term sequences of configurations of an articulated form (typically human) that are characteristic to specific *actions* such as grasping or touching an object. By *activity* we mean a long-term symbolic spatio-temporal scenario (e.g. having lunch from beginning to end) performed by one or several interacting individuals in the physical world. In the computer vision community the term *video event* is often used as a general term for both a *short-term action* and a *long-term activity*.

High level understanding Concerning related work in high level understanding for activity recognition we focus here on two aspects which correspond to two important open issues: event recognition and ontology-based knowledge acquisition.

An increasing effort is done to recognize events from videos [162, 282, 144, 207], in particular using motion-based techniques [171, 251, 228, 250]. For instance several studies have been done on home-care applications [215, 173]. Along the work achieved in event recognition, two main categories of approaches are used to recognize temporal scenarios based on (1) a probabilistic network [264, 217, 284] (especially using Bayesian techniques [263, 234] or Hidden Markov models HMMs [201, 147]) or a neural network [159, 216] combining potentially recognized scenarios, (2) a symbolic network that stores all previously recognized scenarios [247, 254, 187, 230]. For instance, for the probabilistic techniques, Hongeng *et al.* [214] propose a scenario recognition method that uses concurrent Bayesian threads to estimate the likelihood of potential scenarios. Most of these techniques allow an effective recognition of scenarios, but there are still some temporal constraints which cannot be processed. Moreover, most of these approaches require that the scenarios are bounded in time. To address this issue and to handle uncertainty, the Stars team will extend the Orion's original approach on scenario recognition based on temporal constraint checking [277, 278].

Knowledge acquisition is still a complex and time-consuming task but it is required to obtain an effective system to recognize complex objects. Most approaches use dedicated knowledge dependent on the application domain. For instance, in [275] object description is based on domain related annotation. We propose to use a perceptual concept ontology to make easier the knowledge acquisition task in the same way as we have done in [235, 12, 236] for static 2D visual concepts.

Learning event patterns Regarding machine learning and event recognition, most approaches learn event patterns using statistical techniques [286, 227, 213, 199] and several approaches can deduce abnormalities from the learned events [281, 288], for instance using probabilistic latent semantic analysis (PLSA) [190]. Regarding action recognition, a great effort has also been devoted to representing actions via their 2D characteristics computed with statistical techniques [246, 179, 229, 279]. For instance, the Lear and Willow projects of INRIA is working on object and action recognition and scene interpretation from static images and video sequences. They use shape and interest point information (i.e. local descriptors) to describe actions in images. This numerical content is learned to build visual models of action classes. Therefore, visual description and action models are mainly based on low-level features. They do not use symbolic information to describe the objects.

3.2.2 Proposed Approach

Our goal is to propose innovative techniques for autonomous and effective scene understanding systems for long-term activity recognition [14, 134]. This objective is ambitious; however the current state-of-the-art techniques in cognitive vision already led to partial solutions [165, 248, 170, 283].

The major issue in semantic interpretation of dynamic scenes is the gap between the subjective interpretation of data and the objective measures provided by the sensors. Our approach to solve this problem is to keep a clear boundary between the application dependent subjective interpretations and the objective analysis of the videos. We thus define a set of objective measures which can be extracted in real-time from the videos, we also propose formal models to enable users to define their activities of interest and we build matching techniques to bridge the gap between the objective measures and the activity models.

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 insights of the scene and its dynamics while handling the low level noise. Moreover, to reach a semantic abstraction level, 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 ontologies are needed.

More precisely we plan to work along two main research axes: high level understanding (recognize the activities of physical objects based on high level activity models) and learning (how to learn the models needed for activity recognition). In addition we are exploring a third approach (related to discrete event systems). The results of this work will be integrated as understanding and learning components in the SUP platform for activity recognition which is described in section 3.3.

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

We propose to define new activity recognition algorithms and activity models. Activity recognition includes the computation of relationships between physical objects. The real challenge is to explore efficiently all the possible spatio-temporal relationships of the objects that may correspond to activities of interest. 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). One challenge is to explore and analyze this large event space without getting lost in combinatorial searches during the on-line processing of the video streams.

Concerning the modeling of activities, there are two kinds of open problems: the introduction of uncertainty in the formalism for expressing a priori knowledge and the development of knowledge acquisition facilities based on ontological engineering techniques. For the first problem we will investigate either classical statistical techniques and fuzzy logics or new approaches mixing the two kinds of techniques such as Markov Logic Networks [185].

We need also to extend our current visual concept ontology (currently limited to color, texture and spatial concepts) with temporal concepts (motion, trajectories, events, ...) and other perceptual concepts (such as audio concepts or physiological sensor concepts). Here, ontology studies are limited to the issues of the spatio-temporal activity recognition domain and do not include a generic ontology.

A short term goal will be to add new activity concepts related to other sensors, both in the framework of the European project VANAHEIM for activity monitoring in subways and in the framework of the French SWEET-HOME [100, 128] project for elderly monitoring. For instance environmental sensors (pressure, contact sensors,...) are necessary for obtaining robust estimation on the frailty of the elderly.

Learning for Activity Recognition The second research axis is to study which machine learning techniques could help to learn the models needed for activity recognition. Since it is still very difficult to build an activity recognition system for a new application, we propose to study how machine learning techniques can automate model building or model enrichment at the understanding level.

At this level two main research topics are planned: a first topic is the learning of primitive event detectors. This can be done in the same way as in Nicolas Maillot's PhD [236, 12], by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual features samples. This work was restricted to color, texture and simple shapes and needs to be extended towards spatio-temporal primitive events, as it has been proposed in [8].

A second topic is the learning of typical composite event models for frequent activities [97, 6] using data mining techniques. We name *composite event* a particular combination of several primitive events. For instance, PhD G. Pusiol's algorithm is able to learn typical composite event models for frequent activities [62, 82, 123] using data mining techniques on short trajectories. In this work we have proposed a framework to recognize and classify loosely constrained activities with minimal supervision. The framework uses basic trajectory information as input and goes up to video interpretation. The work reduces the gap between low-level information and semantic interpretation, building an intermediate layer of Primitive Events. The proposed representation for primitive events aims at capturing small coherent units of motion over the scene with the advantage of being learned in an unsupervised manner. An activity is then defined as a particular combination of several spatio-temporal primitive events. The activity model is built in a semi-supervised way using only real tracking data [99]. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. lowering the human annotation task) and whether we can do it on-line? This topic is very important for healthcare activity monitoring where large amounts of data are available but little a priori knowledge is formalized. Coupling learning techniques with a priori knowledge techniques is very promising but still at a very preliminary stage for activity recognition and, in particular, recognition of on-line meaningful semantic activities.

Activity Recognition and Discrete Event Systems The two 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 3.3.3.

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

3.3 Software Engineering for Activity Recognition

Participants[9]: **Sabine Moisan, Annie Ressouche**, **Jean-Paul Rigault** and François Brémond.

Concerning Software Engineering, Stars intends to study generic systems for activity recognition, to elaborate a methodology for their design, and to assist designers in configuring complex video systems. In the long term, we propose to contribute in the setting up of a complete software engineering methodology for the development of activity recognition systems. This methodology will be based on Model Engineering and formal techniques. The approach will be validated by applying it to the future Stars platform (a successor of SUP).

Applying the platform to new problems, the new experience will in turn be abstracted, enriching generic models and platform architecture. Therefore, the three Stars research directions (perception, semantic activity recognition and software architecture) consolidate each other.

3.3.1 Software Engineering Aspects

Activity recognition systems are complex to design due to a number of characteristic properties. First, their *algorithmic structure* is complex: these systems encompass *many abstraction levels*, from pixel-based processing to high level knowledge and data handling; all levels and their interactions impact the quality of the overall system. Secondly, their *architecture* is inherently *distributed* and often *heterogeneous* at almost all levels: many sensors of various kinds, local processing, distributed resources, distributed knowledge and data, etc. Thirdly, these systems have to be *efficient* since they have to work more and more in *real time*. Fourthly, the *target applications* are demanding in terms of *dependability*, *safety* and *privacy* (e.g., transport, medical domains). Finally, and

[9]The main permanent research scientists working in this research direction are in bold font, other permanent research scientist working only partly in this research direction are in regular font

for all the previous reasons, these systems have *requirements* that are *difficult to express and validate*.

A single characteristic property in the previous list makes software development a toil. Combining them constitutes a real challenge which calls for sophisticated Software Engineering (SE) techniques to set up a clear methodology and develop adequate supporting tools. It is not sufficient, however, to adopt well proven traditional SE methods; one needs to invent new ones, possibly picking, adapting, and developing state of the art research in this domain. Hence, our goal is not only to provide operational software tools, but also to draw new methodologies to build these tools.

In particular two bodies of SE techniques, both of which have drawn considerable interest in the recent years, may be fruitfully applied to the design of activity recognition systems: *genericity* for organizing design and code, and *model engineering* for methodological guidance.

Developing generic entities has been a major task of software engineering since its creation in 1969. Indeed such entities can be highly reusable, saving both development and testing/validation time. Consequently, they also promote extensibility, dependability, and maintainability. Subprograms, modules, components, objects, frameworks, all of these contribute to the domain. There is a danger though: genericity implies raising the abstraction level, possibly at the expense of specific application requirements or demanded performance. But software engineering has always striven to produce techniques and tools that try to reconcile reusability with adaptability to a specific task. Nowadays, the key studies in genericity involve the use of patterns [202] and object-oriented or component-based frameworks as well as the numerous techniques hidden behind the general term of *generative programming* [177] (metaprogramming, aspect-oriented programming, etc.).

A few years ago, the Object Management Group (OMG) launched an initiative to develop a so-called *Model-driven Architecture* (MDA) [155]. In the strictest sense, MDA is an attempt to favor a seamless development of "business" software using models (based on the UML, another production of OMG) and automatic transformations of these models, but establishing a clear separation between the high level application requirements and the middle and low level implementations, which are platform dependent. The current research broadens the scope of MDA, envisioning to design systems by an automatic (or semi-automatic, but controlled and validated) sequence of *model transformations*. With this meaning the approach is best designated as *Model Engineering* or *Model Driven Development* (MDD). Also, we do not underestimate the important work of the *Software Architecture* community [145] and think that both approaches could benefit from being merged.

The Software Architecture research direction proposed by Stars—as a continuation of previous Pulsar work—concerns the study of generic systems for activity recognition as well as the definition of a methodology for their design. We target genericity, modularity, reusability, extensibility, dependability, and maintainability, but still ensuring adequate (real time) performance. To tackle this challenge, we rely on state of the art Software Engineering practice. In particular we shall explore the new techniques mentioned above, those related to genericity on the architectural side as well as model engineering approaches on the methodological side.

It is almost impossible to achieve genericity from scratch, thus our approach is pragmatic: it is indeed fuelled by the problems that Stars has to face in video understanding and more generally in activity recognition. Now that we have reached some maturity in scene understanding, we can consider building more general solutions and tools. Therefore, we need to abstract from our experience with operational systems to provide more generic solutions and user-friendly tools. Once the corresponding generic entities are available, they can be applied to new problems. This new experience will in turn be abstracted, enriching generic models and platform architecture. Therefore, the three Stars research directions (perception, semantic activity recognition and software architecture) consolidate each other in a "virtuous cycle": an operationalization flow produces effective systems from the generic platform whereas an abstraction flow enriches the platform from the lessons learned with concrete applications.

Another important characteristic of this research direction is that we strive to develop models with formally correct foundations. This makes it possible to build dependable systems relying on formal verification techniques, since traditional testing is not sufficient in these domains.

This approach proved feasible with the development of the LAMA platform in the Orion project-team ([245]). Nowadays Model Driven Engineering appears as a promising way to master the description, the deployment, and the run time control of a complex component platform such as SUP. It appears as complementary to the knowledge-based approach of LAMA. We have already started to explore this direction by modeling video-surveillance activities and SUP components using Feature Models [69, 103]. This methodological work should still go with formal correction and thus the possibility of (semi-)automatic verifications.

3.3.2 Related Work

Activity recognition methods have been extensively studied during these last decades mainly in the cognitive computer science domain. Scenarios are an interesting way to express required behaviors of a system. A popular way to define scenarios comes from the UML community with the use of sequence diagrams. Message Sequence charts or Live Sequence Charts [180] are specification languages for scenarios with a clear graphical layout that immediately gives an intuitive understanding of the intended system behavior. Artificial intelligence proposes a declarative representation of scenarios defined as a set of spatio-temporal and logic constraints [276, 277] and relies on constraint resolution techniques to recognize scenarios.

Concerning safeness, our goal is not to develop original research, we rather intend to rely on well known formal techniques (synchronous models [154, 256], model-checking [220]) in the line of our previous work with the Blocks framework [243, 244]. The domain of synchronous model of time has been explored since two decades. However, only few research works have studied the efficient and modular compilation of synchronous programs relying on a model-driven approach. In our knowledge only some work has been dedicated to distributed code generation [233, 261].

On the methodological side, Model-driven engineering (MDE) seems mature enough to be confronted with real applications. A crucial question is to determine its current limits, especially regarding variability issues. This issue is difficult to master, even by experienced users [225, 252]. Feature models are one of the most common formalisms used to model commonalities and variabilities. They are not only relevant to requirement engineering [270] but also to design or implementation. Since the original definition [222], a plethora of notations and extensions have been proposed [178, 266, 150]. A generalization of all the variants of feature diagrams has been proposed in [262].

To transform a source model into a target one, MDE favors transformation rules that describes how constructs in the source model can be transformed into constructs in the target one [224]. Transformation rules must be expressive enough and preferably verifiable. OCL or systems like ALLOY are candidates for expressing these rules. Moreover, transformation engines (imperative ones such as KERMETA [141], more declarative like ATL[219]) or even inference rule engines as in artificial intelligence, are necessary to make them executable. An other alternative is to use a constraint solver to derive a correct and optimal configuration. However, means to efficiently translate specification rules into concrete features are still an open issue [280].

Another challenge concerns the exploitation of models at runtime to cope with the dynamic aspects of video surveillance systems. An emerging approach for dynamic adaptation in Software Product Lines [255] is to use Dynamic Software Product Lines [212] which try to achieve the self-modification of a system. In [195], the authors propose to combine the use of models at run time and aspect-oriented modeling techniques. A number of techniques have been proposed in the literature to capture and express adaptations: [196, 289, 197]. Finally, our models at runtime may deal with values, provided for example by sensors. To handle them we need an extended formalism of feature models, such as proposed in [152].

3.3.3 Proposed Approach

As indicated before we intend to study generic systems for activity recognition and to elaborate a methodology for their design. We wish to ensure genericity, modularity, reusability, extensibility, dependability, and maintainability.

This research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. Altogether it is a long term goal although some of its aspects can be addressed in the near future.

