INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Orion

Intelligent Environments for Problem Solving by Autonomous Systems

Sophia Antipolis - Méditerranée

THEME COG

Table of contents

1. Team

Head
Monique Thonnat [DR1 Inria, HdR]

Team Assistant
Catherine Martin

Research Scientists
François Brémond [CR1 Inria, HdR]
Guillaume Charpiat [CR2 Inria since December 2007]
Sabine Moisan [CR1 Inria, HdR]
Annie Ressouche [CR1 Inria]

Long Term Invited Professor
Jean-Paul Rigault [Professor, Nice Sophia-Antipolis University, from September 2005 to September 2007]

External collaborator
Jean-Paul Rigault [Professor, Nice Sophia-Antipolis University, from October 2007]

Technical Staff
Bernard Boulay [Collaboration with ST Micro Electronics, since February 2007]
Etienne Corvée [European Project CARETAKER]
Ruihua Ma [European project SERKET]
José Luis Patino Vilchis [European Project CARETAKER]
Tomi Raty [European project SERKET, since March 2007]
Valery Valentin [SIC Project, since May 2007]
Thinh Van Vu [European project SERKET, up to June 2007]

PhD Students
Bernard Boulay [Paca Lab Grant, Nice Sophia-Antipolis University, HdR]
Binh Bui [CIFFRE RATP Grant, Nice Sophia-Antipolis University]
Mohamed Bécha Kaâniche [Paca Lab Grant, Nice Sophia-Antipolis University]
Naoufel Khayati [ENSI Tunis]
Lan Le Thi [Hanoi University and Nice Sophia-Antipolis University]
Anh Tuan Nghiem [Nice Sophia-Antipolis University, from 1 December 2006]
Vincent Martin [Regional Grant, Nice Sophia-Antipolis University]
Nadia Zouba [Nice Sophia-Antipolis University]
Marcos Zúñiga [CONYCIT Grant, Nice Sophia-Antipolis University]

Post-doctoral fellows
Sundaram Suresh [ERICM PostDoc, since August 2007]

Intern Students
Yoshimura Masaki [EGIDE Inria since April 2007 up to October 2007]
Raoudha Chebil [Master ENSI Tunis]
Makrem Djebali [Master ENI Tunis]

Visitors
Wolfgang Ponweiser [Vienna University from 27 August 2007 up to 30 August 2007]

2. Overall Objectives

2.1. Presentation

Orion is a multi-disciplinary team at the frontier of computer vision, knowledge-based systems(KBS), and software engineering.

The Orion team is interested in research on **reusable intelligent systems** and **cognitive vision**.

2.1.1. Research Themes

More precisely, our objective is the design of intelligent systems based on knowledge representation, learning and reasoning techniques.

We study two levels of reuse: the reuse of programs and the reuse of tools for knowledge-based system design. We propose an original approach based on **program supervision** techniques which enables to plan modules (or programs) and to control their execution. Our researches concern knowledge representation about programs and their use as well as planning techniques. Moreover, relying on state-of-the-art practices in software engineering and in object-oriented languages we propose a platform that facilitates the construction of **cognitive systems**.

In cognitive vision we focus on two research areas of **automatic image understanding**: *video sequence understanding* and *complex object recognition*. Our researches thus concern knowledge representation of objects, of events and of scenarios to be recognized, as well as knowledge about the reasoning processes that are necessary for image understanding, like categorization for object recognition.

2.1.2. International and Industrial Cooperation

Our work has been applied in the context of 3 European projects: AVITRACK, SERKET, CARETAKER. We have industrial collaborations in several domains: transportation (CCI Airport Toulouse Blagnac, SNCF, INRETS, ALSTOM, RATP, Roma ATAC Transport Agency (Italy)), banking (Crédit Agricole Bank Corporation, Eurotelis and Ciel), security (THALES R&T FR, THALES Security Syst, INDRA (Spain), EADS, Capvidia, Multitel, FPMs, ACIC, BARCO, VUB-STRO and VUB-ETRO (Belgium)), multimedia (Multitel (Belgium), Thales Communications, IDIAP (Switzerland), SOLID software editor for multimedia data basis (Finland)), civil engineering sector (Centre Scientifique et Technique du Bâtiment (CSTB)), computer industry (BULL), software industry (SOLID software editor for multimedia data basis (Finland), Silogic S.A) and hardware industry (ST-Microelectronics). We have international cooperations with research centers such as Reading University (UK), ARC Seibersdorf research GMBHf (Wien Austria), ENSI Tunis (Tunisia), National Cheng Kung University (Taiwan), National Taiwan University (Taiwan), MICA (Vietnam), IPAL (Singapore), I2R (Singapore), NUS (Singapore), University of Southern California (USC), University of South Florida (USF), University of Maryland.

2.2. Highlights of the year

Orion will end after 12 years in December 2007. We will continue within a new project-team Pulsar which will be focused on activity recognition.

3. Scientific Foundations

3.1. Introduction

The research topics we study within Orion concern reusable intelligent systems and cognitive vision. The work we conduct for reusable intelligent systems is mainly based on software engineering and on artificial intelligence techniques. The work we conduct for cognitive vision is mainly based on computer vision and artificial intelligence techniques. In the following sections we describe two levels of reusable systems: program supervision and software platform for cognitive systems, two kinds of cognitive vision problems for automatic image understanding: automatic interpretation of image sequences and design of a cognitive vision platform.

3.2. Program Supervision

Keywords: *planning*, *program reuse*, *program supervision*.

Participants: Sabine Moisan, Monique Thonnat.

Program supervision aims at automating the reuse of complex software (*e.g.* image processing program library), by offering original techniques to plan and control processing activities.

Knowledge-based systems are well adapted for the program supervision research domain. Indeed, these techniques achieve the twofold objective of program supervision: to favor the capitalization of knowledge about the use of complex programs and to operationalize this utilization for users not specialized in the domain. We study the problem of modeling knowledge specific to program supervision, in order to define, on the one hand, knowledge description languages and knowledge verification facilities for experts, and, on the other hand, tools (e.g., inference engines) to operationalize program supervision knowledge into software systems dedicated to program supervision. To implement different program supervision systems, we have developed a generic and customizable framework: the LAMA *platform [8], which is devoted both to knowledge base and inference engine design.*

Program supervision aims at automating the (re)use of complex software (for instance image processing library programs). To this end we propose original techniques to plan and control processing activities. Most of the work that can be found in the literature about program supervision is generally motivated by application domain needs (for instance, image processing, signal processing, or scientific computing). Our approach relies on knowledge based systems techniques. A knowledge-based program supervision system emulates the strategy of an expert in the use of the programs. It typically breaks down into:

- a library of executable programs in a particular application domain (e.g., medical image processing),
- a knowledge base for this particular domain, that encapsulates expertise on programs and processing; this primarily includes descriptions of the programs and of their arguments, and also expertise on how to perform automatically different actions, such as initialization of program parameters, assessment of program execution results,
- a general supervision engine, that uses the knowledge stored in the knowledge base for effective selection, planning, execution and control of execution of the programs in different working environments,
- interfaces that are provided to users to express initial processing requests and to experts to browse and modify a knowledge base, as well as to trace an execution of a knowledge-based system.

Program supervision is a very general problem, and program supervision techniques may be applied to any domain that requires complex processing and where each sub-processing corresponds to proper sequencing of several basic programs. To tackle this generality, we provide both knowledge models and software tools. We want them to be both general, i.e., independent of any application and of any library of programs, and flexible, which means that the absence of certain type of knowledge has to be compensated by control mechanisms, like powerful repairing mechanisms.

Program Supervision Model

To better understand the general problem of program supervision and to improve the (re)use of existing programs, the knowledge involved in this activity has to be modeled independently of any application. The knowledge model defines the structure of program descriptions and what issues play a role in the composition of a solution using the programs. It is thus a guideline for representing reusable programs. We have thus used knowledge modeling techniques to design an explicit description of program supervision knowledge to allow the necessary expertise to be captured and stored for supporting of a novice user or an autonomous system performing program supervision. We have modeled concepts and mechanisms of program supervision first for the OCAPI [4] engine, and then for our more recent engines. A preliminary work with KADS expertise model has been improved by using recent component reuse techniques (from Software Engineering), planning techniques (from Artificial Intelligence), existing program supervision systems, and our practical experience in various applications such as obstacle detection in road scenes, medical imaging, or galaxy identification.

Knowledge Base Description Language

In the LAMA platform we have developed the YAKL language that allows experts to describe all the different types of knowledge involved in program supervision, independently of any application domain, of any program library, or of the implementation language of the knowledge-based system (in our case Lisp or C++).

The objective of YAKL is to provide a concrete means to capitalize in a both formal (system-readable) and informal (user-readable) form the necessary skills for the optimal use of programs, for user assistance, documentation, and knowledge management about programs. First, a readable syntax facilitates communication among people (*e.g.*, for documenting programs) and, second, a formal language facilitates the translation of abstract concepts into computer structures that can be managed by software tools.

YAKL is used both as a common storage format for knowledge bases and as a human readable format for writing and consulting knowledge bases. YAKL descriptions can be checked for consistency, and eventually translated into operational code. YAKL is an open extensible language which provides experts with a user-friendly syntax and a well defined semantics for the concepts in our model. [45]

3.3. Software Platform for Cognitive Systems

Keywords: *component reuse*, *frameworks*, *library*, *software engineering*.

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

The LAMA software platform provides a unified environment to design not only knowledge bases, but also inference engines variants, and additional tools. It offers toolkits to build and to adapt all the software elements that compose a knowledge based system (or Cognitive System).

The LAMA *software platform allows us to reuse all the software elements that are necessary to design knowledge-based systems (inference engines, interfaces, knowledge base description languages, verification tools, etc.). It gathers several toolkits to build and to adapt all these software elements. The platform both allows to design program supervision and automatic image interpretation knowledge-based systems and it facilitates the coupling between the two types of systems.*

Designing dedicated tools for a particular task (such as program supervision) has two advantages: on the one hand to focus the knowledge models used by the tools on the particular needs of the task, and, on the other hand to provide unified formalisms, common to all knowledge bases dealing with the same task.

We want to go one step further in order to facilitate also the *reuse of elements that compose a knowledge-based system*. That is why we decided to design a generic and adaptable software development platform, namely the LAMA platform [8]. It gathers several toolkits to build and to adapt all these software elements. Such a platform allows us to tackle the problem of adapting a task – like program supervision, as well as planning, data interpretation, or classification – and tuning it in order to fulfill, for instance, specific domain requirements. LAMA provides both experts and designers with task-oriented tools, i.e. tools that integrate a model of the task to perform, which help to reduce their efforts and place them at an appropriate level of abstraction. The platform thus provides a unified environment to design not only expert knowledge bases, but also variants of inference engines, and additional tools for knowledge-based systems.

LAMA relies on recent techniques in software engineering. It is an object-oriented extensible and portable software platform that implements different layers. First, a general layer, common to a large range of knowledge-based systems and tasks, implements, for instance, inference rules, structured frames, and state management. On top of this common layer, each task has an attached dedicated layer, that implements its model and specific needs. LAMA provides "computational building blocks" (toolkits) to design dedicated tools. The toolkits are complementary but independent, so it is possible to modify, or even add or remove a tool without modifying the rest. Another objective of the platform is to be able to couple knowledge based systems performing different complementary tasks in a unified environment.

We have already used LAMA to design different program supervision engines and variants of them. The platform has substantially simplified the creation of these engines, compared to the amount of work that had been necessary for the previous implementation of OCAPI.

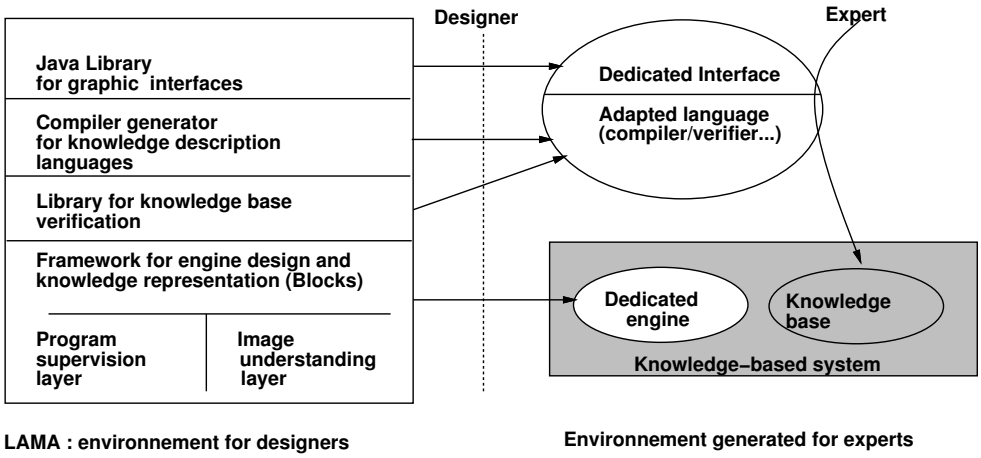


Figure 1. LAMA *architecture and tools for engine design, knowledge base description, verification, and visualization*

The core of the platform (see figure 1) is a *framework* of re-usable components, called BLOCKS, it provides designers with a software framework (in the sense of software engineering). For instance, the program supervision part of the framework offers reusable and adaptable components that implement generic data structures and methods for supporting a program supervision system. BLOCKS also supplies the knowledge concepts of a task ontology (*e.g.*, program supervision ontology) to build knowledge bases. Dedicated description languages that operationalize the conceptual models described in task ontologies, can also be developed. They provide experts with a human readable format for describing, exchanging, and consulting knowledge, independently of any implementation language, any domain, or any application. Additional toolkits are also provided in the platform: a toolkit to design knowledge base editors and parsers – to support the dedicated description language –, a knowledge verification toolkit – adapted to the engine in use –, a toolkit to develop graphical interfaces – both to visualize the contents of a knowledge base and to run the solving of a problem. The most important toolkits are briefly described below.

Framework for Engine Design

BLOCKS (Basic Library Of Components for knowledge-based Systems) is a framework (in the software engineering sense), that offers reusable and adaptable components implementing generic data structures and methods for the design of knowledge-based systems' engines. The objective of BLOCKS is to help designers create new engines and reuse or modify existing ones without extensive code rewriting.

The components of BLOCKS stand at a higher level of abstraction than programming language usual constructs. BLOCKS thus provides an innovative way to design engines. It allows engine designers to speed-up the development (or adaptation) of problem solving methods by sharing common tools and components. Adaptation is often necessary because of evolving domain requirements or constraints.

Using BLOCKS, designers can conveniently compose engines (or, in other words, problem solving methods) by means of basic reasoning components. They can also test, compare or modify different engines in a unified framework. Moreover, this platform allows the reuse of (parts of) existing engines, to develop variants of engines performing the same task.

This approach allows as well a unified vision of various engines and supplies convenient comparisons between them.

Engine Verification Toolkit

From a software engineering point of view, in order to ensure a safe reuse of BLOCKS components, we are working on a verification toolkit for engine behavior. To design new engines, BLOCKS components can be used by composition and/or by sub typing. Among the existing formal methods of verification, we do not choose testing methods, since they are not complete, nor theorem prover techniques, since there are not totally automatic. We prefer *model-checking* techniques, because they are exhaustive, automatic and well-suited to our problem. The goal is to produce safe environments for knowledge based system engine design.

We propose a mathematical model and a formal language to describe the knowledge about engine behaviors. Associated tools may ensure correct and safe reuse of components, as well as automatic simulation and verification, code generation, and run-time checks.

Knowledge Base Verification Toolkit

Knowledge-based systems require a safe building methodology to ensure a good quality. This quality control can be difficult to introduce into the development process due to its unstructured nature. The usual verification methods focus on syntactic verification based on formalisms that represent the knowledge (knowledge representation schemes, like rules or frames) .

Our aim is to provide tools to help experts during the construction of knowledge bases, in order to guarantee a certain degree of reliability in the final system. For this purpose we can rely not only on the knowledge representation schemes, but also on the underlying model of the task that is implemented in the knowledge based system (tasks supported by the LAMA platform are currently program supervision, model calibration and data interpretation).

The toolkit for verification of knowledge bases is composed of a set of functions to perform knowledge verifications. These verifications are based on the properties of the modes of representation of the knowledge used in the knowledge based systems (frames and rules), but it can be adapted to check the role which the various pieces of knowledge play in the task at hand. Our purpose is not only to verify the consistency and the completeness of the base, but also to verify the adequacy of the knowledge with regard to the way an engine is going to use it.

Graphic Interface Framework

Interfaces are an important part of a knowledge-based system. The graphic interface framework is a Java library that follows the same idea as BLOCKS: it relies on a common layer of graphic elements, and specific layers for each task. It allows to customize interfaces for designing and editing knowledge bases and to run them, according to the task and the engine. Thanks to Java, a distributed architecture can also be developed for remote users.

3.4. Automatic Interpretation of Image Sequences

Keywords: *image interpretation*, *image sequences*, *pattern recognition*.

Participants: Francois Brémond, Monique Thonnat.

> **Automatic Image Interpretation** consists in extracting the semantics from data based on a predefined model. This is a specific part of the perception process: automatic interpretation of results coming from the image processing level.

One of the most challenging problems in the domain of computer vision and artificial intelligence is automatic interpretation of image sequences or video understanding. The research in this area concentrates mainly on the development of methods for analysis of visual data in order to extract and process information about the behavior of physical objects in a real world scene. We focus on two main issues: *general solutions for video understanding* and *recognition of complex activities*.

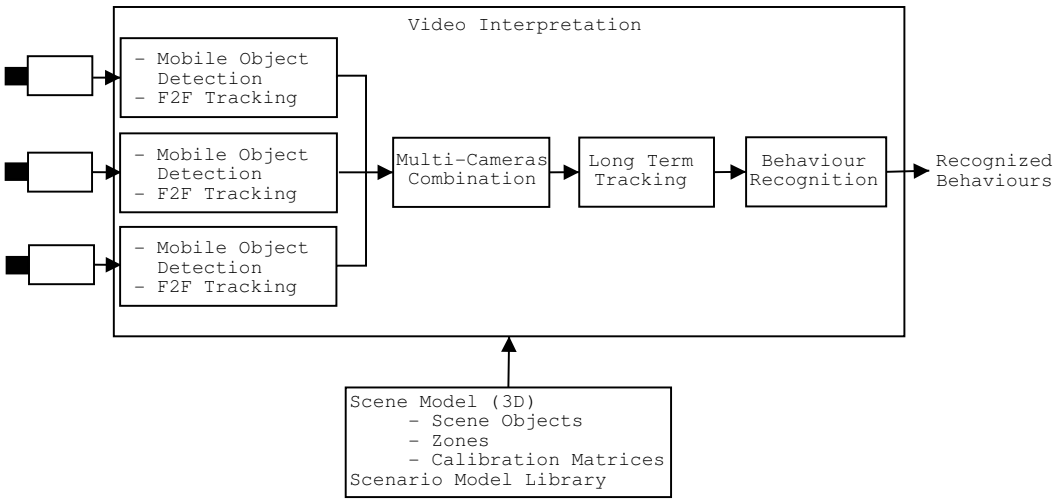


Figure 2. Overview of the interpretation of image sequences.

General Solutions for Video Understanding

In fact the design of general and robust video understanding techniques is still an open problem. To break down this challenging problem into smaller and easier ones, a possible approach is to limit the field of application to specific activities in well-delimited environments. So the scientific community has led researches on automatic traffic surveillance on highways, on pedestrian and vehicle interaction analysis in parking lots or roundabouts, or on human activity monitoring outdoor (like streets and public places) or indoor (like metro stations, bank agencies, houses) environments.

We believe that to obtain a reusable and efficient activity monitoring platform, a single sophisticated piece of program OR software containing all the operations is not adequate because it cannot handle the large diversity of real world applications. We propose to use software engineering and knowledge engineering techniques to combine and integrate several algorithms to handle such diversity.

Other issues remain. Video understanding systems are often difficult to configure and install. To have an efficient system handling the variety of the real world, extended validation and tuning is needed. Automatic capability to adapt to dynamic environments should be added to the platform, which is a new topic of research.

Recognition of Complex Activities

Moreover the recognition of complex activities is also an open problem. Most approaches in the field of video understanding include methods for detection of simple events. We propose a two-step approach to the problem of video understanding:

1. A visual module is used to extract visual cues and primitive events.
2. This information is used in a second stage for the detection of more complex and abstract events also called scenarios.

By dividing the problem into two sub-problems we can use simpler and more domain-independent techniques in each step. The first step makes usually extensive usage of computer vision and stochastic methods for data analysis while the second step conducts structural analysis of the symbolic data gathered in the preceding step. Examples of this two-level architecture can be found in the works of [15].

To solve scenario recognition issues, we study languages to describe scenario models and real-time scenario recognition methods based for instance on temporal constraint resolution techniques. Other issues are still open concerning for instance the learning of primitive events from visual data and the learning of complex scenarios from a large sets of video sequences.

Proposed Approach

To address these issues we thus propose a general model for video understanding based on it knowledge (containing the scene model and a library of scenario models) and on the cooperation of 4 tasks (see figure 2): 1) mobile object detection and frame to frame tracking, 2) multi-cameras combination, 3) long term tracking, and 4) behavior recognition. For each camera the first task detects the mobile objects evolving in the scene and tracked them on 2 consecutive images. The second one combines the detected mobile objects from several cameras. This task is optional in the case of one camera. The third task tracks the mobile objects on a long term basis using model of the expected objects to be tracked. The last task consists, thanks to artificial intelligence techniques, in identifying the tracked objects and in recognizing their behavior by matching them with predefined models of one or several scenarios. Our goal is to recognize in real time behaviors involving either isolated individuals, groups of people or crowd from real world video streams coming from a camera network. Thus in this model video understanding takes as input video streams coming from cameras and generates alarms or annotations about the behaviors recognized in the video streams.

To validate this model in the recent years we have designed a platform for image sequence understanding called VSIP (Video Surveillance Interpretation Platform) (see 5.3 section). VSIP is a generic environment for combining algorithms for processing and analysis of videos which allows to flexibly combine and exchange various techniques at the different stages of the video understanding process. Moreover, VSIP is oriented to help developers describing their own scenarios and building systems capable of monitoring behaviors, dedicated to specific applications.

At the first level, VSIP extracts primitive geometric features like areas of motion. Based on them, objects are recognized and tracked. At the second level those events in which the detected objects participate, are recognized. For performing this task, a special representation of events is used which is called event description language [15]. This formalism is based on an ontology for video events which defines concepts and relations between these concepts in the domain of human activity monitoring. The major concepts encompass different object types and the understanding of their behavior (e.g. "Fighting", "Blocking", “Vandalism”, “Overcrowding”) from the point of view of the domain expert.

3.5. Cognitive Vision Platform

Keywords: *classification*, *cognitive vision*, *image formation*, *learning*, *scenario recognition*.

Participants: Sabine Moisan, Monique Thonnat.

The Cognitive Vision Platform is based on reasoning, learning and image processing mechanisms.

We study the problem of semantic image interpretation which can be informally defined as the automatic extraction of the meaning (semantics) of an image. This complex problem can be simply illustrated with the example shown in figure 3.

When we look at the image on the left of figure 3, we have to answer to the following question: *what is the semantic content of this image?* According to the level of knowledge of the interpreter, various interpretations are possible:(1) a white object on a green background; (2) an insect; or (3) an infection of white flies on a rose leaf. All these interpretations are correct and enable us to conclude that semantics is not inside the image. Image interpretation depends on a priori knowledge and contextual knowledge. Our approach for the semantic image interpretation problem involves the following aspects of cognitive vision : knowledge acquisition and representation, reasoning, machine learning and program supervision. We want to design a generic and reusable cognitive vision platform dedicated to semantic image understanding. Currently, we have

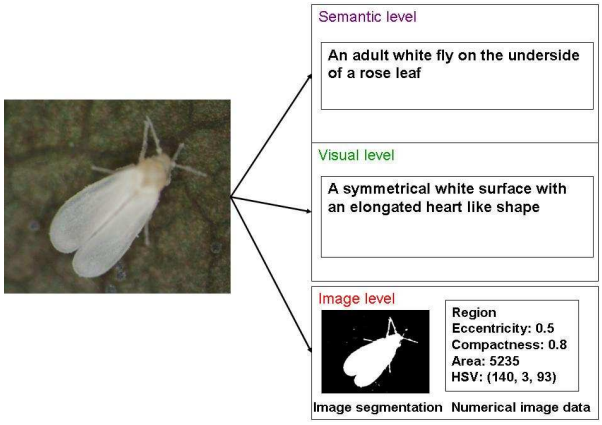


Figure 3. Illustration of the three abstraction levels of data corresponding to the sub-problems of semantic image interpretation. The image is a microscopic biological image.

restricted our works to 2D object recognition and 2D static scene understanding. By *cognitive vision*, we refer, according to the ECVision[1] roadmap, to *the attempt to achieve more robust, resilient and adaptable computer vision systems by endowing them with cognitive faculties: the ability to learn, adapt, weight alternative solutions, and even the ability to develop new strategies for analysis and interpretation*.
We have focused our attention on :

- **the design of a minimal architecture** : more than a solution for a specific application, the platform is a modular system which provides reusable and generic tools for applications involving semantic image interpretation needs;
- **the specification of goals** : to be intelligent a system must deal with goals. It has to be able to choose itself, according to an priori knowledge and contextual knowledge, actions to perform to accomplish the specified goals;
- **the interactivity of the platform with its environment** : the cognitive vision platform has to be able to adapt its behavior by taking into account end-user specifications. In particular, a high level language based on an ontology allows to describe new classes of objects. The work on ontological engineering presented above takes part on this requirement.
- **the development of learning capabilities** : As explained in the ECVision roadmap, *cognitive systems are shaped by their experiences*. That is why the development of learning capabilities is crucial for cognitive vision systems.