We plan to 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 shall follow state of the art Software Engineering practices and, if needed, we shall attempt to set up new ones.

Discrete Event Models of Activities As mentioned in the previous section (3.2.2) 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. We shall continue with the synchronous approach since it facilitates scenario validation.

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 [253, 164] and helps to produce a code that has been formally proved. Our verification method follows a compositional approach [244], a well-known way to cope with scalability problems in model-checking. Particularly, we intend to rely on a previous work we have led concerning the CLEM framework [140, 3, 33] that supplies a set of tools to design dependable and reusable applications relying on a synchronous modeling. Indeed, CLEM supports a modular compilation to deal with large systems and model-based approach to perform formal validation.

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 [226]. New model checking techniques can be used as verification means [265]. 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[168].

Model-Driven Engineering for Configuration 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 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 model to model transformations [69]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify model to model 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 [223]. 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 [68, 126, 117].

Platform Architecture for Activity Recognition. 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

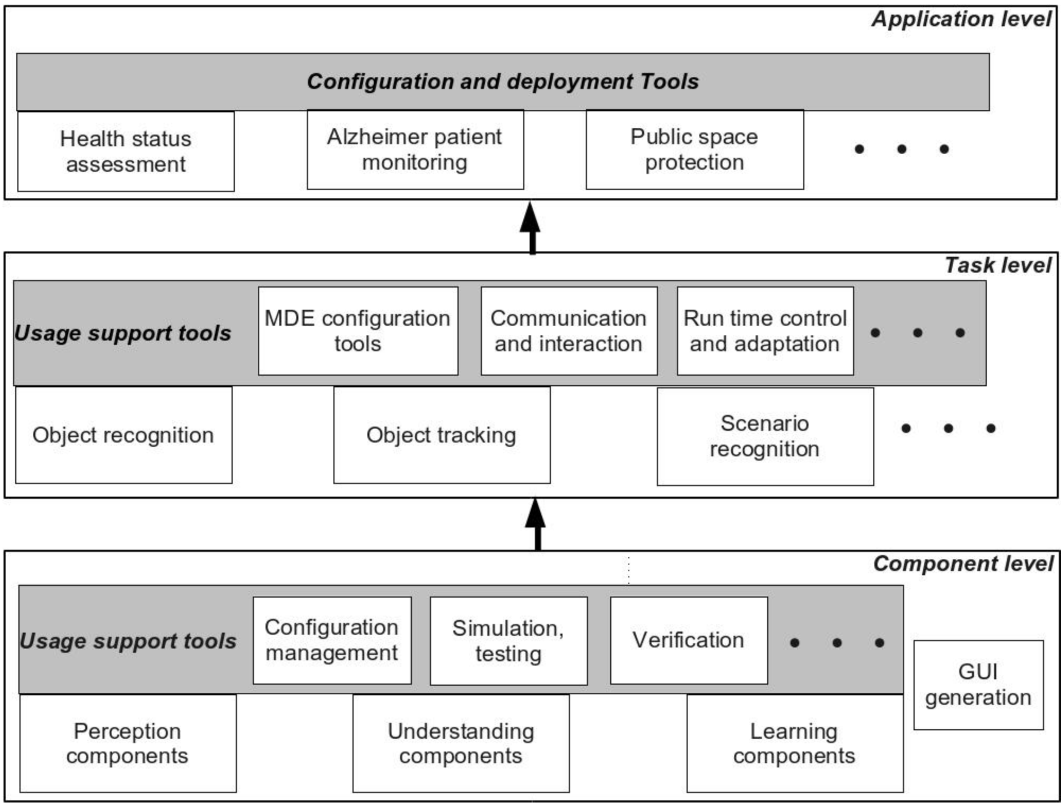

Figure 4: Conceptual architecture of an activity recognition platform.

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., health status monitoring, Alzheimer patient monitoring, public space protection, 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.

In the figure, the hashed areas contain software engineering support modules whereas the other modules correspond to software components (in the bottom two levels) or to generated systems (at Application level).

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

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

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving, etc. Note that, in figure 4, some of the boxes can be naturally adapted from Lama and 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).

Platform Adaptivity As it is mentioned in 3.1.2, our goal is to offer a multi sensor approach in activity recognition system generation. An underlying problem is to take into account the possible appearance and disappearance of sensors when doing such a generation. In collaboration with the Rainbow team (Nice-Sophia Antipolis University) we proposed to federate the inherent constraints of an activity recognition platform like SUP with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team proposes a component-based adaptive middleware (WComp) to dynamically adapt and recompose assemblies of components. The SUP platform gathers a set of modules devoted to design applications in the domain of activity recognition. Indeed, the services provided by SUP can be seen as complex high-level services associated to components in WComp. Then these SUP components will be discovered dynamically by WComp. However, some services turn out to be critical. Thus, according to our safeness concern, we studied how to maintain consistency of applications designed in WComp in spite of changes of critical components and conflicts that may appear when we compose some component assemblies. Relying on a synchronous modeling , we provide a well defined representation of component behaviors. In such a setting, model checking techniques are applied to ensure that concurrent access does not violate expected and acceptable behaviors of critical components [124]. This complementary research axis needs to be improved (using others formal methods as Abstract Interpretation technique) to be efficient. It is a short-term goal.

4 Application Domains

Stars will focus its work on two main application domains: **healthcare monitoring** and **visualsurveillance**.

4.1 Healthcare monitoring

As described in section 3, we have initiated a new strategic partnership (called CobTek) with Nice hospital [181, 34] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring. These new studies will address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

To achieve this objective, four topics need to be tackled. These topics have been previously described in section 3 and can be summarized within two points: finer activity description and longer analysis. Finer activity description is needed for instance, to discriminate the activities (e.g. sitting, walking, eating) of Alzheimer patients from the ones of healthy older people. It is essential for the early diagnosis of dementia and for providing a better and more specialised care. Longer analysis is required when people monitoring aims at measuring the evolution of patient behavioral disorders. Setting up such long experimentation with dementia people has never been tried before but is necessary to have real-world validation. This is one of the challenge of the European FP7 project Dem@Care where several patient homes should be monitored over several months.

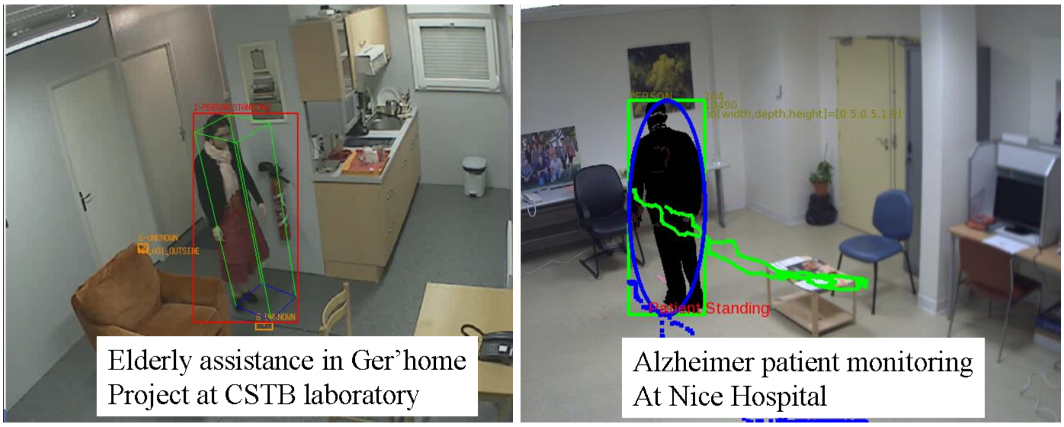

Figure 5: Examples of activity monitoring for healthcare

4.2 Visualsurveillance

Our experience in visualsurveillance [163, 148, 13] (also referred as video analytics) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

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

Moreover new challenges are arising from the visualsurveillance community. For instance, people detection and tracking in a crowed 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.

5 Ethical and Privacy Issues

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

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

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

6 Objectives for the next 4 years

The general objectives presented in the previous sections are long term goals. Depending on the human resources and funding we will focus our activity on the following priorities:

1. Perception for Activity Recognition:

   - Video segmentation and tracking: the image segmentation algorithms currently used do not take fully into account the temporal information. We plan to study 3D volumic representation of videos (2D space + 1D time), which rewrites the problems of segmentation and tracking as a search for coherent trajectories. A PhD student, Ratnesh Kumar, has started to work on this topic, since January 2011.
   - Optimization: to improve segmentation quality and speed, we regularly have to deal with optimization techniques and theory. Depending on needs, we may work on different generic optimization tools such as graph-cuts.
   - Shape statistics: the information retrieved from the shape of the objects detected is currently limited. We plan to use the shape statistics developed in [73] for the segmentation of images or videos with shape prior, or equivalently, depending on the needs of further activity recognition, for a finer classification of postures and gestures.
   - Activity recognition: currently Stars work is limited to coarse motion detection, global 3D features and simple posture detection. We plan to refine activity description with respect to two different approaches. First, we plan to use new sensors such as 3D motion sensors (e.g. Kinect) to take advantage of precise 3D information. Second, based on recent work [96] on real time gesture recognition, we will design novel spatio-temporal features with PhD student (P. Bilinski) [107] to detect actions in complex scenes (e.g. static and dynamic

occlusions, moving background, interactions between moving objects and interactions with contextual objects). These kinds of detailed spatio-temporal features are required for instance for patient monitoring.

- Controlling tracking algorithms: to study for which hypotheses tracking algorithms are effective, and to understand their limits. The goal is to be able to manage a library of video processing programs to select or combine the best programs depending on the scene conditions. This topic is currently under process with PhD D.P. Chau [75] and J. Badie who are studying the relationships between algorithm parameters, scene conditions and algorithm performance. The objective is then to optimize the use of tracking algorithms and their combinations.

2. Semantic Activity Recognition:

   - Uncertainty management: concerning the modeling of activity we plan to design a general formalism (based on statistical techniques) for characterizing the uncertainty of the recognized activities. Managing the uncertainty of vision features (in particular the lost of tracked objects) is a first crucial point. To reach this objective, we will extend the Pulsar Scenario Description Language for modeling explicitly the uncertainty of vision features. Together with this language extension, we would like also to propose mechanisms to propagate this uncertainty through all the layers of the activity recognition process. Another issue is to keep the real-time performance of the recognition process.
   - Activity learning: for activity learning, we have proposed with PhD M. Zuniga and PhD G. Pusiol [62, 82] several algorithms to cluster people trajectories and learn activity models in monitoring applications. However, the learned frequent models are not always interesting and scenarios of interest are not necessary frequent. Thus, we are planning to refine our previous approach on learning simple activities (e.g. primitive events) [139] by adding more semantic information in order to learn more high level activities. For instance, contextual information (i.e. zones of interest) and more meaningful vision features (e.g. local motion descriptors) can be combined into spatio-temporal features in order to guide the extraction of the activities of interest. Another challenge is to understand at which granularity level the activities should be learned. This topic is starting with PhD G. Pusiol [99, 123] and will expand through the older person monitoring projects. This issue becomes essential while dealing with video monitoring applications involving a large amount of everyday activities.

3. Software Engineering for Activity Recognition:

   - Scenario recognition: we shall complete our approach based on formal models and we shall explore how to take uncertainty into account. This implies to improve our framework for synchronous modeling and verification.
   - Methodology: we shall continue to explore the applicability of model driven development (MDD) to activity recognition systems. Of course, setting up a complete methodology is a long term objective and within a period of 4 years we only expect feasibility results.

- Run-time adaptation: to react to environment changes we shall favor the "model at run-time" approach combined with Artificial Intelligence techniques gained from our past experience with Program Supervision [242, 14].
- Large scale test with SUP: we intend to instrument (a subset of) the SUP platform with models and tools supporting the configuration, deployment and run-time adaptation of video surveillance systems following the previous methodology.
- Improvement of platform architecture: such an instrumentation may require a study of internal concurrency, functional distribution, and real-time performance. It also necessitates user interfaces and tools in particular to support the methodological aspects. These are indeed time-consuming activities and development in these domains will depend on the available work power.

7 Evaluation

We are aware that activity recognition is an ambitious topic. Our concrete goal in terms of academic impact is to publish in computer vision (in particular CVIU and IJCV (or TPAMI) international journals, (as well as ICCV, CVPR, ECCV international conferences) in artificial intelligence (IJCAI conference, AI Journal) and in machine learning (Machine Learning journal, NIPS conference).

We also consider conferences and journals linking Software Engineering and Artificial Intelligence (such as SEKE or Data and Knowledge Engineering) as well as the ones more specifically devoted to Software engineering (ICSE) with a particular emphasis on those promoting the model-driven approaches (MoDeLs conference, SoSym journal). We also intend to be actors in the major conferences in vision and cognitive systems (such as ICVS) and in visual surveillance (such as AVSS).

The Stars evaluation should also include the assessment of the produced software. However activity recognition systems are difficult to evaluate, since their performance depends directly on the test videos and on the system configuration phase. A robust algorithm tested in new environmental conditions or with inadequate parameters can get poor performance. Moreover the application domains require to take into account a huge variety of real-world conditions (e.g. night/day, crowded scenes). Thus, we are collaborating in the QUASPER evaluation program to better define the evaluation criteria (e.g. rules to build the ground-truth) and to design novel evaluation metrics taking into account the dependencies of video characteristics (e.g. contrasted shadows, illumination changes) including their acquisition conditions and the configuration phases (e.g. parameter tuning, object model learning) of the tested algorithms. We will also continue to participate through industrial cooperations to performance evaluation campaigns made with data taken with the existing operational video cameras on real sites.

We have also set up two short term (4 years) concrete goals in terms of activity recognition systems, which are :

- to design manually a real-time recognition system for complex and long activities involving interactions between a person and his/her environment based on postures with fixed large field of view cameras.
- to automatically quantify the performances of a new activity recognition system based on benchmarks and ground-truth.

We have also set up two long term concrete goals in terms of activity recognition systems, which are :

- to automate the set-up (parametrization) of an activity recognition system (such as an activity monitoring system to prevent people behavior disorders) for a new site.
- to generate semi-automatically new activity recognition systems using MDE-based methodologies and high level specifications (in terms of activity to recognize and constraints).

8 Positioning and Collaborations

8.1 Positioning within INRIA

Stars has an original multidisciplinary approach mixing techniques from the domain of perception, cognitive systems and software engineering for activity recognition.

Perception through video streams

Several INRIA teams are working on video analysis for human activity description and thus are close to Stars objectives. Currently we do not have effective collaborations with these teams (except with the PRIMA team), but we intend to develop new collaborations on certain topics.