Object recognition and scene understanding are difficult problems. Both can be divided into the following more tractable sub-tasks (fig. 4):

1. high-level semantic interpretation;
2. mapping between high level representations of physical objects and image numerical data (i.e. symbol grounding problem);
3. image processing (i.e. segmentation and feature extraction).

[1] The European Research Network for Cognitive Computer Vision Systems, www.ecvsion.org

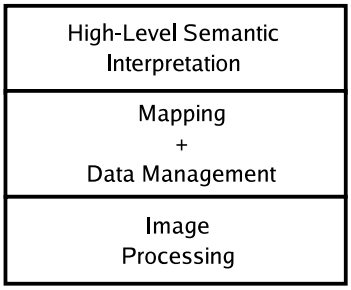


Figure 4. The problem of image interpretation is divided into three sub-tasks.

For each sub-task, the abstraction level of data, the level of knowledge and the reasoning is different as illustrated in figure 3. To separate the different types of knowledge and the different reasoning strategies involved in the object recognition and scene understanding processes, we propose a distributed architecture based on three highly specialized modules :

- a semantic interpretation module;
- a visual data management module;
- an image processing module.

We are interested in both the cognitive and the software engineering issues involved in the design of such a platform. One strong point of the proposed cognitive vision platform is its modularity. This means that each sub-task can be treated by different approaches and that additional functionalities can be added easily. The current implementation is based on the development platform LAMA (3.3).

4. Application Domains

4.1. Overview

Keywords: *astronomy*, *bioinformatics*, *environment*, *health*, *multimedia*, *transportation*, *visual surveillance*.

While in our research the focus is to develop techniques, models and platforms that are generic and reusable, we also make effort in the development of real applications. The motivation is twofold. The first is to validate the new ideas and approaches we introduced. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, the applications we achieved cover a wide variety of domains: automatic classification of galaxies in astronomy, intelligent visual surveillance of underground stations, or applications in medical domain.

4.2. Astronomic Imagery

The complete automation of galaxy description and classification with respect to their morphological type based on images is an historic application in our team [13] [52]. We are expert in this domain both concerning the image processing of galaxies field and theoretical models for morphological classification. This application is a reference to validate our models and software related to interpretation for complex objects understanding and to program supervision [53], [54].

4.3. Video Surveillance

The growing feeling of insecurity among the population led the private companies as well as the public authorities to deploy more and more security systems. For the safety of the public places, the video camera based surveillance techniques are commonly used, but the multiplication of the camera number leads to the saturation of transmission and analysis means (it is difficult to supervise simultaneously hundreds of screens). For example, 1000 cameras are now used for monitoring the subway network of Brussels. In the framework of our works on automatic video interpretation, we have studied since 1994 the conception of an automatic platform which can assist the video-surveillance operators.

The aim of this platform is to act as a filter, sorting the scenes which can be interesting for a human operator. This platform is based on the cooperation between an image processing component and an interpretation component using artificial intelligent techniques. Thanks to this cooperation, this platform automatically recognize different scenarios of interest in order to alert the operators. These works have been realized with academic and industrial partners, like European projects Esprit Passwords, AVS-PV, AVS-RTPW, ADVISOR and AVITRACK and more recently, European projects CARETAKER and SERKET, industrial projects RATP, CASSIOPEE, ALSTOM and SNCF. A first set of very simple applications for the indoor night surveillance of supermarket (AUCHAN) showed the feasibility of this approach. A second range of applications has been to investigate the parking monitoring where the rather large viewing angle makes it possible to see many different objects (car, pedestrian, trolley) in a changing environment (illumination, parked cars, trees shacked by the wind, etc.). This set of applications allowed us to test various methods of tracking, trajectory analysis and recognition of typical cases (occlusion, creation and separation of groups, etc).

Since 1997, we have studied and developed video surveillance techniques in the transport domain which requires the analysis and the recognition of groups of persons observed from lateral and low position viewing angle in subway stations (subways of Nuremberg, Brussels, Charleroi and Barcelona). We worked in cooperation with Bull company in the Dyade Telescope action, on the conception of a video surveillance intelligent platform which is independent of a particular application. The principal constraints are the use of fixed cameras and the possibility to specify the scenarios to be recognized, which depend on the particular application, based on scenario models which are independent from the recognition system. The collaboration with Bull has been continued through the European project ADVISOR until March, 2003. Also, we experimented in the framework of a national cooperation, the application of our video interpretation techniques to the problem of the media based-communication. In this case, the scene interpretation is a way to decide which information has to be transmitted by a multimedia interface.

In parallel of the video surveillance of subway stations, since 2000, new projects based on the video understanding platform have started for new applications, like bank agency monitoring, train car surveillance and aircraft activities monitoring to manage complex interactions between different types of objects (vehicles, persons, aircrafts). A new challenge consists in combining video understanding with data mining as it is done in the CARETAKER project to infer new knowledge on observed scenes.

4.4. Early Detection of Plant Diseases

In the Environment domain, Orion is interested in the automation of the early detection of plant diseases. The goal is to detect, to identify and to accurately quantify the first symptoms of diseases or pest initial presence. As plant health monitoring is still carried out by humans, the plant diagnosis is limited by the human visual capabilities whereas most of the first symptoms are microscopic. Due to the visual nature of the plant monitoring task, computer vision techniques seem to be well adapted. We make use of complex object recognition methods including image processing, pattern recognition, scene analysis, knowledge based systems. Our work takes place in a large-scale and multidisciplinary research program (IPC: Integrated Crop Production) ultimately aimed at reducing pesticide application. We focus on the early detection of powdery mildew on greenhouse rose trees. Powdery mildew has been identified by the Chambre d'Agriculture as a major issue in ornamental crop production. As the proposed methods are generic, the expected results concern all the horticultural network.

Objects of interest can be fungi or insects. Fungi appear as thin networks more or less developed and insects have various shapes and appearances. We have to deal with two main problems: the detection of the objects and their semantic interpretation for an accurate diagnosis. In our case, due to the various and complex structures of the vegetal support and to the complexity of the objects themselves, a purely bottom up analysis is insufficient and explicit biological knowledge must be used. Moreover, to make the system generic, the system has to process images in an intelligent way, i.e. to be able to adapt itself to different image processing requests and image contexts (different sensors, different acquisition conditions). We proposed a generic cognitive vision platform based on the cooperation of three knowledge based systems.

This work is taking part in a two year research agreement between the Orion team and INRA (Institut National de Recherche Agronomique) started in November 2002. This research agreement continues the COLOR (COoperation LOcale de Recherche) HORTICOL started in September 2000.

4.5. Medical Applications

In the Medical domain, Orion is interested in the long-term monitoring of a person at home, which aims to support the caregivers by providing information about the occurrence of worrying change in the person's behavior. We are especially involved in the GER'HOME project, funded by the PACA region, in collaboration with two local partners: CSTB and Nice City hospital. In this project, an experimental home that integrates new information and communication technologies is built in Sophia Antipolis City. The purpose concerns the issue of monitoring and learning about person activities at home, using autonomous and non-intrusive sensors. The goal is to detect the sudden occurrence of worrying situations, such as any slow change in a person frailty. The aim of the project is to design an experimental platform, providing services and allowing to test their efficiency.

Some other monitoring applications related to medical domains are also investigated. In collaboration with Nice City hospital (Dr. Nicolas Sirvent, Archet 2), we study the issue of monitoring children in sterile rooms equipped with video cameras. The aim is to learn the features of a typical day, so that unusual situations can be detected at any time. The context of monitoring epileptic patients using videos is also investigated in collaboration with Marseille City hospital (Prof. P. Chauvel, La Timone). One purpose in terms of activity monitoring is to model the changes in behavior for a given patient when in ceasure. The ultimate goal is to cluster and/ or classify ceasures of patient groups, so that the localization of the brain lesion of a given patient is more easily determined.

5. Software

5.1. Ocapi

Until 1996 the Orion team has developed and distributed the OCAPI version 2.0 program supervision engine. The users belong to industrial domains (NOESIS, Geoimage, CEA/CESTA) or academic ones (Observatoire de Nice, Observatoire de Paris-Meudon, University of Maryland).

5.2. Pegase

Since September 1996, the Orion team distributes the program supervision engine PEGASE, based on the LAMA platform. The Lisp version has been used at Maryland University and at Genset (Paris). The C++ version (PEGASE+) is now available and is operational at ENSI Tunis(Tunisia) and at CEMAGREF in Lyon (France).

5.3. VSIP

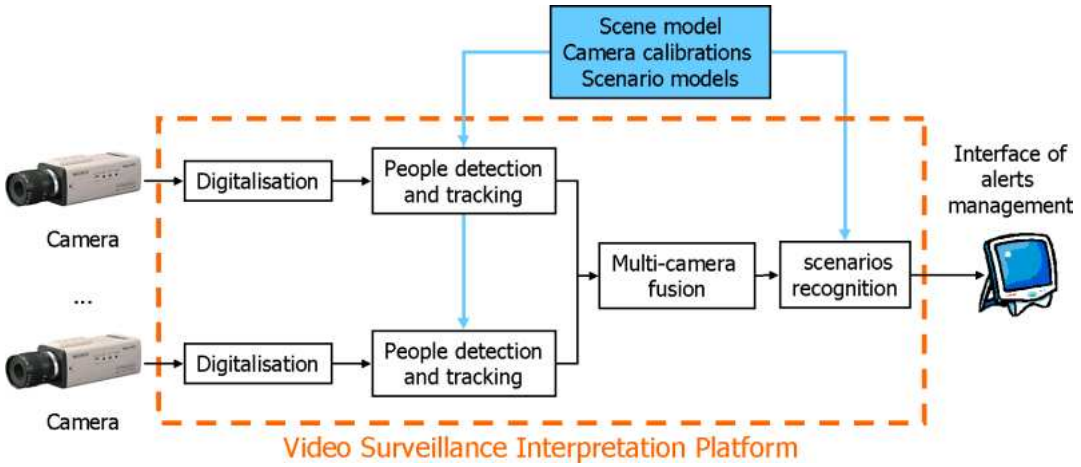


Figure 5. Components of the Video Surveillance Interpretation Platform (VSIP).

VSIP (detailed in 3.4) is a real-time Intelligent Videosurveillance Software Platform written in C and C++ (see figure 5). Actually, four modules of the VSIP platform have been registered at APP (the French agency for patrimony protection) in 2005. These modules are:

1. VSIP-DMM contains the global architecture for data and module management;
2. VSIP-OD contains the image processing algorithms in charge of a video stream of one camera (mobile object detection, classification and frame to frame tracking);
3. VSIP-STA contains the multi-camera algorithms for the spatial and temporal analysis (4D) of the detected mobile objects;
4. VSIP-TSR contains the high level scenario recognition algorithms and scenario representation parsers.

Several versions of VSIP have been transfered to industrial partners: in 2003 to **Bull**, to **Thales**, and to the integrator **Ciel, Toulon**, in 2004 in bank agencies of **Crédit Agricole** and to **Vigitec, Bruxelles** a specialist in Videosurveillance, in July 2005 to **Reading, UK**. VSIP has been exploited by **Keeneo** the Start-up created since July 2005 by the Orion research team.

5.4. PFC

PFC is a real-time 4D software for counting and classification of passengers; this software has been transfered to the Paris subway **RATP**.