- The **PRIMA** research team is very close to Stars as it shares with Stars a cognitive system approach. They work more on 2D appearance-based approaches whereas Stars will design 4D spatio-temporal video understanding approaches on long term period (e.g. several days) to recognize semantic activities.
- The **WILLOW** research team is interested in developing geometric, physical, and statistical models for all components of the image interpretation process, including illumination, materials, objects, scenes, and human activities. More recently they have been working on motion event recognition (cf Y. Laptev's work on repetitive motion detectors). Stars's work is complementary to these approaches since we are using 3D models and human expertise. A new cooperation with WILLOW for learning new motion event detectors would be certainly fruitful.
- The **LEAR** research team is working on object recognition for static images and video sequences. They focus on learning low level visual features. In particular they use shape and affine interest point information to describe objects in images. This numerical content is learned to build visual models of object classes. A new cooperation with LEAR is certainly interesting for learning new shape and texture descriptors.

Personally Assisted Living (PAL)

We are involved in a large scale INRIA initiative (called Personally Assisted Living) gathering several INRIA teams : in robotics (COPRIN, AROLAG, LAGADIC, E-MOTION, DEMAR), in vision (PRIMA, STARS), in artificial intelligence (MAIA) and in software engineering (TRIO). The objective of this project is to create a research infrastructure that will enable experiments with technologies for improving the quality of life for persons who have suffered a loss of autonomy through age, illness or accident. In particular, the project seeks to enable development of technologies that can provide services for older and fragile persons, as well as their immediate family, caregivers and social groups. For instance, these services should allow older people to stay safely at home, to benefit from an automated medical supervision and should delay their entrance in nursing homes.

For Stars, one of the objectives of this collaboration is the early detection of deteriorated health status. For example, Stars is interested to collaborate with the **COPRIN** and **AROLAG** teams to combined the analysis of cameras mounted on a robot with cameras installed in the apartment of an older people. The finer analysis could enable a better understanding of the older people behavior and more appropriate services.

Perception through other sensors

For audio sensor understanding cooperations with the projects of the INRIA theme *Audio, Speech, and Language Processing* are possible and interesting, in particular with PAROLE and/or METISS.

Knowledge management

We share common interests in knowledge representation, knowledge acquisition, ontologies, machine learning and reasoning with several INRIA project-teams mainly in INRIA theme *Knowledge and Data Representation and Management*. Among them we can cite the following ones:

- The **MAIA** project is focused on autonomous real-time robotic systems. We share common objectives on real-time artificial intelligence techniques for signal understanding and on healthcare applications.
- The **WIMMICS** project develops knowledge management and ontologies techniques mainly for Semantic Web applications. We had informal cooperations with WIMMICS on ontology languages and Human Machine Interfaces which could be important to continue.

Optimization and Statistical Learning

We share common interests in machine learning, artificial intelligence and optimization with several INRIA project-teams in the theme *Optimization, Learning and Statistical Methods*. Among them we can cite for instance the following one:

- The **TAO** project focuses on research that draws on the insight that machine learning and optimization are closely related and need to be considered together: on one hand, typical machine learning problems (e.g., identifying the best model) are basically optimization problems, and on the other hand, optimization problems require adaptation and thus machine learning for best performance. Their two biggest research topics are "Learning and Uncertainty" and "Stochastic Optimization and Automatic Parameter Tuning". Cooperation in any of these two fields would be fruitful.

Software engineering

We have close contacts with AOSTE concerning synchronous models and with TRISKELL about software modeling and meta-modeling.

- The **AOSTE** project tackles several topics in the design methodologies for synchronous reactive systems. Particularly, it has proposed the MARTE UML profile which allows to define the temporal semantic of models as logical clocks with constraints. A cooperation with the AOSTE project could be useful to help us defining the different models of time that we need for setting up models of scenarios.

- The **TRISKELL** project focuses on the reliable and efficient design of applications by assembling software components. In particular, TRISKELL has developed the KERMETA workbench which allows to define both the structure and the behavior of metamodels. We belong to the same community of model-driven engineering.

Several other INRIA project-teams within the themes *Embedded and Real Time Systems* (ESPRESSO) and *Distributed Systems and Services* (ADAM, PHOENIX) deal with component-based programming and run-time adaptation.

3D simulation of real-world scenes
We have similar interests for the 3D simulation of real-world scenes with INRIA project-teams in the theme *Interaction and Visualization*. For instance MIMETIC is proposing an original approach to allow real and virtual humans to naturally interact in a shared virtual environment. It will be also interesting to collaborate with REVES to study the behaviors of older people especially when they are observed in contact with virtual environment.

8.2 Academic collaborations and positioning

8.2.1 National

We share common interests with several research teams in France, outside INRIA.

Video Analysis

- We cooperate with Thales TSS in the Video-Id project, with Thales Com in the VANAHEIM project and with Thales Theresis in the VICOMO and QUASPER projects. We use Thales expertise in videosurveillance to design novel algorithms such as people re-identification [89] or data-mining for activity discovery [19].
- We regularly cooperate with the research team directed by Patrick Sayd at CEA in Saclay in the context of videosurveillance [42]. Currently, we work together in the context of the Quasper and VICOMO projects. They focus in 2D video analysis for people tracking over networks of video cameras while we are in charge of high level scenario understanding.

Software Engineering

- We plan to continue our strong cooperation with the Modalis team (I3S-UNSA-CNRS) at Sophia Antipolis on the model engineering approach which has already produced joint papers [69, 68, 103].
- For dynamic component management, we are exploring the component-based adaptive middleware (WComp [269, 268, 78]) developed in the Rainbow team (I3S-UNSA-CNRS) at Sophia Antipolis.
- We plan to continue our tight collaboration with the MCSOC team (LEAT-UNSA-CNRS).

Ambient Assisted Living

We have initiated a series of new cooperations in the domain of healthcare. These cooperations are a first step which need to be enriched and supported.

- We will deepen our collaboration with CSTB (Centre Scientifique et Technique du Batiment) and the Nice City Hospital (Groupe de Recherche sur la Trophicité et le Viellissement) in the framework of the CIU-santé project, funded by the DG-CIS (i.e. industry minister). This project is devoted to experiment and develop techniques that allow long-term monitoring of people at home. In this project an experimental room has been designed in Nice City Hospital and is relying on the research of the Stars team concerning supervised and unsupervised event learning and recognition.

Energy

- We have a fruitful collaboration with CEA at Cadarache for nuclear fusion reactor. We have developed a prototype system set up on the French tokamak Tore Supra able to recognize automatically thermal events [24, 32] observed during plasma operation in order to protect the plasma facing components (real-time issue) and to better understand the plasma-wall interactions (plasma physics issues). The system may take multimodal data as inputs: plasma parameters (plasma density, injected power by heating antenna, plasma position...), infrared and visible images, and signals coming from other sensors as spectrometers and bolometers.

8.2.2 International

We have a precise view of the main groups working in our domains of interest through international networks gathering our main competitors and collaborators.

Europe: We are in competition (and in collaboration) with several European research teams (e.g. Prof A Cohn and Prof D. Hoggs from University of Leeds, UK, Prof Neumann from University of Hamburg) on activity recognition and (e.g. Prof. R. Cucchiara from University of Modena, Italy, Prof. G. Jones from University of Kingston, UK) on videosurveillance. We are also involved with some of these European research teams (e.g. University of Leeds, University of Hamburg) in the network of excellence EU-CogII - 2nd European Network (see http://www.eucognition.org) for the Advancement of Artificial Cognitive Systems, Interaction and Robotics. We also cooperate with many European research teams through European Projects (see next section on industrial collaborations and funding).

USA: We are competing with US academic research teams (e.g., Prof. Ram Nevatia from University of Southern California, Prof. A. Bobick for Georgia Tech, Prof. Larry Davis from University of Maryland) on the topic of video event recognition. We also cooperate with these teams for defining standard video event ontology (e.g., Prof. Nevatia from University of Southern California, Prof. R. Chelappa from University of Maryland) and for comparing video understanding algorithm performances (e.g. Prof. M. Shah from University of South Florida and Prof. Larry Davis from University of Maryland).

We also collaborate with Guillermo Sapiro's team, University of Minnesota, on the design of new shape distances and shape statistics.

We are exploring possible collaboration on software engineering for video analysis with the Colorado State University (Prof. R. France for software engineering and prof. B. Draper for video analysis).

Asia: We have been cooperating with the Multimedia Research Center in Hanoi MICA on semantics extraction from multimedia data and on videosurveillance issues [20]. Currently we are working in the Sweet-Home project on healthcare applications together with the National Cheng Kung University in Taiwan.

We share also common interests with the I2R-CVIU team in Singapore on many subjects such as videosurveillance and ambient assisted living.

8.2.3 General

Our team is participating to Quasper and CIU-santé, two academic and industrial platforms for the evaluation of video analysis techniques on videosurveillance and healthcare applications. These projects are supposed to involve a large number of international organisations in the future.

8.3 Industrial Collaborations and Funding

Stars will rely on a strong network of industrial partners mainly in the domain of videosurveillance and healthcare monitoring.

National Initiatives

- **CIU-Santé**: CIU-Santé (Centre d'Innovation et d'Usages en Santé) is a DGCIS French project from December 2008 to April 2012 on long-term monitoring of elderly people at Hospital with CSTB, the Nice City Hospital, Pole SCS, SagemCom, Movea, CEA and UNSA. In this project two experimental rooms should be built in Nice-Cimiez Hospital for monitoring Alzheimer patients.
- **SWEET-HOME**: SWEET-HOME is a ANR TECSAN French project from Nov 1st 2009 to 2012 (3 years) on long-term monitoring of elderly people at Hospital with Nice City Hospital, MICA Center (CNRS unit - UMI 2954) in Hanoi, Vietnam, SMILE Lab at National Cheng Kung University, Taiwan and National Cheng Kung University Hospital. INRIA Grant is 240 Keuros out of 689 Keuros for the whole project. SWEET-HOME project aims at building an innovative framework for modeling activities of daily living (ADLs) at home. These activities can help assessing elderly disease (e.g. Alzheimer, depression, apathy) evolution or detecting precursors such as unbalanced walking, speed, walked distance, psychomotor slowness, frequent sighing and frowning, social withdrawal with a result of increasing indoor hours.
- **Quasper** : QUAlification et certification des Systèmes de PERception

  Quasper R&D is a FUI project with national funding, from June 2010 to may 2012. It gathers 3 objectives to serve companies and laboratories:

  1. to encourage R&D and the design of new perception systems;
  2. to develop and support the definition of European standards to evaluate the functional results of perception systems;

3. to support the qualification and certification of sensors, software and integrated perception systems.

Target domains are Security, Transportation and Automotive. INRIA participates to the development of the evaluation platform. The coordinator is THALES SERVICES SAS and other partners are AFNOR, AKKA, DURAN, INRETS, SAGEM SECURITE, ST MICROELECT., THALES RT, VALEO VISION SAS, CEA, CITILOG, INSTITUT d'OPTIQUE, CIVITEC, SOPEMEA, ERTE, HGH.

- **HWB** : Sharing resources for HWB Living labs is a 1 year ANR project starting at the end of 2011. The main objective of this project is to bring applications closer to market through the use of (shared) Experience & Living Lab (ELL) facilities, and to attract a wider range of (SME) companies to the benefits of ELL facilities. These common labs and testbeds are experimentation environments (facilities) that can help companies (especially SMEs) and research institutions to improve the assessment of novel ideas and concepts. The French ecosystem will bring close interaction between the medical specialists in geriatrics and Alzheimer (e.g Prof Philippe Robert team Centre Memoire de Ressources et de Recherche CMRR at Nice CHU hospital) and the advanced research institutes in information technology (e.g. INRIA). The leader of the project is Novay (The Netherlands) and the other partners are: Philips (The Netherlands), TU Eindhoven (The Netherlands), KTH (Sweden), INRIA (France), TU Lulea (Sweden), FhG (Germany) and Ericsson (Sweden).

European Projects

- **ViCoMo** is a ITEA 2 European Project which has started on the 1st of October 2009 and will last 36 months.

  The project is executed by a strong international consortium, including large high-tech companies (e.g. Philips, Acciona, Thales), smaller innovative SMEs (CycloMedia, VDG Security, Keeneo) and is complemented with relevant research groups and departments from well-known universities (TU Eindhoven, University of Catalonia, Free University of Brussels) and research institutes (INRIA, CEA List). Participating countries are France, Spain, Finland, Turkey and Netherlands.

  The ViCoMo project is focusing on the construction of realistic context models to improve the decision making of complex vision systems and to produce a faithful and meaningful behavior. Its goal is to find the context of events that are captured by the cameras or image sensors, and model this context such that reliable reasoning about an event can be established.

- **VANAHEIM** (Video/Audio Networked surveillance system enhAncement through Human-cEntered adaptIve Monitoring) is a FP7-ICT-2009-4 Cognitive Systems and Robotics IP European project which started on February 1st 2010 and will end in 2013 (42 months). The aim of this proposal is to study innovative surveillance components for the autonomous monitoring of multi-sensory and networked infrastructure such as underground transportation environment. The prime partner is Multitel (Belgium) and others partners are: INRIA Sophia-Antipolis (France), Thales Communications (France), IDIAP (Switzerland), Torino GTT (Italy) and Régie Autonome des Transports Parisiens RATP (France), Ludwig Boltzmann Institute for Urban Ethology (Austria) and Thales Communications (Italy). INRIA Grant is 700 Keuros out of 3 718 Keuros for the whole project.

- **SUPPORT** (Security UPgrade for PORTs) is a FP7-SEC-2009-1 IP European project which started in June 2010 and will end in 2014 (48 months). The aim of this proposal is to study innovative surveillance components for Autonomous Monitoring Of Multi-Sensory And Networked Infrastructure Such As Underground Transportation Environment. INRIA Grant is 367.2 Keuros out of 9920 Keuros for the whole project. The prime partner is BMT Group (UK) and other partners are: INRIA Sophia-Antipolis (France), Swedish Defence Research Agency (SE), Securitas (SE), Technical Research Centre of Finland (FI), MARLO (NO), INLECOM Systems (UK)...

- **EIT ICT Labs** is one of the first three Knowledge and Innovation Communities (KICs) selected by the European Institute of Innovation & Technology (EIT) to accelerate innovation in Europe. EIT is a new independent community body set up to address Europe's innovation gap. It aims to rapidly emerge as a key driver of EU's sustainable growth and competitiveness through the stimulation of world-leading innovation. Among the partners, there are strong technical universities (U Berlin, 3TU / NIRICT, Aalto University, UPMC - Université Pierre et Marie Curie, Université Paris-Sud 11, Institut Telecom, The Royal Institute of Technology); excellent research centres (DFKI, INRIA, Novay, VTT, SICS) and leading companies (Deutsche Telekom Laboratories, SAP, Siemens, Philips, Nokia, Alcatel-Lucent, France Telecom, Ericsson). This project is largely described at `http://eit.ictlabs.eu`. For the EIT ICTLabs TAL Health and Well Being, Stars is involved in one action in 2012 (Indoor Physical Activity).

- **Dem@care** (Dementia Ambient Care) is a FP7-ICT-2011.5.1 IP European project (Personal Health Systems) started on November 2011 and will end in 2015 (48 months). This project addresses Challenge 5: Digital ICT for Health, Ageing Well, Inclusion and Governance.

  The objective of Dem@Care is the development of a complete system providing personal health services to persons with dementia, as well as medical professionals, by using a multitude of sensors, for context-aware, multiparametric monitoring of lifestyle, ambient environment, and health parameters. Multisensor data analysis, combined with intelligent decision making mechanisms, will allow an accurate representation of the person's current status and will provide the appropriate feedback, both to the person and the associated medical professionals. Multi-parametric monitoring of daily activities, lifestyle, behaviour, in combination with medical data, can provide clinicians with a comprehensive image of the person's condition and its progression, without their being physically present, allowing remote care of their condition. There will be several pilots in France, Ireland and Sweden. The prime partner is the Centre for Research and Technology Hellas (G) and other partners are: INRIA Sophia-Antipolis, University de Bordeaux 1, Cassidian, Nice Hospital, LinkCareServices (FR), Lulea Tekniska Universitet (SE), Dublin City University (IE), IBM Israel (IL), Philips (NL) and Vistek ISRA Vision (TR).

Industrial Cooperations

- **Thales**: we are involved in a long-term research collaboration with Thales through several contracts, including a direct industrial contract, 3 past European projects (Advisor, SERKET, CARETAKER), 2 running European projects (VICOMO, VANA-

HEIM), 2 past national projects (the SIC project within the security "pole de compétivité" SYSTEM@TIC, Video-ID) and a running one (QUASPER).

- **Keeneo**: the start-up Keeneo was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in videosurveillance (VSIP library). Stars will continue to maintain a close cooperation with Keeneo for impact analysis of VSIP and for exploitation of new results. We are currently continuing this strategic cooperation through the common supervision of a PhD student M. Souded.

- **Link Care Services**: this SME enterprise develops and operates video monitoring solutions for dependent people. An engineer of Link Care Services has been working within Stars since December 2011.

9 Teaching

Stars members have been involved for many years in various teaching activities in software engineering and in the domain of video interpretation, both at the master and doctorate level. Below are few examples:

- Stars is a hosting team for the Master of Computer Science of UNSA.

- François Brémond : Teaching at EURECOM Master at Sophia Antipolis, 3hours per year, class on video understanding techniques.

- Annie Ressouche : Teaching at Master of Computer Science at Polytechnic School of Nice Sophia Antipolis University, lecture on Synchronous Languages and Verification (6h).

- Annie Ressouche: Teaching at Ecole des Mines (Sophia Antipolis), lecture on Scade Toolkit and Verification (4h).

- Jean-Paul Rigault : Full professor at Polytech'Nice Sophia, the engineering school of the University of Nice-Sophia Antipolis ; teaching C, C++ (beginner and advance levels), system programming, software engineering (about 200 hrs per year) ; also teaching advanced C++ programming and object-oriented design for several industrial companies.

Stars members will continue their teaching activities in video interpretation systems, including the supervision of master, doctorate and post-doctorate students.

10 Team Bibliography since 2006

10.1 Book Chapters

2007

[1] Vincent Martin and Monique Thonnat. *Scene Reconstruction, Pose Estimation and Tracking*, chapter A Learning Apporach for Adaptive Image Segmentation, pages 431– 454. ISBN 978-3-902613-06-6, Rustam Stolkin edition, 2007.

2008

[2] Vincent Martin and Monique Thonnat. *A Cognitive Vision Approach to Image Segmentation*, chapter 16, pages 265–294. InTech Education and Publishing, Aug 2008.

[3] Annie Ressouche, Daniel Gaffé, and Valerie Roy. Modular compilation of a synchronous language. In Roger Lee, editor, *Software Engineering Research, Management and Applications*, volume 150 of *Studies in Computational Intelligence*, pages 151–171. Springer, 2008.

2009

[4] S. Moisan, , and J.-P. Rigault. Teaching object-oriented modeling and uml to various audiences. In Sudipto Ghosh, editor, *Workshops and Symposia at MODELS 2009, Reports and Revised Selected Papers*, volume 6002 of *LNCS*. Springer-Verlag, Denver, CO, USA, 2009. Subseries: Programming and Software Engineering.

2010

[5] Guillaume Charpiat, Matthias Hofmann, and Bernhard Schölkopf. Kernel methods in medical imaging. In Paragios, N., J. Duncan, and N. Ayache, editors, *Handbook of Biomedical Imaging*. Springer, August 2010.

[6] Jose Luis Patino, Francois Bremond, and Monique Thonnat. Incremental learning on trajectory clustering. In Dr. Paolo Remagnino, editor, *Intelligent Paradigms in Safety and Security*. Springer-Verlag, 12 2010.

2011

[7] Guillaume Charpiat, Ilja Bezrukov, Yasemin Altun, Matthias Hofmann, and Bernhard Schölkopf. Machine Learning Methods for Automatic Image Colorization. In R. Lukac Editor, editor, *Computational Photography: Methods and Applications*. CRC Press, January 2011.

[8] M. Zuniga, F. Bremond, and M. Thonnat. *Uncertainty control for reliable video understanding on complex environments*, chapter Video Surveillance. InTech, Vienna Austria, February 2011. ISBN 978-953-307-436-8.

10.2 Articles in Referred Journals

2006

[9] B. Boulay, F. Bremond, and M. Thonnat. Applying 3d human model in a posture recognition system. *Pattern Recognition Letter, Special Issue on vision for Crime Detection and Prevention*, 27(15):1788–1796, 2006.

[10] F. Bremond, M. Thonnat, and M. Zúniga. Video Understanding Framework For Automatic Behavior Recognition. *Behavior Research Methods Journal*, 38(3):416–426, 2006.

[11] N. Khayati, W. Lejouad-Chaari, S. Moisan, and J.P. Rigault. Distributing knowledge-based systems using mobile agents. *WSEAS Transactions on Computers*, 5(1):22–29, January 2006.

[12] N. Maillot and M. Thonnat. Ontology based complex object recognition. *Image and Vision Computing Journal, Special Issue on Cognitive Computer Vision*, 2006. To appear, available on line July 2006.

2007

[13] Florent Fusier, Valery Valentin, François Bremond, Monique Thonnat, Mark Borg, David Thirde, and James Ferryman. Video understanding for complex activity recognition. *Machine Vision and Applications Journal*, 18:167–188, 2007.

[14] Benoit Georis, François Bremond, and Monique Thonnat. Real-time control of video surveillance systems with program supervision techniques. *Machine Vision and Applications Journal*, 18:189–205, 2007.

[15] J.P. Vidal, S. Moisan, and D. Faure, J.B.and Dartus. River model calibration, from guidelines to operational support tools. *Environmental Modelling and Software*, 22(11):1628–1640, November 2007.

2008

[16] P. Boissard, V. Martin, and S. Moisan. A cognitive vision approach to early pest detection in greenhouse crop s. *International Journal of Computers and Electronics in Agriculture*, 62(2):81–93, July 2008.

[17] Paul Boissard, Vincent Martin, and Sabine Moisan. A cognitive vision approach to early pest detection in greenhouse crops. *Computer and Electronics in Agriculture*, 62(2):83–93, April 2008.

[18] M. Hofmann, F. Steinke, V. Scheel, G. Charpiat, J. Farquhar, P. Aschoff, M. Brady, B. Schölkopf, and B. J. Pichler. MR-based attenuation correction for PET/MR: A novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine*, 11 2008.

[19] Jose Luis Patino, hamid Benhadda, Etienne Corvee, Francois Bremond, and Monique Thonnat. Extraction of activity patterns on large video recordings. *IET Computer Vision*, 2(2):108–128, june 2008.

2009

[20] Thi Lan Le, Monique Thonnat, Alain Boucher, and François Bremond. Surveillance Video Indexing and Retrieval using Object Features and semantic Events. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(7):pp 1439–1476, November 2009.

[21] Monique Thonnat and Olivier Guérin. Vision cognitive : Au-delà de l'apparence par Monique Thonnat. Santé à domicile, entretien avec Olivier Guérin, propos recueillis par Dominique Chouchan. *La Recherche. Les Cahiers de l'Inria*, (428), March 2009.

[22] Nadia Zouba, Francois Bremond, Monique Thonnat, Alain Anfosso, Eric Pascual, Patrick Mallea, Véronique Mailland, and Olivier Guerin. A computer system to monitor older adults at home: Preliminary results. *Gerontechnology Journal*, 8(3):129–139, July 2009.

2010

[23] Slawomir BAK, Jacek Blazewicz, Grzegorz Pawlak, Maciej Plaza, Edmund K. Burke, and Graham Kendall. A parallel branch and bound approach to the rectangular guillotine strip cutting problem. *INFORMS Journal on Computing, JOC*, 2010.

[24] V. Martin, JM. Travere, V. Moncada, F. Bremond, and G. Dunand. Thermal event recognition applied to protection of tokamak plasma facing components. *IEEE Transactions on Instrumentation and Measurement*, 59(5):1182–1191, May 2010. Digital Object Identifier: 10.1109/TIM.2009.2038032.

[25] Sabine Moisan. Generating Knowledge-Based System Generators: a Software Engineering Approach. *International Journal of Intelligent Information Technologies*, 6(1):1–17, January 2010.

[26] Monique Thonnat. Comment vieillir chez soi. *La Recherche. Les Cahiers de l'Inria*, (446 novembre 2010), November 2010.

[27] Monique Thonnat. Special issue on Intelligent Vision Systems. *Computer Vision and Image Understanding*, 114(5):pp 501–608, May 2010.

[28] Nadia Zouba, François Bremond, Alain Anfosso, Monique Thonnat, Eric Pascual, and Olivier Guerin. Monitoring elderly activities at home. *Gerontechnology*, 9(2), May 2010.

2011-2012

[29] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Boosted human re-identification using Riemannian manifolds. *Image and Vision Computing*, August 2011.

[30] Ahmed Gamal Eldin, Xavier Descombes, Guillaume Charpiat, and Josiane Zerubia. Multiple Birth and Cut Algorithm for Multiple Object Detection. *Journal of Multimedia Processing and Technologies*, 2012.

[31] M. Kaaniche and F. Bremond. Recognizing gestures by learning local motion signatures of hog descriptors. *Transactions on Pattern Analysis and Machine Intelligence(TPAMI)*, 2012.

[32] V. Martin, JM. Travere, V. Moncada, G. Dunand, and F. Bremond. Towards intelligent video understanding applied to plasma facing component monitoring. *Contributions to Plasma Physic*, 51(2-3):252–255, March 2011.

[33] Annie Ressouche and Daniel Gaffé. Compilation Modulaire d'un Langage Synchrone. *Revue des sciences et technologies de l'information, série Théorie et Science Informatique*, 4(30):441–471, June 2011.

[34] R. Romdhane, E. Mulin, A. Derreumeaux, N. Zouba, J. Piano, L. Lee, I. Leroi, P. Mallea, R. David, M. Thonnat abd F. Bremond, and P.H. Robert. Automatic video monitoring system for assessment of alzheimer's disease symptoms. *The Journal of Nutrition, Health and Aging Ms(JNHA)*, JNHA-D-11-00004R1, 2011.

[35] Marcos Zuniga, Francois Bremond, and Monique Thonnat. Hierarchical and incremental event learning approach based on concept formation models. *Neurocomputing, Special Issue: Behaviours in Video*, 2011.

[36] Marcos Zuniga, Francois Bremond, and Monique Thonnat. Real-time reliability measure-driven multihypothesis tracking using 2D and 3D features. *EURASIP EURASIP Journal on Advances in Signal Processing*, 2011:142, December 2011.

10.3 Publications in International Referred Conferences

2006

[37] C. Carincotte, X. Desurmont, B. Ravera, F. Bremond, J. Orwell, S. Velastin, J-M. Odobez, B. Corbucci, J. Palo, and J. Cernocky. Toward generic intelligent knowledge extraction from video and audio: the eu-funded caretaker project. In *IET conference on Imaging for Crime Detection and Prevention (ICDP 2006)*, London, Great Britain, June 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[38] B. Georis, M. Mazière, F. Bremond, and M. Thonnat. Evaluation and knowledge representation formalisms to improve video understanding. In *Proceedings of the International Conference on Computer Vision Systems (ICVS'06)*, New-York, NY, USA, January 2006.

[39] M. B. Kaâniche, F. Bremond, and M. Thonnat. Monitoring trichogramma activities from videos : An adaptation of cognitive vision system to biology field. In *International Cognitive Vision Workshop, (ICVW'2006) in conjunction with the 9th European Conference on computer Vision (ECCV'2006)*, Graz, Austria, 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICVW06.html.

[40] Vincent Martin, Nicolas Maillot, and Monique Thonnat. A learning approach for adaptive image segmentation. In *Proceedings of the International Conference on Computer Vision Systems (ICVS'06)*, New-York, NY, USA, January 2006.

[41] A. Toshev, F. Bremond, and 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.

[42] T. Vu, F. Bremond, G. Davini, M. Thonnat, QC. Pham, N. Allezard, P. Sayd, JL. Rouas, S. Ambellouis, and A. Flancquart. Audio video event recognition system for public transport security. In *IET conference on Imaging for Crime Detection and Prevention (ICDP 2006)*, London, Great Britain, June 2006. http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[43] M. Zhan, P. Remagnino, S. Velastin, F. Bremond, and M. Thonnat. Matching gradient descriptors with topological constraints to characterise the crowd dynamics. In *The 3rd International Conference on Visual Information Engineering (VIE2006)*, pages 441–446, Bangalore, India, September 26-28 2006.

[44] M. Zúniga, F. Bremond, and M. Thonnat. Fast and reliable object classification in video based on a 3d generic model. In *The 3rd International Conference on Visual Information Engineering (VIE2006)*, pages 433–441, Bangalore, India, September 26-28 2006.

2007

[45] H. Benhadda, J.L. Patino, E. Corvee, F. Bremond, and M. Thonnat. Data mining on large video recordings. In *Veille Strategique Scientifique et Technologique VSST 2007*, Marrakech, Morocco, 21st -25th October 2007.

[46] A. T. Nghiem, F. Bremond, M. Thonnat, and R. Ma. A new evaluation approach for video processing algorithm. In *Proceedings of WMVC 2007 IEEE Workshop on Motion and Video Computing*, Austin, Texas, USA, February 2007.