6. New Results

6.1. Software Platform for Cognitive Systems

Participants: Raoudha Chebil, Makrem Djebali, Naoufel Khayati, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault.

This year we have improved our work on Program Supervision with the developement of both a web Server and a medical Program Supervision knowlegde base. We have continued our work on distributed Knowledge-based Systems. We have also started the study and implementation of an hybrid reactive synchronous language and of an efficient and re-usable scenario recognition engine.

6.1.1. Introduction

Efficient design of knowledge-based systems is a major research topic in Orion. To this end, the devoted platform LAMA provides a unified environment for the design of knowledge bases, inference engines and additional tools. LAMA defines computational building blocks and toolkits to design dedicated tools. The toolkits are complementary but independent. So it is possible to modify, or even add or remove a tool without modifying the rest. In the last years, we have experienced the profit of using LAMA mainly for developing program supervision engines and variants of them. We now tackle other tasks such as classification. Moreover the platform makes it possible to combine knowledge-based systems performing different tasks: for instance classification together with program supervision were used in a cognitive vision platform [19] applied to agronomy (see 6.3.2).

The core of the platform is a *framework* of re-usable components, called BLOCKS (Basic Library Of Components for Knowledge-based Systems). It offers reusable and adaptable components implementing generic data structures and methods for the design of knowledge-based system engines. The objective of BLOCKS is to help designers create new engines and reuse or modify existing ones without extensive code rewriting. From a software engineering point of view, in order to ensure a safe reuse of BLOCKS components, we continue to develop a toolkit for verifying engine behavior: graphic interfaces, simulation and analysis tools, model-checking, etc.

This year, we have continued to study the distribution of knowledge-based systems. Our example application is Program Supervision in medical imaging, more precisely for osteoporosis detection. In this line, we have focussed on three topics: developing a Web server for Program Supervision, improving osteoporosis knowledge base and medical imaging algorithms, and addressing security and privacy issues in distributed medical applications.

This year, we have also continued to develop an event-driven language devoted to activity recognition. This language is reactive and we now offer a complete toolkit to specify (both graphically and textually) with the language and to compile it in an efficient way.

Concerning scenario specification and recognition, no major practical improvements have been done this year. We have continued to study the application of formal methods to build a real time scenario recognition engine, reusable in different settings and not specific to video based scenarios.

6.1.2. Web Server for Program Supervision

Participants: Sabine Moisan, Jean-Paul Rigault.

Several years ago, we developped a first prototype of a Web server to allow remote access to program supervision techniques for image processing. This server became obsolete due to the Web technologies used. Thus we have completly revisited the server architecture and implementation. The new version relies on current Web standard tools (such as Apache and Tomcat servers) and modern Web programming facilities (such as PHP and Java Server Pages). Through the use of standards, this version aims at better scalability and ease of maintenance and evolution. It also required to modify the program supervision engine (PEGASE+) to improve its communication with the Web server and ultimately with the users.

The primary purpose of this server is to allow physicians to remotely submit their own radiographic images. In the case of osteoporosis detection, bone physicians can thus access to state of the art image processing programs without having to perform low level computer tasks. This approach will be evaluated through our partnership with rheumatologists from the CHR in Orléans, who also provide us with test images.

A secondary goal is to allow exploring and even modifying the knowledge bases that control the use of programs. Of course such features are accessible to knowledge experts only, thus requiring priviledged access rights. Figure 6 illustrates the overall architecture of the server.

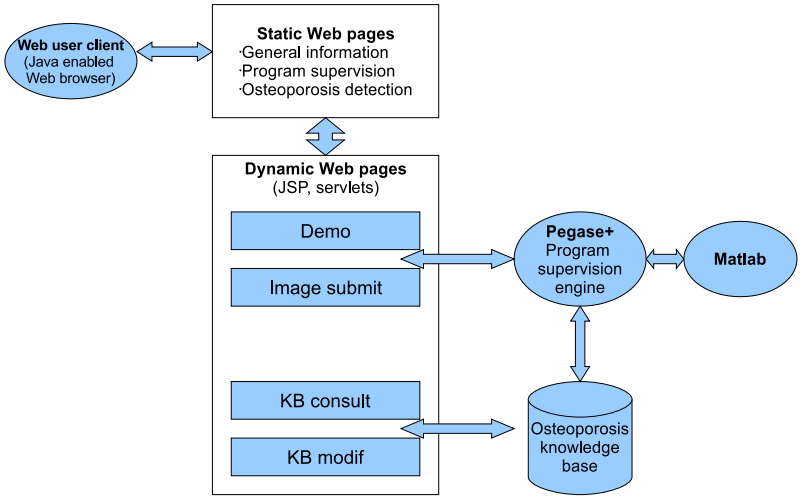


Figure 6. Overview of the Program Supervision Web Server

To fullfil users' needs, not only was the Web technology upgrade necessary but also improvements in medical imaging algorithms as well as in the corresponding program supervision knowledge base were suitable. This is the topic of the next section.

6.1.3. Medical Program Supervision

Participants: Raoudha Chebil, Makrem Djebali, Sabine Moisan, Jean-Paul Rigault.

For our osteoporosis detection application, the image processing algorithms and the knowledge base have been developped in collaboration with ENSI (SOIE laboratory) and ENIT (U2S team), two engineering schools in Tunis. Both algorithms and knowledge base needed revamping.

The original image processing algorithms were witten in Matlab, a language convenient for prototyping, but incurring performance problems when it comes to interactive usage, especially as the size of images tends to increase. A first improvement was to translate some time consuming algorithms from Matlab to C. A dramatic gain of several order of magnitude has been reached. For instance, one particular operation (skeletonization of a 256×256 image) which lasted about 2 minutes in Matlab takes as little as 3 seconds when rewritten in C.

Since we expect to handle much bigger images in the near future (say 1024×1024 or more), we decided to go a little further by proposing concurrent implementation and execution. Indeed multiprocessor (multi core) systems are now available at low cost, and the same parallelization techniques can be applied to GRID architecture. As a matter of example, we have studied how to parallelize the execution of the already mentioned skeletonization algorithm and we implemented it. The result was an execution time divided by about the number of available processors (4 in our case).

Concerning the Program Supervision knowledge base, its modification appeared mandatory for several reasons: existing Matlab programs had evolved and been reorganized, others had been translated into C, and new decision rules emerged from discussions with image processing experts. Consequently the base has been reworked to support new programs and to handle mixing C and Matlab programs. Several new decision rules

have been added (1 parameter initialization, 3 choices, 5 evaluation rule sets). As a consequence, the control behaviour of the KBS is now richer and better fits imaging expertise.

Another extension to the knowledge base aims at introducing statistical processing of the resulting osteoporosis images. For this we interface PEGASE+ with the R system[2], in a similar way as it has been interfaced with Matlab.

6.1.4. Distributed Knowledge-based Systems

Participants: Naoufel Khayati, Sabine Moisan, Jean-Paul Rigault.

In the last years, collaborating with ENSI Tunis, we have developed a prototype of distributed knowledge-based system based on mobile agents. Our example application is still medical imaging and more precisely osteoporosis detection. We use the Aglets system which provides a portable platform for mobile agents, written in Java.

Last year we have improved the architecture and the scenarios of use and tested the whole system [25]. However, security and privacy issues are yet to be studied and solved. In fact these are two different, although related, topics. On the one hand, multi-agent systems are rather vulnerable to security attacks led by malicious agents or other programs, either from the outside or from the inside of the system. Many works have attempted to address these problems, but this is not our main concern. Indeed our system is restricted to highly trusted users who can access to it only after authentication. So we think that the regular security mechanisms of the execution platform (Java) combined with those of the agent platform (Aglets), although not totally bullet-proof, are sufficient to cope with this aspect. On the other hand, our application is dealing with medical data, most of which are confidential. These data concern patients, their pathologies, the corresponding diagnoses, associated images, etc. This second aspect, the privacy of information, is the most important one in our opinion.

In an agent system like Aglets, there are two means for carrying information over the network. One can define "transporter agents" which encapsulate the information in their data members or use inter-agent communications where the information is embedded into messages. In both cases, the information is by default sent in clear, which is not acceptable for sensitive data. Thus we started to study the possibility of encrypting this information in a safe and efficient way. This is work in progress, which has yet to be finalized.

In an other line, it appears that the multi-agent platform that we have chosen might not be ideal. Indeed Agelts is a very simple system, providing the basic mechanisms of agent mobility. However its way too basic, its programming becomes delicate for complex applications, and moreover the Aglets project is frozen and no evolution or improvements can be expected in the future. Thus we are studying the possibility of using more recent and powerful multi-agent platforms. Two obvious candidates are the Jade [3] and the ProActive [4] systems. A comparison between them is underway.

6.1.5. Component Framework Verification

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

We had defined both a dedicated language (BDL) to express BLOCKS component behavior and a *synchronous* mathematical model to give a semantics to BDL programs [46]. We formally characterized the notion of behavioral substitutability and proved that the corresponding preorder is stable with respect to the BDL constructions and thus that substitutability can be verified in a compositional way. In 2006, we have completed our synchronous formalism and interfaced with the latest version of the NUSMV model checker. The theoretical aspect of this work has been completed. Pursuing this work, we now aim at giving effective verification means to use the LAMA platform in a safe way. The technical realization requires to implement a substitutability analyzer for LAMA components.

[2] http://www.r-project.org/
[3] http://www.jadeproject.ca/cdmp.html
[4] http://proactive.inria.fr

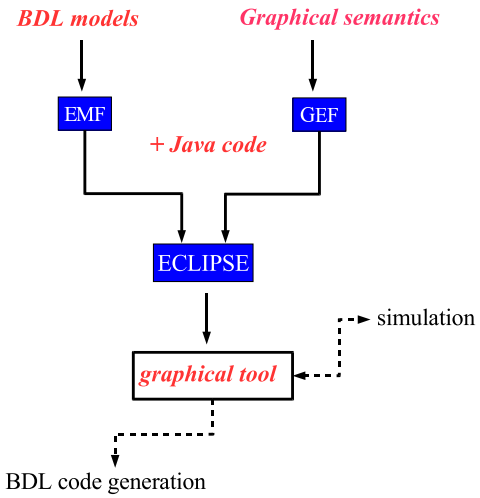


Figure 7. Overview of the Developemnt Scheme for BDL Graphical Interface

This year, the development of a graphical interface has been started through a master course [37]. This tool has been built to express BDL programs and to simulate them with respect to their synchronous semantics. Instead of developing a specific tool, we are developing a generic graphical interface relying on Eclipse and its Graphical Editing Framework (GEF) paired with its Modeling Framework (EMF) to edit hierarchical and parallel automata. This developement scheme is shown in figure 7. Such an approach will allow to obtain dedicated graphical interfaces by straightforward specialization. For instance, we began to specialize this generic graphical interface to enter BDL programs.

6.1.6. Hybrid Event-driven Language

Participant: Annie Ressouche.

The aim of this research is to develop a special purpose language devoted to activity recognition. This high level language allows to specify applications reacting to events comming from different sensors. Such a language naturally belongs to the family of synchronous reactive languages. On the other hand, Model Driven Software Developement is now well known as a way to manage complexity, to achieve high re-use level, and to significantly reduce the development effort. Thus, in collaboration with V. Roy (CMA Ecole des Mines) and D. Gaffé (CNRS and UNSA), we have designed a new specification model based on reactive synchronous approach. Therefore, we benefit from a formal framework well suited to compilation and formal validation. In practice, we designed and implemented a special purpose language (LE) together with its two semantics: a *behavioral semantics* to define a program by the set of its behaviors, avoiding ambiguities in program interpretations; an *execution equational semantics* to allow modular compilation of programs into software and hardware targets (C code, VHDL code, FPGA synthesis, observers...). Our approach fulfills two main requirements of critical realistic applications: modular compilation to deal with large systems and model-based approach to perform formal validation.

A first consequence of such an approach is that program debugging, testing, and validation is made easier. In particular, formal verification is possible with techniques like model checking. Another consequence is that we are able to automatically generate code from specifications. To deal with values in labeled transition systems, we rely on static analysis and abstract interpretation methods [41]. Abstract interpretation makes it possible to define interpretation functions mapping an hybrid program to a synchronous one—where data are abstracted as boolean values— and to apply verification methods to the control part of the hybrid program.

To summarize, we defined a new synchronous language LE that supports separated compilation. Its behavioral semantics provides each program with a meaning, allowing formal verification; its equational semantics allows to compile programs in a truly separated way.