[47] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin. Etiseo: an evaluation project for video surveillance systems. In *Proceedings of AVSS2007, IEEE International Conference on Advanced Video and Singal based Surveillance*, London, UK, September 2007.

[48] J.L. Patino, H. Benhadda, E. Corvee, F. Bremond, and M. Thonnat. Video-data modelling and discovery. In *International Conference on Visual Information Engineering VIE 2007*, London, UK, 25th -27th July 2007.

[49] J.L. Patino, E. Corvee, F. Bremond, and M. Thonnat. Management of large video recordings. In *Ambient Intelligence Developments AmI.d 2007*, Sophia Antipolis, France, 17th -19th September 2007.

[50] Lan Le Thi, Alain Boucher, and Monique Thonnat. Subtrajectory-based video indexing and retrieval. In *13th International MultiMedia Modeling Conference (MMM)*, Singapore, 2007.

[51] N. Zouba, F. Bremond, M. Thonnat, and V. T. Vu. Multi-sensors analysis for everyday elderly activity monitoring. In *Proceedings of SETIT 2007 the 4th International Conference Sciences of Electronics, Technologies of Information and Telecommunications*, Tunis, Tunisia, March 2007.

2008

[52] G. Charpiat, M. Hofmann, and B. Schölkopf. Automatic image colorization via multimodal predictions. In D. Forsyth, P. Torr, and A. Zisserman, editors, *10th European Conference on Computer Vision*, pages 126–139, Berlin, Germany, 10 2008. Springer.

[53] D. Gaffé and A. Ressouche. The clem toolkit. In *Proceedings of 23rd IEEE/ACM International Conference on Automated Software Engineering (ASE 2008)*, L'aquila, Italy, September 2008.

[54] Thi-Lan Le, Alain Boucher, Monique Thonnat, and Francois Bremond. A framework for surveillance video indexing and retrieval. In *International Workshop on Content-Based Multimedia Indexing (CBMI'08)*, pages 338–345, UK, 18-20 June 2008.

[55] Thi-Lan Le, Monique Thonnat, Alain Boucher, and Francois Bremond. A query language combining object features and semantic events for surveillance video retrieval. In *The International MultiMedia Modeling Conference (MMM'08)*, pages 307–317, Kyoto, Japan, 9-11 January 2008.

[56] V. Martin and S. Moisan. Early pest detection in greenhouses. In *Visual Observation and Analysis of Animal and Insect Behavior (VAIB) Workshop at ICPR*, Tampa, Fl. USA, December 2008.

[57] V. Martin, S. Moisan, B. Paris, and O. Nicolas. Towards a video camera network for early pest detection in greenhouses. In *International Conference Endure 2008*, La Grande Motte, France, October 2008.

[58] V. Martin and M. Thonnat. Learning contextual variations for video segmentation. In *Proceedings of the 6th IEEE Int. Conf. on Computer Vision Systems (ICVS'08)*, pages 464–473, Santorini, Greece, May 2008. IEEE Computer Society.

[59] S. Moisan. Component-based support for knowledge-based systems. In *ICEIS, 10th International Conference on Enterprise Information Systems*, Barcelona, Spain, June 2008.

[60] Anh-Tuan Nghiem, Francois Bremond, and Monique Thonnat. Shadow removal in indoor scenes. In *IEEE International Conference on Advanced Video and Signal based Surveillance*, Santa Fe, USA, 1-3 September 2008.

[61] J.L. Patino, E. Corvee, F. Bremond, and M. Thonnat. Data mining for activity extraction in video data. In *Extraction et Gestion des Connaissances EGC 2008*, Sophia Antipolis, France, 29th January - 1st February 2008.

[62] G. Pusiol, J.L. Patino, E. Corvee, F. Bremond, M.Thonnat, and S. Suresh. Optimizing trajectories clustering for activity recognition. In *1st International Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA 08)*, Marseille, France, 12th October - 18th October 2008.

[63] A. Sanchez, S. Moisan, and J.-P. Rigault. Dynamic first-class relations for knowledge-based systems. In *Workshop on Relationships and Associations in Object-Oriented Languages (RAOOL)*, Nashville, TN, USA, October 2008. OOPSLA 2008. http://www.mcs.vuw.ac.nz/raool/.

[64] M. Thonnat. Semantic activity recognition. In IOS Press, editor, *Proceedings of the18th European Conference on Artificial Intelligence(ECAI)*, pages 3–7, Greece, July 2008.

[65] P. Verney, M. Perrin, M. Thonnat, and J.F. Rainaud. An approach to seismic interpretation based on cognitive vision. the horizon identification as an application for the cognitive vision. In *Proceedings of the 70th EAGE Conference and Exhibition*, Rome, Italy, June 2008.

[66] P. Verney, J.F. Rainaud, M. Perrin, and M. Thonnat. A knowledge-based approach of seismic interpretation : Horizon and dip-fault detection by means of cognitive vision. In *SEG, International Exposition and Seventy-Eighth Annual Meeting*, Las Vegas, Nevada, USA, November 2008.

[67] N. Zouba, F. Bremond, and M. Thonnat. Monitoring activities of daily living (ADLs) of elderly based on 3d key human postures. In *In the 4th International Cognitive Vision Workshop (ICVW 2008)*, Santorini, Greece, 12-15 May 2008.

2009

[68] M Acher, P Collet, F Fleurey, P Lahire, Sabine Moisan, and Jean-Paul Rigault. Modeling Context and Dynamic Adaptations with Feature Models. In *Models@run.time Workshop*, Denver, CO, USA, October 2009.

[69] M Acher, P Lahire, Sabine Moisan, and Jean-Paul Rigault. Tackling high variability in video surveillance systems through a model transformation approach. In *ICSE'2009 - MISE Workshop*, Vancouver, Canada, May 2009.

[70] Slavomir Bak, Sundaram Suresh, Etienne Corvee, François Bremond, and Monique Thonnat. Fusion of motion segmentation and learning based tracker for visual surveillance. In *The International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, Lisbon, Portugal, 5 - 8 February 2009.

[71] Piotr Bilinski, François Bremond, and Mohamed-Becha. Kaâniche. Multiple object tracking with occlusions using hog descriptors and multi resolution images. In *Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP'09)*, Kingston University, London, UK, December 2009.

[72] Cyril Carincotte, François Bremond, Jean-Marc Odobez, Jose Luis Patino, Bertrand Ravera, and Xavier Desurmont. Multimedia Knowledge-based Content Analysis over Distributed Architecture. In *The Networked and Electronic Media, 2009 NEM Summit - "Towards Future Media"*, Saint- Malo, France, September 2009.

[73] Guillaume Charpiat. Learning Shape Metrics based on Deformations and Transport. In *Second Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment*, Kyoto, Japan, September 2009.

[74] Duc-Phu Chau, François Bremond, Etienne Corvee, and Monique Thonnat. Repairing people trajectories based on point clustering. In *The International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, Lisbon, Portugal, 5 - 8 February 2009.

[75] Duc-Phu. Chau, François Bremond, and Monique Thonnat. Online evaluation of tracking algorithm performance. In *The 3rd International Conference on Imaging for Crime Detection and Prevention*, London, United Kingdom, 3 December 2009.

[76] Etienne Corvee and François Bremond. Combining face detection and people tracking in video surveillance. In *Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention, ICDP 09*, London UK, December 2009.

[77] Carolina Garate, Piotr Bilinski, and François Bremond. Crowd event recognition using hog tracker. In *Proceedings of the Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, (invited paper), Winter-PETS 2009*, Snowbird, Utah, USA, December 2009.

[78] Valery Hourdin, Jean-Yves Tigli, Stephane Lavirotte, and Michel Riveill. Context-sensitive authorization for asynchronous communications. In *4th International Conference for Internet Technology and Secured Transactions (ICITST)*, London UK, November 2009.

[79] Mohammed-Bécha Kaâniche and François Bremond. Tracking hog descriptors for gesture recognition. In *The 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Genoa, Italy, 2 - 4 September 2009.

[80] Vincent Martin, François Bremond, J-M. Travere, V Moncada, and G Dunand. Thermal event recognition applied to tokamak protection during plasma operation. In *IEEE International Instrumentation and Measurement Technology Conference*, pages 1690–1694, May 2009.

[81] An-Tuan Nghiem, François Bremond, and Monique Thonnat. Controlling background subtraction algorithms for robust object detection. In *The 3rd International Conference on Imaging for Crime Detection and Prevention*, London, United Kingdom, 3 December 2009.

[82] Guido Pusiol, François Bremond, and Monique Thonnat. Trajectory-based primitive events for learning and recognizing activity. In *Second IEEE International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences (THEMIS2009)*, Kyoto Japan, October 2009.

[83] Rim Romdhane, François Bremond, and Monique Thonnat. Handling uncertainty for video event recognition. In *The 3rd International Conference on Imaging for Crime Detection and Prevention*, London, United Kingdom, 3 December 2009.

[84] Mohamed Siala, N Khlifa, François Bremond, and K Hamrouni. People detection in complex scene using a cascade of boosted classifiers based on haar-like-features. In *IEEE Intelligent Vehicle Symposium (IEEE IV 2009)*, June 2009.

[85] Nadia Zouba, François Bremond, and Monique Thonnat. Multisensor fusion for monitoring elderly activities at home. In *The 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Genoa, Italy, 2 - 4 September 2009.

[86] Nadia Zouba, François Bremond, Monique Thonnat, A Anfosso, E Pascual, P Mallea, V Mailland, and O Guerin. Assessing computer systems for the real time monitoring of elderly people living at home. In *The 19th IAGG World Congress of Gerontology and Geriatrics (IAGG)*, Paris, France, 5 - 7 July 2009.

[87] Marcos Zuniga, François Bremond, and Monique Thonnat. Incremental video event learning. In *Proceedings of the 7th IEEE International Conference on Computer Vision Systems*, Liege Belgium, October 2009.

2010

[88] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Person re-identification using haar-based and dcd-based signature. In *2nd Workshop on Activity Monitoring by Multi-Camera Surveillance Systems, AMMCSS 2010, in conjunction with 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance*. AVSS, 2010.

[89] Slawomir Bak, Etienne Corvee, Francois Bremond, and Monique Thonnat. Person re-identification using spatial covariance regions of human body parts. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Boston USA, 2010.

[90] I. Bechar and S. Moisan. On-line counting of pests in a greenhouse using computer vision. In *Visual observation and analysis of animal and insect behavior (VAIB) Workshop at ICPR*, Istanbul, August 2010.

[91] I. Bechar, S. Moisan, M. Thonnat, and F. Bremond. On-line video recognition and counting of harmful insects. In *International Conference for Pattern Recognition (ICPR 2010)*, Istanbul, Turkey, August 2010.

[92] Siqi Chen, Guillaume Charpiat, and Richard J. Radke. Converting level set gradients to shape gradients. In *11th European Conference on Computer Vision*, Hersonissos, Greece, September 2010.

[93] Etienne Corvee and Francois Bremond. Body parts detection for people tracking using trees of Histogram of Oriented Gradient descriptors. In *AVSS - 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Boston USA, 09 2010.

[94] Thi Lan Le, Alain Boucher, Monique Thonnat, and François Bremond. Surveillance video retrieval: what we have already done? In *Third International Conference on Communications and Electronics (ICCE 2010)*, Nha Trang, Viet Nam, 2010.

[95] Vincent Martin, Gwenaël Dunand, Victor Moncada, Jouve Michel, and Jean-Marcel Travere. New FPGA image-oriented acquisition and real-time processing applied to plasma facing component thermal monitoring. In Rev. Sci. Instrum., editor, *18th Topical Conference on High-Temperature Plasma Diagnostics*, volume 81 of *10E113*, pages 1–4, Wildwood, NJ, United States, October 2010. AIP.

[96] MB.Kaâniche and F. Brémond. Gesture recognition by learning local motion signatures. In *Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition, (CVPR'2010)*, June 2010.

[97] JL. Patino, F. Bremond, and M. Thonnat. Activity discovery from video employing soft computing relations. In *IEEE International Joint Conference on Neural Networks and Computational Intelligence (IJCNN 2010)*, Barcelona, Spain, July 2010.

[98] Jose Luis Patino, François Bremond, Murray Evans, Ali Shahrokni, and James Ferryman. Video Activity Extraction and Reporting with Incremental Unsupervised Learning. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Boston, United States, August 2010.

[99] Guido Pusiol, Francois Bremond, and Monique Thonnat. Trajectory Based Activity Discovery. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Boston USA, 2010.

[100] Rim Romdhane, Francois Bremond, and Monique Thonnat. A framework dealing with Uncertainty for Complex Event Recognition. In *AVSS 2010*, Boston USA, 08 2010.

[101] N. Zouba, F. Bremond, A. Anfonso, M. Thonnat, E. Pascual, and O. Guerin. Monitoring elderly activities at home. In *7th ISG 2010 Conference (International Society for Gerontechnology*, Vancouver, Canada, May 2010.

[102] Nadia Zouba, Francois Bremond, and Monique Thonnat. An Activity Monitoring System for Real Elderly at Home: Validation Study. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance AVSS10*, Boston États-Unis, 08 2010.

2011-2012

[103] M. Acher, P. Collet, P. Lahire, S. Moisan, and J-P. Rigault. Modeling Variability from Requirements to Runtime. In *16th Int. Conf. on Engineering of Complex Computer Systems (ICECCS'11)*, Las Vegas, 2011. IEEE Press.

[104] S. Bak, D.P. Chau, J. Badie, E. Corvee, F. Bremond, and M. Thonnat. Multi-target tracking by discriminative analysis on riemannian manifold. In *Proceedings of the 19th International Conference on Image Processing, IEEE Computer Society, ICIP 2012*, Orlando, USA, Sep 2012.

[105] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Multiple-shot human re-identification by mean riemannian covariance grid. In *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2011*, Klagenfurt, August 2011.

[106] Slawomir Bak, Krzysztof Kurowski, and Krystyna Napierala. Human Re-identification System On Highly Parallel GPU and CPU Architectures. In Andrzej Dziech and Andrzej CzyÅ¼ewski, editors, *Multimedia Communications, Services and Security*, volume 149, Krakow, Poland, June 2011. Springer Berlin Heidelberg.

[107] P. Bilinski, F. Bremond, and M. Thonnat. Evaluation of local descriptors for action recognition. In *8th International Conference on Computer Vision Systems, ICVS 2011*, Sophia Antipolis, France, September 2011.

[108] Guillaume Charpiat. Exhaustive Family of Energies Minimizable Exactly by a Graph Cut. In *Computer Vision and Pattern Recognition*, Colorado Springs, United States, June 2011.

[109] Duc Phu Chau, François Bremond, and Monique Thonnat. A multi-feature tracking algorithm enabling adaptation to context variations. In *The International Conference on Imaging for Crime Detection and Prevention (ICDP)*, London, Royaume-Uni, November 2011.

[110] Duc Phu Chau, François Bremond, Monique Thonnat, and Etienne Corvee. Robust Mobile Object Tracking Based on Multiple Feature Similarity and Trajectory Filtering. In *The International Conference on Computer Vision Theory and Applications*

(VISAPP), Algarve, Portugal, March 2011. This work is supported by The PACA region, The General Council of Alpes Maritimes province, France as well as The ViCoMo, Vanaheim, Video-Id, Cofriend and Support projects.

[111] Etienne Corvee, Slawomir Bak, and François Bremond. People detection and re-identification for multi surveillance cameras. In *VISAPP - International Conference on Computer Vision Theory and Applications -2012*, Rome, Italy, February 2012. Biometrics Groups at TELECOM SudParis, Multimedia Image processing Group of Eurecom and T3S (Thales Security Systems and Solutions S.A.S.).

[112] Etienne Corvee and Francois Bremond. Haar like and LBP based features for face, head and people detection in video sequences. In *International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, Sophia Antipolis, France, September 2011.

[113] C. Crispim-Junior, V. Joumier, Y.L. Hsu, P.C Chung M.C. Pai, A. Dechamps, P. Robert, and F. Bremond. Sensors data assessment for older person activity analysis. In *Proceedings of the ISGISARC 2012 Conference (International Society for Gerontechnology, the 8th world conference in cooperation with the ISARC, International Symposium of Automation and Robotics in Construction)*, Eindhoven, The Netherlands, Jun 2012.

[114] Ahmed Gamal Eldin, Xavier Descombes, Guillaume Charpiat, and Josiane Zerubia. A Fast Multiple Birth and Cut Algorithm using Belief Propagation. In *ICIP 2011 (International Conference on Image Processing)*, Bruxelles, Belgium, September 2011.

[115] Véronique Joumier, Rim Romdhane, Francois Bremond, Monique Thonnat, Emmanuel Mulin, P. H. Robert, A. Derreumaux, Julie Piano, and J. Lee. Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients. In *International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, Sophia Antipolis, France, September 2011.

[116] Vincent Martin, Victor Moncada, Jean-Marcel Travere, Thierry Loarer, François Bremond, Guillaume Charpiat, and Monique Thonnat. A Cognitive Vision System for Nuclear Fusion Device Monitoring. In James L. Crowley, Bruce Draper, and Monique Thonnat, editors, *8th International Conference, ICVS 2011*, volume 6962 of *Lecture Notes in Computer Science*, pages 163–172, Sophia Antipolis, France, September 2011. Springer.

[117] S. Moisan, J.-P. Rigault, M. Acher, P. Collet, and 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.

[118] Sabine Moisan. Intelligent monitoring of software components. In *RAISE 2012 Workshop at ICSE*, Zurich, CH, June 2012. to appear.

[119] Sabine Moisan, jean Paul Rigault, and Mathieu Acher. A feature-based approach to system deployment and adaptation. In *MISE 2012 Workshop at ICSE*, Zurich, CH, June 2012. to appear.

[120] Luis Patino, Hamid Benhadda, Nedra Nefzi, Bernard Boulay, François Bremond, and Monique Thonnat. Abnormal behavior detection in video protection systems.

In *International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, Sophia Antipolis, France, September 2011.

[121] Luis Patino, François Bremond, and Monique Thonnat. Incremental learning on trajectory clustering. In Dr. Paolo Remagnino, editor, *Intelligent Paradigms in Safety and Security*. Springer-Verlag, 2011.

[122] Luis Patino, Murray Evans, James Ferryman, François Bremond, and Monique Thonnat. Unsupervised Activity Extraction on Long-Term Video Recordings employing Soft Computing Relations. In *8th International Conference on Computer Vision Systems, ICVS 2011*, Sophia Antipolis, France, September 2011.

[123] G. Pusiol, F. Bremond, and M. Thonnat. Unsupervised discovery, modeling and analysis of long term activities. In *8th International Conference on Computer Vision Systems, ICVS 2011*, Sophia Antipolis, France, September 2011.

[124] Annie Ressouche, Jean-Yves Tigli, and Oscar Carrillo. Toward Validated Composition in Component-Based Adaptive Middleware. In Ethan Jackson Sven Apel, editor, *SC 2011*, volume 6708 of *LNCS*, pages 165–180, Zurich, Suisse, July 2011. Mario Sudhot, Springer.

[125] Leonardo Rocha, Sagar Sen, Sabine Moisan, and Jean-Paul Rigault. Towards Lightweight Dynamic Adaptation. A Framework and its Evaluation. In *12th Argentine Symposium on Software Engineering (ASSE 2011), JAII0 2011*, Córdoba, Argentina, 2011.

[126] Leonardo Manuel Rocha, Sabine Moisan, Jean-Paul Rigault, and Sen Sagar. Girgit: A Dynamically Adaptive Vision System for Scene Understanding. In *ICVS*, Sophia Antipolis, France, September 2011.

[127] Leonardo Manuel Rocha, Sen Sagar, Sabine Moisan, and Jean-Paul Rigault. Girgit: A Framework for Component-based Dynamic Adaptation in Python. In *EuroSciPy 2011*, Paris, France, August 2011.

[128] R. Romdhane, F. Bremond, and M. Thonnat. Probabilistic recognition of complex event. In *8th International Conference on Computer Vision Systems, ICVS 2011*, Sophia Antipolis, France, September 2011.

[129] Malik Souded, Laurent Giulieri, and Francois Bremond. An Object Tracking in Particle Filtering and Data Association Framework, Using SIFT Features. In *International Conference on Imaging for Crime Detection and Prevention (ICDP)*, London, United Kingdom, November 2011.

[130] Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc Phu Chau, Etienne Corvee, and Francois Bremond. Group interaction and group tracking for video-surveillance in underground railway stations. In *International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, Sophia Antipolis, France, September 2011.

10.4 HDR and PhD Theses Defended in Pulsar and Stars Teams

[131] Bernard Boulay. *Human Posture Recognition for behavior Understanding*. PhD thesis, Nice-Sophia Antipolis University, January 2007.

[132] Francois Bremond. *Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition.* Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, juillet 2007.

[133] D.P. Chau. *Dynamic and Robust Object Tracking for Activity Recognition*. PhD thesis, Université de Nice Sophia Antipolis, Mar 2012.

[134] Benoit Georis. *Program Supervision Techniques for Easy Configuration of Video Understanding Systems*. PhD thesis, Louvain Catholic University, January 2006.

[135] Mohamed-Becha Kâaniche. *Gesture Recognition from video sequences*. PhD thesis, Université de Nice - Sophia Antipolis - ED - STIC, October 2009.

[136] Lan Le Thi. *Indexation et Recherche de Video pour la Videosurveillance*. PhD thesis, Université de Nice-Sophia Antipolis, Feb 2009.

[137] Anh-Tuan Nghiem. *Algorithmes adaptatifs d'estimation du fond pour la détection des objets mobiles dans les séquences vidéos*. PhD thesis, Université de Nice Sophia-Antipolis, 06 2010.

[138] Nadia Zouba. *Multisensor Fusion for Monitoring Elderly Activities at Home*. PhD thesis, Université de Nice Sophia Antipolis, January 2010.

[139] Marcos Zuniga. *Incremental Learning of Events in Video using Reliable Information*. PhD thesis, Université de Nice-Sophia Antipolis, Nov. 2008.

11 References

[140] Clem home page. http://www.inria.fr/sophia/pulsar/projects/Clem.

[141] Kermeta home page. http://www.kermeta.org/.

[142] A. Agarwal and B. Triggs. A local basis representation for estimating human pose from cluttered images. In *Asian Conference on Computer Vision*, January 2006.

[143] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43:16:1–16:43, April 2011.

[144] Umut Akdemir, Pavan Turaga, and Rama Chellappa. An ontology based approach for activity recognition from video. In *Proceeding of the 16th ACM international conference on Multimedia*, MM '08, pages 709–712, New York, NY, USA, 2008. ACM.

[145] R. Allen and D. Garlan. A formal basis for architectural connection. *ACM Transactions on Software Engineering and Methodology*, 6(3):213–249, 1997.

[146] E. Andrade, S. Blunsden, and R. Fisher. Performance analysis of event detection models in crowded scenes. In *Workshop on Towards Robust Visual Surveillance Techniques and Systems at Visual Information Engineering 2006*, pages 427–432, Bangalore, India, September 2006.

[147] E. L. Andrade, S. Blunsden, and R. B. Fisher. Hidden markov models for optical flow analysis in crowds. In *International Conference on Pattern Recognition*, Hong Kong, August 2006.

[148] A. Avanzi, F. Bremond, C. Tornieri, and M. Thonnat. Design and assessment of an intelligent activity monitoring platform. *EURASIP Journal on Applied Signal Processing, special issue in "Advances in Intelligent Vision Systems: Methods and Applications"*, 2005(14):2359–2374, August 2005.

[149] Lamberto Ballan, Marco Bertini, Alberto Bimbo, Lorenzo Seidenari, and Giuseppe Serra. Event detection and recognition for semantic annotation of video. *Multimedia Tools Appl.*, 51:279–302, January 2011.

[150] Don S. Batory. Feature models, grammars, and propogsitional formulas. In *SPLC*, volume 3714 of *LNCS*, pages 7–20, 2005.

[151] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15:1373–1396, 2003.

[152] D. Benavides, P. Trinidad, and A. Ruiz-Cortés. Automated reasoning on feature models. In *LNCS, ADVANCED INFORMATION SYSTEMS ENGINEERING: 17TH INTERNATIONAL CONFERENCE, CAISE 2005*. Springer, 2005.

[153] J. Berclaz, F. Fleuret, and P. Fua. Robust people tracking with global trajectory optimization. In *Conference on Computer Vision and Pattern Recognition*, 2006.

[154] G. Berry. The foundations of esterel. In G. Plotkin, C. Stearling, and M. Tofte, editors, *Proof, Language, and Interaction, Essays in Honor of Robin Milner*. MIT Press, 2000.

[155] Sami Beydeda and Matthias Book, editors. *Model-Driven Software Development*. Springer, 2005.

[156] M. Borg, D. J. Thirde, J. M. Ferryman, K. D. Baker, J. Aguilera, and M. Kampel. Evaluation of object tracking for aircraft activity surveillance. In *The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2005)*, Beijing, China, October 2005.

[157] B. Boulay, F. Bremond, and M. Thonnat. Human posture recognition in video sequence. In *Proceedings Joint IEEE International Workshop on VS-PETS, Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 23–29, OCT 11-12 2003.

[158] Bernard Boulay, Francois Bremond, and Monique Thonnat. Posture recognition with a 3d human model. In *Proceedings of IEE International Symposium on Imaging for Crime Detection and Prevention*, London, UK, 2005. Institution of Electrical Engineers.

[159] H. Buxton. Learning and understanding dynamic scene activity. In *ECCV Generative Model Based Vision Workshop*, Copenhagen, 2002.

[160] A. Cavallaro. Event detection in underground stations using multiple heterogeneous surveillance cameras. In *International Symposium on Visual Computing, Intelligent Vehicles and Autonomous Navigation*, Lake Tahoe, Nevada, December 2005.

[161] A. Cavallaro, O. Steiger, and T. Ebrahimi. Tracking video objects in cluttered background. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):575–584, April 2005.

[162] R. Chellappa, A. R. Chowdhury, and S. Zhou. *Recognition of Humans and Their Activities Using Video*. Morgan Claypool, 2005.

[163] N. Chleq, F. Bremond, and M. Thonnat. *Advanced Video-based Surveillance Systems*, chapter Image Understanding for Prevention of Vandalism in Metro Stations, pages 108–118. Kluwer A.P. , Hangham, MA, USA, C.S. Regazzoni, G. Fabri, and G. Vernazza edition, November 1998.

[164] A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV 2: an OpenSource Tool for Symbolic Model Checking. In Ed Brinksma and Kim Guldstrand Larsen, editors, *Proceeeding CAV*, number 2404 in LNCS, pages 359–364, Copenhagen, Danmark, July 2002. Springer-Verlag.

[165] A. G. Cohn, D. C. Hogg, B. Bennett, V. Devin, A. Galata, D. R. Magee, C. Needham, and P. Santos. Cognitive vision: Integrating symbolic qualitative representations with computer vision. *Cognitive Vision Systems: Sampling the Spectrum of Approaches, Lecture Notes in Computer Science*, pages 221–246, 2006.

[166] Toby Collins. Analysing Video Sequences using the Spatio-temporal Volume. Technical Report November, University of Edinburgh, 2004.

[167] C. Coué, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessière. Bayesian occupancy filtering for multitarget tracking: an automotive application. *Journal of Robotics Research*, 25(1):19–30, January 2006.

[168] P. Cousot and R. Cousot. On Abstraction in Software Verification. In Ed Brinksma and Kim Guldstrand Larsen, editors, *Proceeeding CAV*, number 2404 in LNCS, pages 37,56, Copenhagen, Danmark, July 2002. Springer-Verlag.

[169] D. Cremers. Nonlinear dynamical shape priors for level set segmentation. *Journal of Scientific Computing*, 35(2-3):132–143, June 2008.

[170] J. L. Crowley. *Cognitive Vision Systems*, chapter Things that See: Context-Aware Multi-modal Interaction. Springer Verlag, March 2006.

[171] J. L. Crowley. Situation models for observing human activity. *ACM Queue Magazine*, May 2006.

[172] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati. The sakbot system for moving object detection and tracking. *Video-based Surveillance Systems: Computer Vision and Distributed Processing (Part II - Detection and Tracking)*, pages 145–158, 2001.

[173] R. Cucchiara, A. Prati, and R. Vezzani. Domotics for disability: smart surveillance and smart video server. In *8th Conference of the Italian Association of Artificial Intelligence - Workshop on Ambient Intelligence*, pages 46–57, Pisa, Italy, September 2003.

[174] R. Cucchiara, A. Prati, R. Vezzani, L. Benini, E. Farella, and P. Zappi. An integrated multi-modal sensor network for video surveillance. *Journal of Ubiquitous Computing and Intelligence (JUCI)*, 2006.

[175] F. Cupillard, F. Brémond, and M. Thonnat. Group behavior recognition with multiple cameras. In *IEEE Proc. of the Workshop on Applications of Computer Vision - ACV2002*, Orlando USA, Dec 2002.

[176] F. Cupillard, F. Brémond, and M. Thonnat. Tracking groups of people for video surveillance. In P. Remagnino, GA. Jones, N. Paragios, and C. Regazzoni, editors, *Video-Based Surveillance Systems*, pages 89–100. Kluwer Academic, 2002.