These results together with the language description are detailed in two forthcoming reports [36]. This year, we have improved the language by introducing automata as a native construct and we have completed a compiler (`clem`) implementing the equational semantic rules. This compiler is the central component of our design toolkit. This toolkit comprises different editors to enter programs (a graphical one named `galaxy` for drawing automata and a textual one supporting theLE language. It has several back-ends:

- *code generation*: C code for software applications and VHDL for hardware targets;
- *simulation tools*: thanks to the *blif* format generation we target our own simulator (*blif_simul*);
- *verification tools*: *blif* is well-suited to target several model-checkers (xeve, sis) and our own automata equivalence verifier (blif2autom, blifequiv).

The toolkit architecture is shown in figure 8. The toolkit is available at http://www.unice.fr/L-EEA/LE.html.

Presently, the reactive part of LE is completed, but we still need to handle sensor values in the language as well as to integrate abstract interpretation techniques.

6.1.7. Scenario Description and Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Thinh Van Vu.

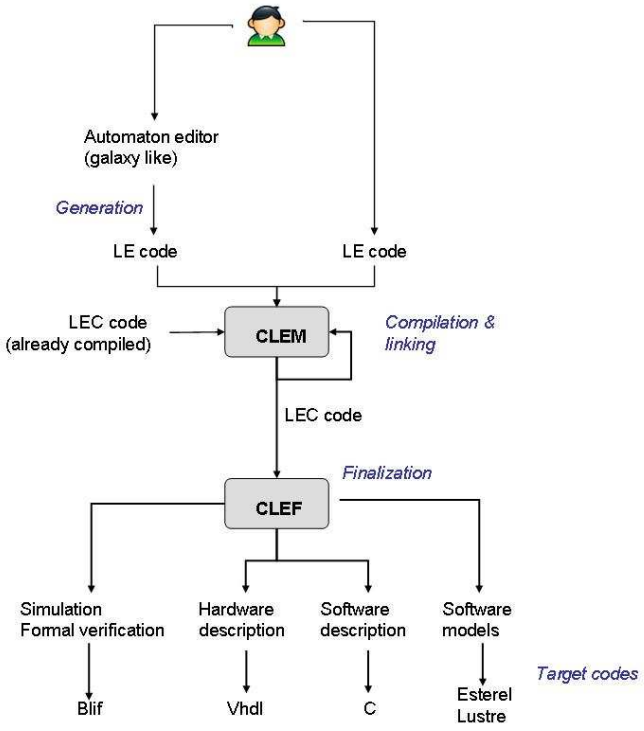


Figure 8. LE Compilation Scheme

Scenarios representation and recognition is a research topic studied for a long time in the Orion project. Recently, Van-Thinh Vu has proposed a language to describe human behaviors together with an algorithm to recognize temporal scenarios from video sequences [15]. The underlying scenario recognition engine is real time, a strong and important feature in the domain.

This year we have studied models of scenarios dealing with both real time (to be realistic and efficient in the recognition phase) and logic time (to benefit from well-known mathematical models allowing re-usability, easy extension and verification). Scenarios are mostly used to specify the way a system may react to sensor inputs. To address these needs (logic time, real time and uncertainty) we have defined a language to express scenarios as modular compositions of automata. Basic scenarios are composed of events, while general scenarios support hierarchy, i.e parallel or sequential sub-scenarios. Temporal constraints can be expressed. This year we started the study of a scenario recognition mechanism that is both efficient and real time.

Moreover, a PhD started at CMA (Ecole des Mines) in October 2006 on the topic: "Formal methods to model and recognize scenarios", co-directed by V. Roy (Ecole des Mines) and A. Ressouche. During this first year, the candidate has particularly studied specific spatio temporal logics (epistemologic logic, probabilistic logic), in order to take into account the uncertainty of sensor results in a scenario recognition engine.

6.2. Automatic Interpretation of Image Sequences

Participants: François Brémond, Bernard Boulay, Binh Bui, Mohamed Bécha Kaâniche, Ruihua Ma, Vincent Martin, Anh Tuan Nghiem, José Luis Patino Vilchis, Tomi Raty, Lan Le Thi, Monique Thonnat, Valéry Valentin, Masaki Yoshimura, Thinh Van Vu, Nadia Zouba, Marcos Zúñiga.

Our goal here is to automate the understanding of the activities happening in a scene by analyzing signals from multiple sensors. Sensors are mainly one or several fixed and monocular video cameras in indoor or outdoor scenes; the observed mobile objects are mainly humans and human-made objects such as vehicles etc. Our objective is to model the interpretation process of image sequences and other perception signals and to validate this model through the design and development of a generic interpretation platform. This year, the techniques developed have been applied in six projects: the European IST project CARETAKER, the European ITEA project SERKET, the SIC project (department of research, pôle de competitivite) and GERHOME project (PACA region), and two industrial projects: Intelligent Cameras (STmicroelectonics) and PFC (RATP).

6.2.1. Introduction

The problem is the interpretation of the behavior of people moving in a scene; i.e. to find a *meaning* to their evolution and their dynamics in the scene. This scene is observed by one or several fixed video cameras and sensors of other types such as contact sensors. To realize the interpretation, we need to solve two sub-problems. The first one is to provide for each frame measures about the scene content. The system in charge of this sub-problem is called "perception" module. The second sub-problem is to understand the scene content. To accomplish this process, we try to recognize predefined scenarios based on perceptual invariants. The system in charge of the second problem is the scenario recognition module. Our approach to image sequence interpretation is based on 3D reasoning in the real world and on the *a priori* model of the observed environment.

This year, we have refined our work on mobile object detection to process crowd scenes and to recognize crowd behaviors in the future. The segmentation task has been enriched with a learning-based approach for adaptive segmentation. This method is particularly interesting for very long sequences (several hours) by adapting the background model to image variations (for instance illumniation changes). We have extended a method for object categorization by taking advantage of the colour distribution and the 3D parallelepiped volume of the observed objects. We have improved our approach for the recognition of human postures by optimizing the combination of motion based, 2D and 3D techniques. We have started to study the combination of video cameras with other sensors (eg., contact sensors), in particular for homecare applications. We have continued designing new unsupervised learning techniques to help users to define scenario models and to extract new types of knowledge from the video signal. We have also continued our work on the evaluation of the video understanding platform on two specific applications: SERKET and SIC. In the framework of the ETISEO

project, we have extended the evaluation methodology to gain more insight into video analysis algorithms. We have started new work on learning trajectory descriptors for video retrieval.

6.2.2. Adaptive Video Segmentation

Participants: Vincent Martin, Monique Thonnat.

In this work, the goal is to detect moving objects (e.g. a person) in the field of view of a fixed video camera. Detection is usually carried out by video segmentation algorithms using background modelling methods. Video segmentation algorithms can be decomposed into two classes: algorithms which rely on a training stage for background modelling (e.g., mixture of Gaussian or codebook models) and the others (e.g., optical flow, running average). In the first case, the quality of the segmentation mostly depends on the quality of the background model training. In the second case, it mostly depends on the parametrization of some key paremeters as the detection threshold. The learning step of our framework for parameter tuning could be used to learn the parametrization of such algorithms. However, this implies to manually segment a lot of training samples with foreground objects. We prefer to save the user from this tedious task and we more focus on the learning-based video segmentation algorithms. In this case, strong efforts have been done to cope with quick-illumination changes or long term changes, but coping with both problems altogether remains an open issue. For this task, we propose an approach for dynamic background model selection based on context analysis.

Our approach is based on a preliminary (off-line) weakly supervised learning module (see Figure 9) during which the knowledge of the context variations is acquired. The role of the user is restricted to establish a training image set composed of background samples that point out context variations. We tackle context modelling problem by performing an unsupervised clustering of the training images. This clustering is based on the analysis of global image characteristics like color variations. At the end of the clustering process, each cluster gathers training images sharing similar global features, i.e. images of the same context.

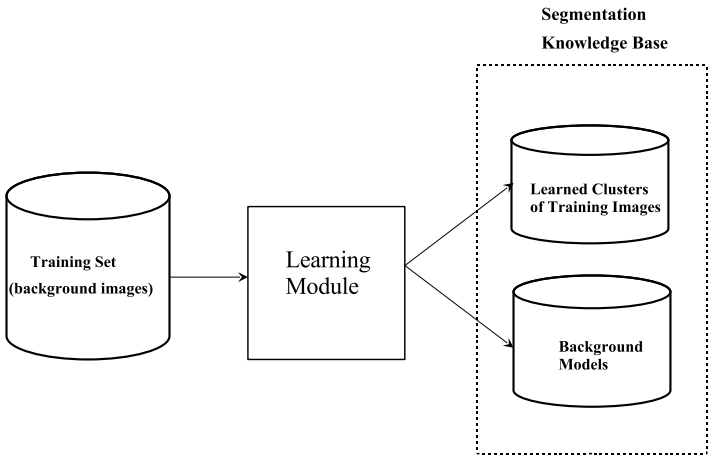


Figure 9. The learning module in video segmentation task.

The goal of the clustering of the training images is to make the background modelling task more reliable by restricting the model parameter space. This approach is particularly interesting for motion segmentation algorithms relying on a training stage of models as mixture of Gaussian [51] or codebook models [44].

In on-line stage, for a new input image, global features are first extracted then a background model is selected and figure-ground segmentation is performed. A temporal context filtering step is applied before segmentation to prevent from incoming erroneous context identification as sketched in Figure 10.

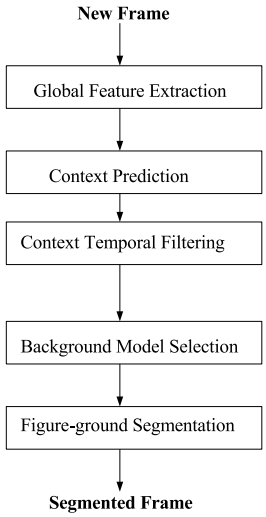


Figure 10. Adaptive figure-ground segmentation schema based on context identification and background model selection.

The first experiments have proved that the dynamic selection of background models is a good approach to deal with adaptation facilities. Nevertheless, it is clear that our approach is still unable to manage unforeseen situations, i.e. new contexts. An extension of this approach to enable continuous learning facility is thus actively needed.

6.2.3. Object Categorization Based on a Video and Optical Cell System

Participants: Binh Bui, François Brémond, Monique Thonnat.

This year, we have completed a new prototype for the classification of people travelling inside the metro station network in Paris. This prototype classifies mobiles into four classes (adult, child, luggage and two adults close to each other) and decides whether these people have to validate a ticket while travelling in the metro. A first prototype was designed in 2004 in a RATP lab and has shown successful classification results. So a new prototype is under construction to validate the approach in a real station. We have tested new hardware (optical fibres, cameras, lenses...) to optimize the prototype performance.

The classification algorithm is based on naive Bayesian classifiers which require large learning data set. Thus, we have completed the database to better estimate the first prototype performance. Now, the new data base contains 5000 samples associated with their ground-truth:

- 1900 for adult;
- 66 for two adults close to each other;
- 1500 for child;
- 1000 for luggage.

A still open issue consists in associating a ground-truth label with a 2D detection in case of overlapping objects. We have decided to consider that two 2D detections get different labels (e.g. adult and luggage) when they are globally separated by a vertical line on one image and they get the same label (e.g. adult + luggage) when they cannot. Thus we could expect to obtain three labelled objects (e.g. adult, luggage and adult + luggage). Ongoing experimentations will tell us soon if this assumption is the correct one.

The performance with the new database is as good as the performance obtained with the first database: 99% of correct classification. Similar experimentations have been performed with different supervised learning techniques: neural network, SVM and other types of Bayesian network.

The ongoing research consists now in establishing the impact of deficient resources on result performance. For that, we will test the prototype with a large variety of deficient resources such as one missing camera, two optical fibres off, several optical points. All the resources will be ordered along their impact on the performance.

6.2.4. Reliable Object Description in Video for Incremental Event Learning

Participants: Marcos Zúñiga, François Brémond, Monique Thonnat.

This year, we have focused our work on two main points:

- The improvement of a previously developed 3D classification method proposed in 2006. We have greatly improved the computation time and added several new functionalities.
- The proposal of a new multiple object tracking approach.