[177] K. Czarnecki and U. W. Eisenecker. *Generative Programming: Methods, Tools, and Applications*. Addison Wesley, 2000.

[178] Krzysztof Czarnecki, Simon Helsen, and Ulrich Eisenecker. Staged Configuration through Specialization and Multilevel Configuration of Feature Models. *Software Process: Improvement and Practice*, 10(2):143–169, 2005.

[179] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *European Conference on Computer Vision*, 2006.

[180] W. Damm and D. Harel. Lscs: Breathing life into message sequence charts. *Formal Methods in System Design*, 19:45–80, 2001.

[181] R. David, E. Mulin, P. Mallea, and P.H. Robert. Measurement of neuropsychiatric symptoms in clinical trials targeting alzheimer's disease and related disorders. *Pharmaceuticals*, 3:2387–2397, 2010.

[182] A. C. Davies and S. A. Velastin. Progress in computational intelligence to support cctv surveillance systems. *International Scientific Journal of Computing*, 4(3):76–84, 2005.

[183] X. Desurmont, J. B. Hayet, J. F. Delaigle, J. Piater, and B. Macq. Trictrac video dataset: Public hdtv synthetic soccer video sequences with ground truth. In *Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE)*, pages 92–100, 2006.

[184] X. Desurmont, I. Ponte, J. Meessen, and J. F. Delaigle. Nonintrusive viewpoint tracking for 3d for perception in smart video conference. In *Three-Dimensional Image Capture and Applications VI*, San Jose, CA USA, January 2006.

[185] P. Domingos and M. Richardson. Markov logic networks. *Machine Learning*, 62(1-2):107–136, February 2006.

[186] W. Dong and A. Pentland. Multi-sensor data fusion using the influence model. In *Body Sensor Networks Workshop*, Boston, MA, April 2006.

[187] C. Dousson, P. Gaborit, and M. Ghallab. Situation recognition: Representation and algorithms. In *The 13rd IJCAI*, pages 166–172, August 1993.

[188] W. Du, J. B. Hayet, J. Piater, and J. Verly. Collaborative multi-camera tracking of athletes in team sports. In *Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE)*, pages 2–13, 2006.

[189] W. Du and J. Piater. Multi-view tracking using sequential belief propagation. In *Asian Conference on Computer Vision*, pages 684–693, Hyderabad, India, 2006. Springer-Verlag.

[190] Rémi Emonet, Jagan Varadarajan, and Jean-Marc Odobez. Extracting and Locating Temporal Motifs in Video Scenes Using a Hierarchical Non Parametric Bayesian Model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, June 2011.

[191] A. Enficiaud, B. Lienard, N. Allezard, R. Sebbe, S. Beucher, X. Desurmont, P. Sayd, and J. F. Delaigle. Clovis - a generic framework for general purpose visual surveillance applications. In *IEEE International Workshop on Visual Surveillance (VS)*, 2006.

[192] M. Enzweiler and D.M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12), December 2009.

[193] A. Ess, B. Leibe, K. Schindler, and L.Van Gool. Robust multiperson tracking from a mobile platform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10), October 2009.

[194] J. M. Ferryman, S. Maybank, and A. Worrall. Visual surveillance for moving vehicles. *International Journal of Computer Vision*, 37(2):187–197, June 2000.

[195] F. Fleurey, V. Delhen, N. Bencomo, B. Morin, and J.-M. Jézéquel. Modeling and validating dynamic adaptation. In *Proceedings of the 3rd International Workshop on Models@Runtime, at MoDELS'08*, Toulouse, France, oct 2008.

[196] F. Fleurey and A. Solberg. A domain specific modeling language supporting specification, simulation and execution of dynamic adaptive systems. In *MoDELS*, pages 606–621, Denver, CO, USA, 2009.

[197] J. Floch, S. Hallsteinsen, E. Stav, F. Eliassen, K. Lund, and E. Gjorven. Using architecture models for runtime adaptability. *IEEE Softw.*, 23(2):62–70, 2006.

[198] G. L. Foresti and C. Micheloni. A robust feature tracker for active surveillance of outdoors scenes. *Electronic Letters on Computer Vision and Image Analysis*, 1(1):21–34, 2003.

[199] G. L. Foresti, C. Piciarelli, and L. Snidaro. Trajectory clustering and its applications for video surveillance. In *Advanced Video and Signal-Based Surveillance (AVSS)*, Como, Italy, September 2005.

[200] G. L. Foresti and C. S. Regazzoni. Multisensor data fusion for driving autonomous vehicles in risky environments. *IEEE Transactions on Vehicular Technology*, 51(5):1165 –1185, September 2002.

[201] A. Galata, N. Johnson, and D. Hogg. Learning variable length markov models of behaviour. *Computer Vision and Image Understanding*, 81(3):398–413, 2001.

[202] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison Wesley, 1995.

[203] D. M. Gavrila and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. *International Journal of Computer Vision (est: February-March 2007 issue)*, 2007.

[204] C. Gomila and F. Meyer. Tracking objects by graph matching of image partition sequences. In *3rd IAPR-TC15 Workshop on Graph-Based Representations in Pattern Recognition*, pages 1–11, Ischia, Italy, May 2001.

[205] N. Gourier, J. Maisonnasse, D. Hall, and J. L. Crowley. Head pose estimation on low resolution images. In *CLEAR 2006*, South Hampton, UK, April 2006.

[206] D. Greenhill, P. Remagnino, and G. A. Jones. *Video Based Surveillance Systems - Computer Vision and Distributed Processing*, chapter VIGILANT: Content-Querying of Video Surveillance Streams, pages 193–204. Kluwer Academic Publishers, 2002.

[207] A. Gupta, A. Kembhavi, and L.S. Davis. Observing human-object interactions: Using spatial and functional compatibility for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10), October 2009.

[208] D. Hall. Automatic parameter regulation for a tracking system with an auto-critical function. In *Proceedings of the IEEE International Workshop on Computer Architecture for Machine Perception (CAMP05)*, pages 39–45, Palermo, Italy, July 2005.

[209] D. Hall, R. Emonet, and J. L. Crowley. An automatic approach for parameter selection in self-adaptive tracking. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, Setubal, Portugal, February 2006.

[210] D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R. B. Fisher, J. Santos-Victor, and J. L. Crowley. Comparison of target detection algorithms using adaptive background models. In *2nd Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 113–120, Beijing, October 2005.

[211] D. Hall, F. Pelisson, O. Riff, and J. L. Crowley. Brand identification using gaussian derivative histograms. *Machine Vision and Applications, in Machine Vision and Applications*, 16(1):41–46, 2004.

[212] Svein Hallsteinsen, Erlend Stav, Arnor Solberg, and Jacqueline Floch. Using product line techniques to build adaptive systems. In *SPLC06*, 2007.

[213] D. C. Hogg, A. G. Cohn, V. Devin, D. Magee, C. Needham, and P. Santos. Learning about objects and activities. In *24th Leeds Annual Statistical Research (LASR) Workshop*, Leeds, UK, 2005.

[214] S. Hongeng, F. Brémond, and R. Nevatia. Bayesian framework for video surveillance application. In *Proceedings of the International Conference on Pattern Recognition (ICPR'00)*, pages 1164–1170, Barcelona, Spain, September 2000.

[215] C. L. Huang, E. L. Chen, and P. C. Chung. Fall detection using modular neural networks and back-projected optical flow. In *The 12th International Conference on Neural Information Processing (ICONIP 2005)*, 2005.

[216] C. L. Huang, E. L. Chen, and P. C. Chung. Using modular neural networks for falling detection. In *The 10th Conference on Artificial Intelligence and Application (TAAI2005)*, 2005.

[217] Y. Ivanov and A. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(8):852–872, 2000.

[218] Ries J.D., Echternach J.L., Nof L., and Gagnon-Blodgett M. Test-retest reliability and minimal detectable change scores for the timed up and go test, the six-minute walk test, and gait speed in people with alzheimer disease. *Physical Therapy*, 89(6), 2009.

[219] Frédéric Jouault, Freddy Allilaire, Jean Bézivin, and Ivan Kurtev. Atl: A model transformation tool. *Sci. Comput. Program.*, 72(1-2):31–39, 2008.

[220] E. M. Clarke Jr, O.Grumberg, and D.Peled. *Model Checking*. MIT Press, 2000.

[221] A. Kale, N. Cuntoor, B. Yegnanarayana, A. N. Rajagoplan, and R. Chellappa. *Using Appearance Matching in Optical and Digital Techniques of Information Security*, chapter Gait-based Human Identification. Springer, 2005.

[222] K. Kang, S. Cohen, J. Hess, W. Novak, and S. Peterson. Feature-Oriented Domain Analysis (FODA) Feasibility Study. Technical Report CMU/SEI-90-TR-21, November 1990.

[223] C. Kästner, S. Apel, S. Trujillo, M. Kuhlemann, and D.S. Batory. Guaranteeing syntactic correctness for all product line variants: A language-independent approach. In *TOOLS (47)*, pages 175–194, 2009.

[224] Anneke Kleppe, Jos Warmer, and Wim Bast. *MDA Explained: The Model Driven Architecture–Practice and Promise*. Addison-Wesley , April 2003.

[225] C. W. Krueger. New methods in software product line development. In *10th Int. Software Product Line Conf.*, pages 95–102, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[226] M. Kwiatkowska, G. Norman, J. Sproston, and F. Wang. Symbolic model checking for probabilistic timed automata. *Information and Computation*, 205(7):1027–1077, 2007.

[227] I. Laptev, B. Caputo, and T. Lindeberg. Local velocity-adapted motion events for spatio-temporal recognition. *Computer Vision and Image Understanding*, 2007.

[228] I. Laptev and T. Lindeberg. Velocity-adaptation of spatio-temporal receptive fields for direct recognition of activities: An experimental study. *Image and Vision Computing*, 22:105–116, 2004.

[229] D. Larlus and F. Jurie. Latent mixture vocabularies for object categorization. In *British Machine Vision Conference*, 2006.

[230] G. Lavee, E. Rivlin, and M. Rudzsky. Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. *IEEE Transactions on Systems, Man, and Cybernetics-Part C:Applications and Reviews*, 39(5):489–504, September 2009.

[231] J. Lezama, K. Alahari, Josef Sivic, and I. Laptev. Track to the Future: Spatio-temporal Video Segmentation with Long-range Motion Cues. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011.

[232] S. N. Lim, A. Mittal, L. S. Davis, and N. Paragios. Fast illumination-invariant background subtraction using two views: Error analysis, sensor placement and applications. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, June 2005.

[233] R. Lublinerman, C. Szegedy, and S. Tripakis. Modular code generation from synchronous block diagrams: Modularity vs. reusability. In *ACM Symposium on Principles of Programming Languages(POPL'09)*, Savannah, GA, USA, 2009.

[234] F. Lv, X. Song, B. Wu, V. K. Singh, and R. Nevatia. Left luggage detection using bayesian inference. In *PETS*, 2006.

[235] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology based cognitive vision. *Machine Vision and Applications (MVA)*, 16(1):33–40, December 2004.

[236] Nicolas Maillot. *Ontology Based Object Learning and Recogntion*. PhD thesis, Nice-Sophia Antipolis University, December 2005.

[237] W. Mark and D. M. Gavrila. Real-time dense stereo for intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):38–50, March 2006.

[238] T. Mathes and J. Piater. Robust non-rigid object tracking using point distribution manifolds. In *28th Annual Symposium of the German Association for Pattern Recognition (DAGM)*, 2006.

[239] G. Medioni, I. Cohen, F. Brémond, S. Hongeng, and G. Nevatia. Activity Analysis in Video. *Pattern Analysis and Machine Intelligence PAMI*, 23(8):873–889, 2001.

[240] C. Micheloni and G. L. Foresti. Fast good feature selection for wide area monitoring. In *IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 271–276, Miami Beach, Florida, July 2003.

[241] T. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90–126, 2006.

[242] S. Moisan. Knowledge representation for program reuse. In *Proc. European Conference of Artificial intelligence (ECAI 2002)*, Lyon, France, July 2002.

[243] S. Moisan, A. Ressouche, and J-P. Rigault. Blocks, a component framework with checking facilities for knowledge-based systems. *Informatica, Special Issue on Component Based Software Development*, 25(4):501–507, November 2001.

[244] S. Moisan, A. Ressouche, and J.-P. Rigault. Towards Formalizing Behavorial Substitutability in Component Frameworks. In *2nd International Conference on Software Engineering and Formal Methods*, pages 122–131, Beijing, China, 2004. IEEE Computer Society Press.

[245] Sabine Moisan. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*. Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, avril 1998.

[246] S. Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1863–1868, 2006.

[247] B. Neumann and T. Weiss. Navigating through logic-based scene models for high-level scene interpretations. In Springer, editor, *3rd International Conference on Computer Vision Systems - ICVS 2003*, pages 212–222, 2003.

[248] R. Nevatia, J. Hobbs, and B. Bolles. An ontology for video event representation. In *IEEE Workshop on Event Detection and Recognition*, June 2004.

[249] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. Springer, December 2005.

[250] T. Ogata, W. Christmas, J. Kittler, and S. Ishikawa. Improving human activity detection by combining multi-dimensional motion descriptors with boosting. In *Proceedings of the 18th International Conference on Pattern Recognition*, volume 1, pages 295–298. IEEE Computer Society, August 2006.

[251] V. Parameswaren and R. Chellappa. Human action-recognition using mutual invariants. *Computer Vision and Image Understanding*, 98:295–325, September 2005.

[252] G. Perrouin, F. Chauvel, J. DeAntoni, and J.-M. Jézéquel. Modeling the variability space of self-adaptive applications. In Steffen Thiel and Klaus Pohl, editors, *2nd Dynamic Software Product Lines Workshop (SPLC 2008, Volume 2)*, pages 15–22, Limerick, Ireland, September 2008. IEEE Computer Society.

[253] P. Pettersson and K. Larsen. Uppaal2k. *Bulletin of the European Association for Theoretical Computer Science*, 70:40–44, 2000.

[254] C. Pinhanez and A. Bobick. Interval scripts: a programming paradigm for interactive environments and agents. *Pervasive and Ubiquitous Computing*, 7(1):1–21, 2003.

[255] Klaus Pohl, Günter Böckle, and Frank J. van der Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[256] Dumitru Potop-Butucaru, Stephen A. Edwards, and Gerard Berry. *Compiling Esterel*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[257] P. Remagnino, A. I. Shihab, and G. A. Jones. Distributed intelligence for multi-camera visual surveillance. *Pattern Recognition, Special Issue on Agent-based Computer Vision*, 37(4):675–689, April 2004.

[258] K. Sage and H. Buxton. Joint spatial and temporal structure learning for task based control. In *International Conference on Pattern Recognition*, 2004.