Both points are developed in the framework of a new primitive events learning approach. In 2005, a new incremental event recognition and learning approach for objects evolving in a video scene has been proposed. This new incremental event learning approach basically consists in updating a hierarchical structure using four operators (merging, splitting, deletion and addition) which are selected according to the operator which gives the best category utility increment. We use this method for hierarchically classifying the most frequent states on a video sequence, where each input instance is represented by the detected object class name (e.g. person, vehicle), a set of attributes associated to this object (e.g. posture, trajectory, location), and reliability measures for the object attributes. These reliability measures will allow the learning approach to put emphasis on the relevant object information in terms of their temporal coherence and visibility in the analysed image frames.

For obtaining an appropriate input for this learning approach, data extracted by robust object detection and tracking are needed, as well as the reliability measures for the information obtained by the video analysis phases. Then, information related to objects present in a scene must be carefully obtained or inferred in prior stages of the video processing framework in order to have the appropriate quality of input data for the learning approach.

Following this direction, in 2006 we designed a new 3D object classification approach for monocular video sequences using a simple 3D model of the expected objects in the scene. The proposed approach allows to classify objects of different nature in a way that is independent from the relative position between the object and the camera. This model has been designed considering the pinhole camera model. Visual reliability measures of the estimated 3D dimensions of detected objects have been calculated to take into account their visibility in the analysed frame. This approach has been published in the VIE conference [55] in September 2006.

The first point we have been working this year is the optimisation of the 3D classification algorithm, also adding new functionalities. After this optimisation process, we have performed a test over four little parking lot sequences containing one object to be classified (a person or a vehicle), processing a total of 80 frames. We have tested the algorithm before and after the optimisation process. Before optimisation, the algorithm performance was of $98.6 frames/sec$. After optimisation the algorithm performs at $428.9 frames/sec$, improving the computational performance of the classification approach in a ratio of 435.0%. This results prove that this classification approach can be used in real-time applications with multiple objects.

The added functionalities for this classification method are:

- A function to improve the selection of the optimal 3D model. This function analyses moving pixels information with the most likely obtained 3D configurations in order to select the best configuration in terms of number of moving pixels which fit with the obtained 3D configuration.
- A coherence verification function for static objects present in the scene (walls and context static objects). This function verifies that a 3D configuration for an object is not in a zone not compliant with physical constraints (for instance, crossing a wall).
- A method for solving the static occlusion problem (a mobile object partially occluded by a pre-defined static object or in an image border).
- The possibility of defining objects with different postures (e.g. a person) in order to be able to classify objects whose posture can change.

These improvements have enabled the integration of this 3D classification approach in VSIP platform and its utilisation for different applications. Figure 11 shows some results for an elderly-care application (GERHOME project). Figure 12 shows classification results for a posture detection application.

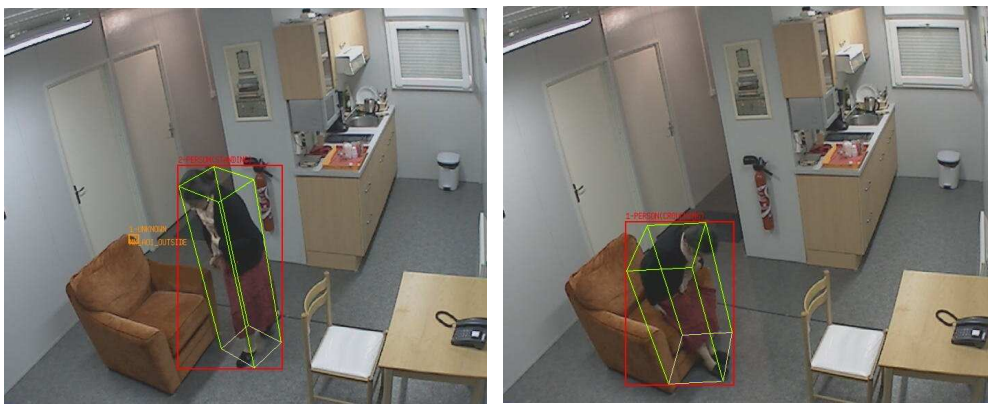


Figure 11. Results for the optimised 3D object classification approach applied to an elderly-care application (GERHOME project). Yellow lines represent the projection of the 3D model of the object in the scene. Figure (a) shows the classification results for a person in standing posture. Figure (b) shows the results for a person in crouching posture.

The second point we have worked this year is a new multiple object tracking algorithm for monocular video sequences. This work focuses on the acquisition of reliable tracking information related to mobile objects present in a scene, together with the reliability measures associated to this information.

Our proposed tracking method has been developed to cope with a wide range of the typical issues present in videos with multiple objects to track. This method maintains a list of likely configurations for the mobile objects present in the scene. The likelihood of these configurations is reinforced or weakened along the time. This approach uses 3D information from generic 3D object models to generate a set of mobile object configuration hypotheses, based on the previous work in 3D object classification. This information corresponds to 3D generic features (e.g. width, height, length, 3D position, orientation) associated with a visual reliability measure to account for the quality of the analysed data. The generated hypotheses are validated or rejected in time according to the information obtained in later frames in the analysed video together with the classification

Figure 12. Results using the optimised 3D object classification approach for a posture detection application. Yellow lines represent the projection of the 3D model of the object in the scene. Three different postures for a person are correctly determined. Figure (a) shows the recognition of a person standing with an arm up. Figure (b) shows the recognition of a crouching posture. Figure (c) shows the recognition of a lying posture.

information of the currently analysed frame. These hypotheses correspond to a set of mobile objects, related to a group of visual evidences in the current frame (e.g. bounding boxes of moving regions, 3D information) or to different path possibilities for a set of tracked mobiles. Each mobile object is represented as a set of statistics (e.g. 2D and 3D object dimensions, 3D position, 3D velocity) inferred from visual evidences of their presence in the scene. The visual evidences are stored as a short-term history buffer. At the same time, reliability measures associated with temporal coherence of visual and temporal features are calculated, together with a global reliability measure for the tracked object.

This tracking algorithm performs with very weak assumptions of initial state of mobile objects in order to prevent losing potential valid solutions. A group of hypotheses is generated according to different possible configurations for a neighbourhood of visual evidences on a given frame. This means that visual evidences can be merged if necessary in order to represent possible mobile configurations according to prior information of the expected objects in the scene. Each group of hypotheses is updated according to visual evidences obtained in later frames, expanding the hypotheses group to take into account different possible mobile object paths if needed. A hypothesis is eliminated if it becomes too unlikely in time, compared with the best current hypothesis in the group.

The likelihood of an hypothesis is calculated based on the weighted mean likelihood of each mobile object in the hypothesis. The weight of a mobile object is directly proportional to its time of presence in the scene, reinforcing the life-time of hypotheses which contain mobile objects strongly validated in time with visual facts. The likelihood for a mobile object is calculated as a summation of the degree of coherence of its features, weighted by the visual reliability of these features. Hypotheses can be separated according to the separability of visual features analysed in the current frame, allowing to separate the tracking procedure into different simpler tracking sub-problems. Separating tracking information is one way to control the combinatorial explosion of hypotheses in time.
After validation of the tracking approach, its output will serve as input to the incremental event learning approach.

6.2.5. Online Learning System for Robust Object Tracking

Participants: Sundaram Suresh, Francois Brémond, Monique Thonnat.

The main challenge in object tracking problem is to faithfully determine the moving object region in each video frame that match with the given target model, under the dynamic change in appearance (both in the object and background), rapid illumination variation, shape/scale change and occlusion. Hence, the objectve of the project is to develop a learning based system to detect and track the objects accurately in a dynamically changing environment. The proposed scheme takes approximate location and bounding box from the object detection mechanism as an initial input

We treat the problem of tracking as a binary classification problem, i.e., detection of object region (class '1') from the background region (class '0'). A new on-line neural network based learning algorithm is proposed to approximate the decision boundary effectively. The estimated posterior probability of the image region from the neural classifier is used to localize the object in the current frame and also used to determine the bounding box. The neural classifier parameters are adapted online based on the tracked object in the current frame to handle the dynamic change in object/background appearance and illumination.

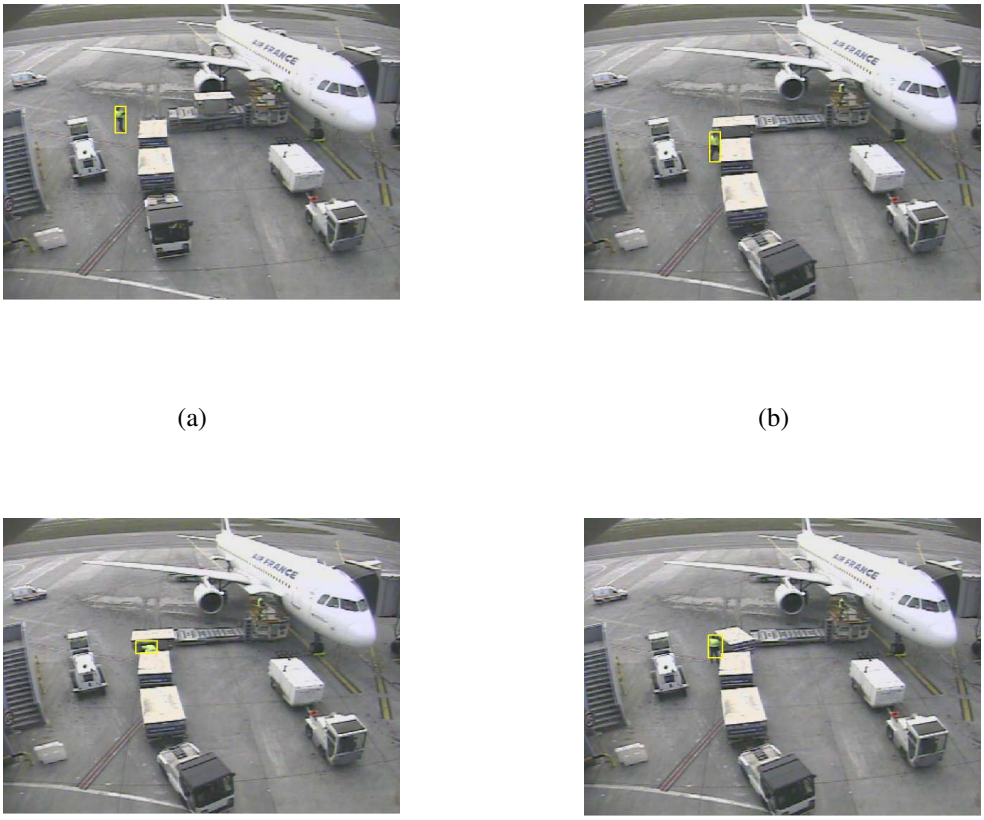
Experimental Results

The approach has been tested on various video sequences obtained using fixed and hand-held camera. The tracking results (Figs. 13-14) clearly indicate the advantage of the proposed on-line learning tracker, which is robust under change in appearance, rapid illumination variation and scale/shape change. Also, the tracking results indicate that the learning based tracker can handle partial occlusion effectively (Fig. 14).

(a) (b) (c)

(d) (e) (f)

Figure 13. Appearance/Illumination: Tracking results for walking sequence in which there is a significant change in illumination and background scene.