[259] E. Salvador, A. Cavallaro, and T. Ebrahimi. Cast shadow segmentation using invariant colour features. *Computer Vision and Image Understanding*, 95(2):238–259, August 2004.

[260] Peter Sand and Seth Teller. Particle Video: Long-Range Motion Estimation Using Point Trajectories. *International Journal of Computer Vision*, 80(1):72–91, May 2008.

[261] Klaus Schneider, Jens Brand, and Eric Vecchié. Modular compilation of synchronous programs. In *From Model-Driven Design to Resource Management for Distributed Embedded Systems*, volume 225 of *IFIP International Federation for Information Processing*, pages 75–84. Springer Boston, January 2006.

[262] Pierre-Yves Schobbens, Patrick Heymans, Jean-Christophe Trigaux, and Yves Bontemps. Generic semantics of feature diagrams. *Computer Networks*, 51(2), 2007.

[263] Y. Sheikh and M. Shah. Bayesian modelling of dynamic scenes for object detection. *IEEE Transactions on PAMI*, October 2005.

[264] Y. Shi, Y. Huang, D. Minnen, A. Bobick, and I. Essa. Propagation networks for recognition of partially ordered sequential action. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR 2004)*, Washington DC, June 2004.

[265] J. Sproston. *Validation of Stochastic System*, volume 2925 of *Lecture Notes in Computer Science*, chapter Model Checking for Probabilistic Timed Systems, pages 189–229. Springer Berlin / Heidelberg, 2004.

[266] Mikael Svahnberg, Jilles van Gurp, and Jan Bosch. A taxonomy of variability realization techniques: Research articles. *Software Practice and Experience*, 35(8):705–754, 2005.

[267] A. M. Taj, E. Maggio, and A. Cavallaro. Multi-feature graph-based object tracking. In *Classification of Events, Activities and Relationships (CLEAR) Workshop*, Southampton, UK, April 2006. Springer.

[268] Jean-Yves Tigli, Stephane Lavirotte, Gaetan Rey, Valery Hourdin, Daniel Cheung, E Callegari, and Michel Riveill. Wcomp middleware for ubiquitous computing: Aspects and composite event-based web services. *Annals of Telecommunications*, 64(3-4), 2009. ISSN 0003-4347 (Print) ISSN 1958-9395 (Online).

[269] Jean-Yves Tigli, Stephane Lavirotte, Gaetan Rey, Valery Hourdin, and Michel Riveill. Lightweight service oriented architecture for pervasive computing. *IJCSI International Journal of Computer Science Issues*, 4(1), 2009. ISSN (Online): 1694-0784 ISSN (Print): 1694-0814.

[270] T. T. Tun, Q. Boucher, A. Classen, A. Hubaux, and P. Heymans. Relating requirements and feature configurations: A systematic approach. In *SPLC'09*, pages 201–210. IEEE, 2009.

[271] R. Urtasun, D. Fleet, and P. Fua. 3d people tracking with gaussian process dynamical models. In *Conference on Computer Vision and Pattern Recognition*, June 2006.

[272] M. Vaillant, M. Miller, A. Trouvé, and L. Younes. Statistics on diffeomorphisms via tangent space representation. *Neuroimage*, 2005.

[273] S. A. Velastin, B. A. Boghossian, and M. A. Vicencio-Silva. A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. *Transportation Research Part C: Emerging Technologies*, 14(2):96–113, 2006.

[274] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[275] S. Von-Wum, L. Chen-Yu, Y. Jaw-Jium, and C. Ching-Chi. Using sharable ontology to retrieve historical images. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries*, pages 197–198, 2002.

[276] V. T. Vu, F. Brémond, and M. Thonnat. Temporal constraints for video interpretation. In *Proc of the 15th European Conference on Artificial Intelligence*, Lyon, France, 2002.

[277] V-T. Vu, F. Brémond, and M. Thonnat. Automatic video interpretation: A novel algorithm for temporal scenario recognition. In *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico, 2003.

[278] Van-Thinh Vu. *Temporal Scenario for Automatic Video Interpretation*. PhD thesis, Nice-Sophia Antipolis University, 2004.

[279] Y. Wang and G.Mori. Human action recognition by semilatent topic models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10), October 2009.

[280] J. White, D. C. Schmidt, E. Wuchner, and A. Nechypurenko. Automating product-line variant selection for mobile devices. In *SPLC*, pages 129–140, 2007.

[281] T. Xiang and S. Gong. Video behaviour profiling and abnormality detection without manual labelling. In *IEEE International Conference on Computer Vision*, pages 1238–1245, Beijing, October 2005.

[282] T. Xiang and S. Gong. Beyond tracking: Modelling activity and understanding behaviour. *International Journal of Computer Vision*, 67(1):21–51, 2006.

[283] T. Xiang and S. Gong. Incremental visual behaviour modelling. In *IEEE Visual Surveillance Workshop*, pages 65–72, Graz, May 2006.

[284] T. Xiang and S. Gong. Optimal dynamic graphs for video content analysis. In *British Machine Vision Conference*, volume 1, pages 177–186, Edinburgh, September 2006.

[285] Q. Yu and G. Medioni. Multiple-target tracking by spatiotemporal monte carlo markov chain data association. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12), December 2009.

[286] S. Zaidenberg, O. Brdiczka, P. Reignier, and J. L. Crowley. Learning context models for the recognition of scenarios. In *In 3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI) 2006*, Athens, Greece, June 2006.

[287] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM MM*, Santa Barbara, CA, USA, 2006.

[288] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan. Semi-supervised adapted hmms for unusual event detection. In *EEE CVPR*, 2005.

[289] J. Zhang and B.H.C. Cheng. Specifying adaptation semantics. In *WADS '05: Proceedings of the 2005 workshop on Architecting dependable systems*, pages 1–7, 2005.

[290] J. Zhang and S. Gong. Beyond static detectors: A bayesian approach to fusing long-term motion with appearance for robust people detection in highly cluttered scenes. In *IEEE Visual Surveillance Workshop*, pages 121–128, Graz, May 2006.

[291] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. *Computer Vision and Pattern Recognition*, 2:406–413, 2004.

[292] Song Chun Zhu and David Mumford. A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, 2(4):259–362, 2006.

[293] F. Ziliani, S. Velastin, F. Porikli, L. Marcenaro, T. Kelliher, A. Cavallaro, and P. Bruneaut. Performance evaluation of event detection solutions: the creds experience. In *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS 2005)*, Como, Italy, September 2005.

A The SUP platform (Scene Understanding Platform)

SUP is a generic software platform for building a large variety of real-time activity recognition systems dedicated to a particular application. This platform enables Stars members designing, experimenting and validating their new algorithms for activity recognition.

SUP is written in python, C and C++ and relies on DTK, a generic platform developed by the DREAM service in the INRIA Research Center Sophia Antipolis - Méditerranée. SUP is oriented to help developers building activity monitoring systems and describing their own scenarios dedicated to specific applications. SUP has been built based on VSIP, the previous Pulsar Platform.

There is an increasing demand in the safety and security domains for video analytics systems. These systems are complicated to build from scratch and difficult to configure due to the large variety of scenes to process and to the numbers of processing steps involved in video sequence analysis. Most of these video analytics systems share common functionalities (as motion detection or object tracking). We thus propose a platform, named SUP (Scene Understanding Platform), to help developers building video analytics systems.

This platform is composed of reusable components which can be combined for different applications. SUP mainly contains: (1) image processing algorithms in charge of mobile object detection, classification and frame to frame tracking; (2) multi-camera algorithms for the spatial and temporal analysis (4D) of the detected mobile objects; (3) high level scenario recognition algorithms.

The SUP platform has the backbone of the Pulsar team experiment to implement the newest algorithms in perception, understanding and learning.

SUP is made as a framework allowing several video analytics workflows to be implemented. Currently, the workflow is static for a given application but our goal is to make it dynamic. A given workflow is the composition of several plugins, each of them implementing an algorithmic step in the video processing (i.e. the segmentation of images, the classification of objects, etc.).

The design of SUP allows to execute at runtime the selected plugins. Currently 15 plugins are available:

- 6 plugins are wrappers on industrial implementations of algorithms (made available by Keeneo). They allow a quick deployment of a video processing chain encompassing image acquisition, segmentation, short-term and long-term tracking. These algorithms are robust and efficient algorithms, but with the drawback that some algorithms can lack accuracy.
- 9 are implementations by the team members which cover the following fields:
  - one segmentation removing the shadows;
  - two classifiers, one being based on postures and one on people detection;
  - four frame-to-frame trackers, using as algorithm: (i) a simple tracking by overlapping, (ii) neural networks, (iii) tracking of feature points, (iv) tracking specialized for the tracking of persons in a crowd.
  - two scenario recognizers, one generic allowing expression of probabilities on the recognized events, and the other one focusing on the recognition of events based on postures.

From a software engineering point-of-view, the goal is to obtain a platform being dynamically reconfigurable for the generated scene understanding systems to be autonomous and adaptable for handling changing environment.

SUP relies on DTK, a generic platform developed by the DREAM service in the INRIA Research Center Sophia Antipolis - Méditerranée. The purpose of DTK is to provide a software infrastructure allowing the generation of a new system by the composition of plugins, each plugin being an algorithmic step of the whole processing chain. SUP is oriented to help developers building activity monitoring systems and describing their own scenarios dedicated to specific applications.

By relying on the DTK software infrastructure, the goals are:

- to simplify the exchanges of algorithms between different INRIA teams using the DTK;
- to use the facilities already provided by the DTK allowing to compose quickly existing plugins. Currently a python interface is almost operational, and we plan to take advantage of the graphical composer to prototype quickly new workflows, or reconfigure existing ones, for the experimentations conducted by the team.

In order to be confident on the results obtained with the SUP platform, an important effort is done to check:

- the correct behavior of the platform from a software engineering point of view, i.e. that the functionalities of the SUP software are correctly provided, or are not broken by modifications;
- a qualitative evaluation tool for the algorithms, which compares the results obtained with the algorithms to ground truth for several reference videos.

Both tests are run on a daily basis.

B Ethical and Privacy Issues

For both studied application domains, Stars has to address several ethical issues.

In particular, for the healthcare domain, a 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 indicating the progression of their condition.

The development and ultimate use of novel assistive technologies by a vulnerable user group such as individuals with dementia, and the assessment methodologies planned by Stars are not free of ethical, or even legal concerns. Thus Stars will pay particular attention to ethical, 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.

Informed consent, privacy and good clinical practice

Throughout the duration of the Stars project, the ethical welfare of the participants will be our primary concern. Privacy will be ensured by design, based on the seven foundational principles: proactive, by default, positive sum, lifestyle protection, visibility/transparency and respect for the users. We will take specific steps to ensure that we always respect the privacy of individuals by immediately anonymising data and only transferring anonymised data between sites. The person with dementia will have the capability of controlling the encryption of the data, as well as to activate or de-activate the monitoring. In the case of video and audio recordings, specific consent will be requested from the person with dementia and their family or caregivers. If such consent is given, all elements identifying the participants will be masked for dissemination.

Ethics and dementia experts, including an Advisory Board, will be consulted on the most appropriate range of mechanisms to acquire informed consent for participating in the studies. This is standard practice from all clinical partners of Stars, which will bring their experience in involving people with dementia in the pilot studies. Mechanisms to withdraw consent without additional stress will also be provided for the participants. Once people with dementia have been selected to participate in the studies, they will need to be approved by a centre director or a responsible committee, such as the Committee for the Protection of Patients in Biomedical Research.

Also, the Charter of Fundamental Rights of the EU will be respected, as it is an important source of rights, including a title that explicitly protects the rights of the elderly.

At this stage of a research project of this nature, it is impossible to anticipate all arising ethical issues associated with the technological goals. Analysing existing related technologies and attempting to project their use into the future does, however, highlight a range of ethical questions, including:

- How should user privacy be protected as they use the applications?
- What data is it ethically acceptable to include in the centralised dementia user profile?
- How should user profile data (as used for inferring user needs, adapting the technology, and personalising the guidance delivered) be protected?
- What are the consequences if incomplete or inaccurate deductions made by the system put the user at increased risk (than would have been the case without the

use of the application)? For example, who is responsible if the user with dementia, following steps recommended by the system, experiences a negative outcome?

- What are the likely effects on user behaviour (both short and longer-term) of using such assistive technologies? For example, will a dependency on the device feedback prevent or encourage more independent, self-initiated behaviour?

Throughout Stars course, we will explore all ethical angles and issues associated with our proposed research objectives. Keeping such issues at the forefront of our thinking, we will engage with dementia experts, ethics experts, legal experts and end-users in discussion on these aspects to elicit their opinions and establish pragmatic approaches to take.

For instance, F. Bremond has been a member of the ODEGAM - "Commission Ethique et Droit" (a local association in Nice area for ethical issues related to older people) from 2010 to 2011 and a member of the French scientifique counsil for the national seminar on "La maladie d'alzheimer et les nouvelles technologies - Enjeux éthiques et questions de société" in 2011. This counsil has in particular proposed a charte and guidelines for conducting researchs with dementia patients.

Legal Issues

To add some concrete direction to these considerations, there are some legal frameworks under which the research project will operate, which can be divided into legal provisions covering data, covering medical devices and covering clinical trials.

Since our proposal involves gathering, storing, analysing and transferring data about individuals, we will be required to comply with the provisions of EU Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data and this will guide our processes. In addition, at a national level, each participant responsible for running a pilot study will need to comply with national legislation. For instance in France, this corresponds to "Loi informatique et libertés", 1978 (Act n°78-17 of 6 January 1978 on Data Processing, Data Files and Individual Liberties amended by the Act of 6 August 2004, relating to the protection of individuals with regard to the processing of personal data). For instance, F. Bremond has been appointed as the Data Controller of the Dem@Care project.

In terms of the legal provisions governing the devices we propose to use in Stars, there are no legal issues arising throughout the implementation of the Stars proposal from a medical law perspective. By law, a medical device is defined as a product used for the diagnosis, prevention, monitoring or treatment of illness or disability. General Medical Devices are regulated by EU Directives 93/42/EEC. The devices used as part of the Stars proposal to sense people activities etc. do not fit this description, and so are unregulated by the directive, meaning that EU and national law does not raise issues for us.

Finally, the issue of whether our pilot trials can be classed as clinical trials is covered by the EU Directive on Clinical Trials 2001/20/EC. An investigational medicinal product is defined as a pharmaceutical form of an active substance or placebo being tested or used as a reference in a clinical trial. The Control of Clinical Trials Acts 1987-1990 regulates clinical trials not falling within the EU Directive on Clinical Trials but this legislation only regulates pharmaceutical products or substances. The devices used as part of the Stars proposal do not fit this description, and so once again, are unregulated by the directive.