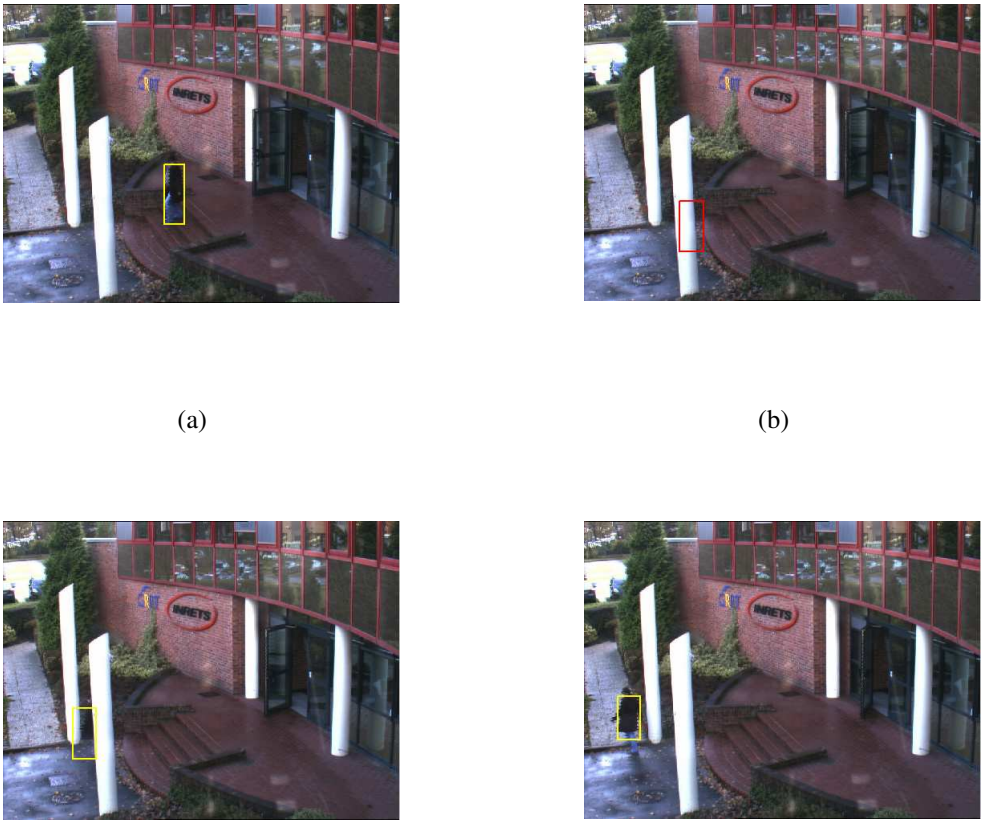


(a) (b)

(c) (d)

Figure 14. Partial occlusion: Tracking results for video sequence in which the airport personnel moves behind the truck and walk away from the camera

We propose a 'hold-and-search' approach to handle complete or brief occlusion. In 'hold-and-search' approach, the tracker hold the position when the object disappear in the current frame. Based on the history of object trajectory, two locations are estimated and neural classifier is tested on these locations to detect the object. If the object is found then the tracking is resumed. Otherwise, the tracker is held at the current position and wait for the next frame. A sample tracking results on complete occlusion is shown in Fig. 15. From the figure, we can see that the object completely disappears in frame 15(b) and reappear, at frame 15(c). During the complete occlusion, the object location is held at the current object position, which is marked in red color. The tracker detects the object at 15(c) and then tracking is resumed.



(a) (b)

(c) (d)

Figure 15. Complete Occlusion: Tracking results for video sequence in which the woman is going behind the pole.

There are several limitations to the proposed approach. First, the current feature space selected is simple color features and intensity. These features are not sufficient to differentiate the object from the background in low contrast video sequences. Enhancing the feature space might help the learning system to differentiate the object from the background. Second, the pruning neurons are used to remove previous learnt parameters to handle the change in appearance and illumination. The rate at which pruning takes place affects the tracking accuracy. The pruning window is determined based on a number of object pixels in the current window, which is a somewhat ad hoc solution to alleviate the tracking accuracy. Finally, the proposed tracker does no include the spatial information, thus making the problem more difficult than it should be.

In the next year, we will try to fix the limitations and also extend the proposed learning based tracker to handle multiple objects.

6.2.6. Human Gesture Recognition

Participants: Mohamed Bécha Kaâniche, François Brémond, Monique Thonnat.

Human Gesture Recognition has several applications in human-computer interaction and video monitoring. The main issue is to get a fine enough description of body parts. Several techniques have been proposed such as body sensors, computational vision using one or more cameras. In this work we are interested in recognizing gestures (e.g. hand raising) and more generally short actions (e.g. falling, bending) accomplished by one or more individual in video sequence captured by one camera. We focus on full-body gestures and specially their categorisation into normal gestures and abnormal gestures for use in home care applications. The reference [43] describes a database of full-body gestures that we want to recognize automatically.

We have developed an algorithm of feature point tracking based on Kanade-Lucas-Tomasi (KLT) tracker in order to detect the motion of the human body parts (e.g. hand, foot, arm). Traditional KLT implementations are usually offline, requiring an important processing time and cannot distinguish the objects in the image. We have designed an online KLT tracking-based algorithm that reduces the processing time needed to track features by aborting consistency checking (using similarity or affine mapping) traditionally used to ensure a more reliable track [49]. Instead, we have improved the time-processing and generalized the tracking process by modifying the three following sub-processes:

- Gradient Generator: the default gradient generator used for traditional KLT trackers is the Derivative of Gaussian Gradient generator. Nevertheless, it generates a computational overhead due to a large kernel for convolution (typically 7x7 DoG operator). Actually, we can use a more simple gradient generator (e.g. a sobel 3x3 operator) to reduce processing overhead.
- Feature Selector: the Shi-Tomasi corner detector is usually associated to the KLT trackers. However, the features selected by this detector are often unstable during the tracking process. Consistency checking is needed to improve stability but it is time expensive. Rather than using consistency checking, a solution consists in using a more consistent corner detector (e.g. FAST corner detector).
- Weights for the Harris matrix and the error vector: the reference [50] introduces two weighting functions: Gaussian for blurred edge and Laplacian of Gaussian for sharp edge. We have introduced the product of Gaussian weight function which, according to preliminary tests, provides a compromise between the Gaussian and Laplacian of Gaussian weighting functions while giving better results than the uniform weight function used in standard KLT implementations.

Moreover, we have added a short-term tracker and a long-term tracker which detect the motion of the human body parts (e.g. hand, foot, arm) by the calculation of coherent motion regions and associate them to silhouette information.

Currently, we are testing the KLT-based feature tracking algorithm on automatically generated video sequences in order to determine the spatial frequency (in pixel/frame) of motion beyond it the tracker fails. In the next step, we have to test it on real video data to determine the robusteness of our algorithm with different configurations (especially for choosing the feature selector and weight functions) depending on different scenes. For future work, we plan to use the global information on the image to identify the human behaviour by applying different learning methods: Support Vector Machine, Dual-State Hidden Markov Model or Neural Networks.

6.2.7. Human Posture Recognition

Participants: Bernard Boulay, François Brémond, Monique Thonnat.

In previous work, we have proposed a human posture recognition approach in video sequence. The approach consists in generating silhouettes with a 3D human model for different postures and orientations. The generated silhouettes are then compared with the detected one to determine the posture. In the following, this treatment part is called ”3D generator part“. More details about the approach can be found in the thesis manuscript [16].

Two main limitations of the proposed approach were identified: the limitation in the quantity of postures and the computation time.
The ”3D generator part“ was designed to overcome the first limitation (quantity of postures). This part is now completely generic:

- new postures can be added. For instance, the standing posture with arms carrying an object and the slumping posture (figure 16) were added for the Gerhome project (section 6.2.9).
- new 3D human model can be added by defining other geometric primitives to model the body parts.
- new techniques to represent and compare silhouettes can be integrated.

Figure 16. New postures added in the set of interesting postures.

The other part of the work was to integrate the proposed human posture recognition approach in the video interpretation platform VSIP. This integration takes into account the second limitation of the proposed approach: the computing time. We decide to distribute on two different computers the 3D computation (the generation of the 3D human model silhouettes and the comparison with the detected silhouette) from the rest of the approach (the interpretation: decision about the recognised posture). The retained solution was to use a socket to communicate between the two processes (the 3D part and the interpretation). The recognition process takes naturally place in the interpretation platform. The main advantages for using a socket are:

- the possibility to physically distribute the processes on different computers. For instance the 3D computation can be done on a distant powerful computer, whereas the interpretation can be done near the camera.
- a real-time computation is possible since the data which is sent through the et on the network is minimal. Typically, the detected silhouette and a list of postures and orientations are exchanged from the interpretation platform to the 3D generator part. Then, the results of the silhouettes comparison (the scores) are exchanged from the 3D generator part to the interpretation platform.

We add the capability for the interpretation platform to decide thanks to tracking information which postures and orientations have to be studied. Thanks to this capability, only few silhouettes are generated based on the 3D human model. Now, less than 20 silhouettes are generated in average instead of 100 silhouettes before.

Few problems still have to be solved. We have to find a solution in case of network failure. Moreover, what the interpretation platform should do if it does not obtain the expected scores? We also want to propose a simplest 3D human model (less realistic) but easier to adapt to the observed person. Moreover, we want to take into account the form of the observed person as well as her/his appearance (colors, textures) in the human posture recognition process.

6.2.8. 3D Visualisation Tool

Participants: Bernard Boulay, Nadia Zouba, François Brémond, Monique Thonnat.

A tool to friendly visualise recognised events is necessary. Such a tool is useful for a demonstration purpose. It can also be used for debugging. For example, we can verify the coherence of a proposed scenario by visualising it. A 3D visualisation tool should display a 3D scene environment, mobile objects (usually persons) and recognised events.

We have developed a prototype of a 3D visualisation tool. We have proposed a 3D engine based on OpenGL to display the 3D scene environment. Each contextual object observable in the scene is manually modelled with 3D colored and textured parallelepipeds (floor, walls, table, cupboard ...). A specific property is associated to the objects which can have interaction with people evolving in the scene (e.g. microwave oven, fridge...). These objects are then highlighted as soon as a detected event involves these objects. A 3D human model can be displayed with the recognised posture at the detected 3D position. Finally, the different recognised events are displayed as overlay in the 3D virtual scene: the location of the detected person, the involved sensors (video camera or other sensors) and the current detected activity. The tool takes as input the video processing results obtained by VSIP. An illustration of a 3D virtual scene for the Gerhome project is shown in section 6.2.9.

The next step is to simplify the modelling stage of the 3D scene environment. In particular, we are planning to build a context file format (e.g. size, location, colors of the contextual objects) which can be used for both interpretation and visualisation purpose.

6.2.9. Multisensor Fusion for Monitoring Activities of Daily Living (ADLs) of Elderly People

Participants: Nadia Zouba, Bernard Boulay, François Brémond, Monique Thonnat.

Introduction: *Medical professionals believe that one of the best ways to detect emerging physical and mental health problems (before it becomes critical - particularly for the elderly) is to look for changes in the activities of daily living (ADLs). Typical ADLs are sleeping, food preparation, eating, housekeeping, bathing or showering, dressing, using the toilet, doing laundry, and managing medications.*

Objective: The objective of this work is to monitor activities of daily living of elderly people by using ambient sensors technologies in order to maintain the autonomy of the elderly and to evaluate their degree of brittleness. The goal is to enable elderly people to live longer in their preferred environment (their own house), to enhance their quality of life and to reduce costs for public health systems. In this framework we have proposed a monitoring system of the daily activities of elderly people by using video cameras installed in the house and environmental sensors attached to house furnishings without disturbing the daily behavior of the residents.

Approach: Our approach consists in collecting data on the home resident to build up a "normal" profile of the elderly person daily activity patterns (go to bed, use the refrigerator, use microwave). The proposed system exploits three major sources of knowledge: the video cameras used to detect and track mobile objects (mostly people) evolving in the scene, the environmental sensors used to collect information about the interactions with the objects in the environment, and the a priori knowledge on frailty scenarios predefined by medical experts.

More details about the approach are described in [35].

In this work, we have proposed a generic sensor model (GSM) for healthcare applications. The goal is to be able to seamlessly integrate a new sensor and link it to the primitive events defined in the system through an event ontology to be used as building blocks by users (e.g. doctors) to define their scenarios of interest. This sensor model describes how the environmental generates measurement on activities. This model can be applied to a large class of sensors including video cameras. We have defined different attributes which characterize the sensors which can be used in healthcare applications.

The important attributes we have defined for a GSM are:

- **Time-stamp**: registers the moment when the event was emitted;
- **Periodicity**: is the duration between each measurement (i.e. 1/frequency);

- **Range**: specifies the limits of the sensor measurement (i.e. minimum to maximum);
- **Uncertainty**: represents the confidence on the sensor measurement. Its value depends on the disturbances occuring on the sensor that are inherited from its environment and its physical limitations;
- **3D position**: indicates the x, y, z 3D coordinates of the sensor;
- **Type**: represents the different type of sensors (e.g. contact sensor);
- **Measure**: is the value provided by the sensor (e.g. “O“: Open, ”C“: Close);
- **Measure type**: represents the class of information provided by the sensor (e.g. temperature, presence);
- **Measure unit**: represents the unit of each measure (e.g. degree celsius);
- **Associated equipment**: represents the equipment where the sensors are installed (e.g. drawer, cupboard).

The proposed system consists of 3 steps: data acquisition, data fusion, and activity recognition.

- **Data acquisition:** It consists in providing the data acquired by the selected sensors already cited. These sensors emit primitive events corresponding to information about the environment. We defined the following format for the sensor output: TimeStamp-SensorID-SensorUnit-SensorLocation-SensorValue-Uncertainty.
- **Data fusion:** It consists in to combining information from multiple sensors and diverses sources. Data fusion can be classified into three categories: low-level fusion (signal), medium-level fusion (features) and high-level fusion (events). In this work we have choosen to perform fusion at the high-level by combining video events and environmental events and by using the event description language (EDL) developed in the Orion team.
- **Activity recognition:** To recognize an event composed of two (or one) sub-events, the recognition algorithm selects a set of physical objects matching the physical object variables of the event model. The event model contains the list of physical objects involved in the primitive event. The activity recognition algorithm is based on the method described in [15].

Experimentation: In order to cope with the different research and domain challenges in parallel, we have set up an experimental laboratory to analyze and evaluate our monitoring system. The experimental Gerhome laboratory (http://gerhome.cstb.fr/) we used is located at the CSTB -Sophia Antipolis. (cf. Figure 17).

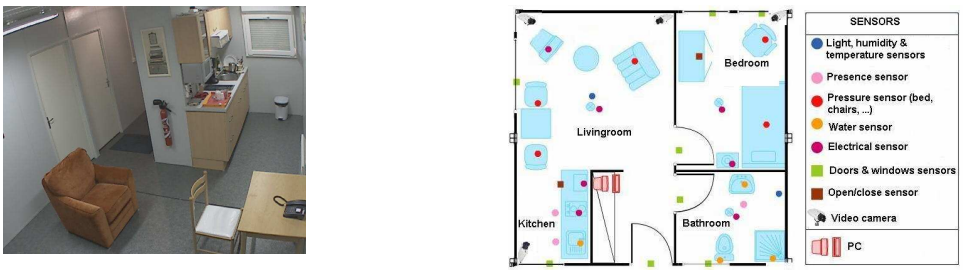


Figure 17. Gerhome laboratory and position of the sensors

Experimental results: To validate our work, we studied a range of activities that may be useful input to a home health monitoring system and we have tested it on Gerhome homecare laboratory. For example: using microwave, using fridge, prepare meal.

When the system correctly claimed an activity occurred, it scored a true positive (TP); an incorrect claim scored a false positive (FP). If an activity occurred and the system did not report it, it scored a false negative (FN). Table 1 shows the results for each ADLs. We then used two standard metrics to summarize the system effectiveness: the precision and the sensitivity. Precision is the probability that a given inference about that activity is correct: TP/(TP + FP). Sensitivity is the probability that the system will correctly infer a given true activity: TP/(TP + FN).

In table 1 videos represent the number of videos we have acquired and events represent the number of events occurring in these videos.

Table 1. Experimental results

Activity	**# Videos**	**# Events**	**TP**	**FN**	**FP**	**Precision**	**Sensitivity**
In the kitchen	10	45	40	5	0	1	0.888
In the livingroom	10	35	40	0	5	0.888	1
Open microwave	8	15	15	0	0	1	1
Open fridge	8	24	24	0	0	1	1
Open cupboard	8	30	30	0	0	1	1
Prepare meal	8	3	3	0	0	1	1

The primitive states ”In the kitchen“ and ”In the livingroom“ are well recognized by video cameras. The primitive events ”Open / cupboard / microwave / fridge“ are correctly recognized by the video and the contact sensors. These events define the composite events such as ”using microwave“, ”using fridge” and “using cupboards”, which defined the ”prepare meal“ scenario. This scenario is well recognized by the system by using video sensors and contact sensors.

We modeled the Gerhome laboratory with the 3D visualisation tool described in section 6.2.8. The different objects (cupboard, armchair, floor, walls, etc.) are textured from the scene images.
After that, we have validated and visualised the recognised events with the 3D visualisation tool. For instance in figure 18, the person is recognised with the posture ”standing with one arm up“, located in the kitchen and using the microwave.

Figure 18. Visualisation of the recognised events in the Gerhome context.

Future work: We will use this recognition process of everyday activities to compute on a long term period (one month) the behavior profile of the observed person and we will detect any pathologic evolution of this profile. Moreover, we propose to incorporate new sensors in order to recognize more complex activities and to add the data uncertainty on sensor measurement.

6.2.10. Tracking and Ontology-Based Event Detection for Knowledge Discovery

Participants: Etienne Corvée, José Luis Patino Vilchis, François Brémond, Monique Thonnat.

At the end of the previous year, the VSIP platform developed within the ORION group was tested with several short videos taken mainly in Torino underground stations. At this stage an early version of the ontology-based event detector was integrated for the CARETAKER project. Few simple and noisy events could be detected from the videos.

Throughout the year 2007, a lot of efforts was put into adapting the vision platform in order to obtain a better tracking algorithm and hence a better input quality for the event detection algorithm. The module in charge of tracking objects is referred to as the 'Long Term Tracker'. The detected and classified mobile objects of interest (such as a person, luggage, crowd or a group of persons). We have contributed to the writing of the deliverable [38] in which we describe how we detect very large objects in a scene and classify them as crowd.

The Long Trem Tracker module produces objects called 'individuals' resulting from the filtering of multiple possible trajectories that mobile objects can take over several successive frames (10 frames in this application). The tracking algorithm is a complex structure of sub algorithms which handles objects features and all their possible trajectories. Filtering of the trajectories was also performed so that relatively reliable tracked individuals can be obtained. The efficiency of the Long Term Tracker in obtaining the right tracks directly depends on the quality of the detected objects. Mis-tracks (ID shifts) or missing tracks always occur in the scenes with rate depending on the video content (for example, these problems occur more often in videos where several persons occlude each other). This algorithm's efficiency also depends on the content of this video information. Dense information on very long processed videos sometimes makes the system unstable. The system was tested throughout Summer 2007 with large datasets of videos recorded during June in Rome. We successfully tracked numerous objects (about 34 000) and detected many simple events (about 235 000) during 2 times 5 hours and half which results were analysed by off-line data mining algorithms. Partners who benefit from our contribution in tracking mutiple objects have written the deliverable document [39].

For the time being, we are mainly focusing our work on improving our event detection in the areas of the subway where individuals interact with each other, with contextual objects (such as dispensing ticket machines, staff office and validating ticket machines) and with specific zones of interest (such as entry and exit zones). Therefore, time was spent in refining the simple events that we can reliably detect depending on the quality of the trajectories given by the long term tracker.

It was decided that the partners involved in the CARETAKER project share multi-media information via the reading and writing of XML files from a database. Thus, we have written codes to perform these tasks on a SQL database for integration purpose. We managed to test our systems along with the vision platform which was continuously and simultaneously reading jpeg images, detecting and classifying objects, tracking individuals and detecting simple events in the observed underground scenes. Two hours of data were put in the database before being read by off-line algorithms. The overall results were published in [29], [31], [24], [30].

At the end of this year 2007, we are working on adapting the vision platform so it can read images from a RTP API which continuously extracts jpeg images from 'mpeg4' format video files provided by the underground cameras.

We will go on working in improving the platform and test it on the many videos of interest already acquired in the Italian subway stations. We also aim to study the obtained trajectories of people in order to finalise a version of the 'Global Tracker' algorithm which is being designed to filter out erroneous trajectories and to force mobile objects to enter a scene from an entry zone and to exit it from an exit zone. We are also planning to obtain more complex events given the fact that trajectories are better filtered. For example, vandalism scenarios or people jumping over turnstiles could be detected as complex events. These events are very interesting for the subway station managers.

As we obtain a large number of trajectory results from people evolving in the scene viewed by the subway cameras, we can self evaluate the quality of these results in terms of noisiness thanks to the data mining algorithm. Once the trajectories are loaded into the off-line system, the statistical results allow us to display the percentage of small to very small trajectories in terms of travel duration of the corresponding persons. For example, it was measured in the Torino station that during 2 hours of recordings, about 32 percent of people spend at least one minute in the Hall where the ticket vending and validating machines are. Moreover a similar percentage of 34 was measured for people spending less than 3 seconds in the Hall. After visualising the trajectories corresponding to these small travel durations, we could conclude that these trajectories were noise as they were mainly due to mis-tracks (change of ID or loss of ID) occurring during occlusion. This auto evaluation achieved at the off-line level will be done for every video we will process.

6.2.11. Unsupervised Behavior Learning and Recognition

Participants: José Luis Patino Vilchis, Etienne Corvée, François Brémond, Monique Thonnat.

In this work we analyse how video information can be processed with the ultimate aim to achieve knowledge discovery of people activity in the video. The proposed approach assumes that a previous tracking system has been able to identify the objects of interest in the video. Then knowledge discovery can be achieved at three different stages: 1) Object statistics from long term analysis 2) Clustering of trajectories 3) Clustering of activity information.

This research has been done in the framework of the CARETAKER project. It aims at studying, developing and assessing multimedia knowledge-based content analysis, knowledge extraction components and metadata management sub-systems in the context of automated situation awareness, diagnosis and decision support. Currently it is being tested on large underground video recordings (GTT metro, Torino, Italy and ATAC metro, Roma, Italy).

1) Object statistics from long term analysis. In order to have a clear and compact representation of the human activity evolving on the video, we compute three semantic tables: mobile objects table, events table and contextual objects table (contextual objects are parts of the empty scene model corresponding to the static environment, for instance, an escalator, an equipment). The proposed representation supports a rich set of spatial topological and temporal relations and captures not only quantitative properties but also higher semantic concepts. The mobile objects table aims at characterising a mobile object by a feature vector which includes the variables 'type' (the class the object belongs to: Person, Group, Crowd or Luggage), 'start' (time the object is first seen), 'duration' (time in which the object is observed) and 'significant event' (this is calculated as the most frequent event related to the mobile object). Similarly, the event table and contextual table give information on the events and contextual objects respectively. The detailed description of all 3 tables can be found in [29]. Statistical information can be obtained from the mobile objects and the contextual objects as well as their interactions. This is a major information source for the end-user. For instance, on large metro video recordings, there is spatial and temporal information on the use of contextual objects. This analysis is detailed in [31]. We have developed a prototype for a graphical off-line analysis tool where the end users select a period of recording time, which they want to monitor and obtain back information such as the most frequent and most rare occurring event, the mean time of use of a contextual object and its use over time.

2) Clustering of trajectories. Recently it has been shown that the analysis of mobile objects'motion detected in the video can give meaningful behavioural information [48], [47], [40]. A people trajectory is made of a dataset of points [xi(t),yi(t)]; whose length is not equal for all objects as the time they spend in the scene is variable. We have analysed two key points of the x and y time series: the beginning and the end, [xi(1),yi(1)] and [xi(end),yi(end)] as they define where the object is coming from and where it is going to. Additionally, we include the directional information given as [cos(Θ), sin(Θ)], where Θ is the angle which defines the vector joining [xi(1),yi(1)] and [xi(end),yi(end)]. We feed the feature vector formed by these six elements to a hierarchical clustering algorithm. (Object trajectories with the minimum distance are clustered together. The final number of clusters is set manually.) This work is detailed in [29], [24]. Figure 19 shows an example of trajectories clustered together after running the agglomerative algorithm on 2h00 of video on the Roma station.

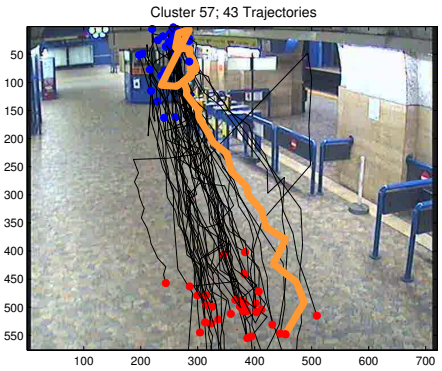


Figure 19. Trajectory cluster. People move from south doors to the office. Red points are entry points, blue points are exit points.

Some knowledge inferred from the analysis of trajectories is for instance:

- 64% of people are going directly to the gates without stopping at the ticket machine;
- 70% of people are coming from north entrance.

3) Clustering of activity information. Once all statistical measures of the activities in the scene have been computed and the prototype trajectories have been found, we aim at discovering complex relationships that may exist between mobile objects themselves, and between mobile objects and contextual objects in the scene. For this task we run a new agglomerative clustering procedure where the data set is the entire mobile object table itself. Each record of the table is thus defined with five features inluding the prototype trajectory that corresponds to each mobile object. It must be noticed that for this clustering process, the set of features contains numeric (for instance the start time and duration time of an object) and symbolic values (for instance, the object type and the significant event) opposed to the clustering of trajectories where all features are numeric. In order to apply the agglomerative clustering algorithm, we have defined a specific metric for the symbolic values. The description of the metrics and the clusters obtained on 45 min of video on the Torino station are presented in [30].

Future work includes analysing a richer set of features to improve the clustering of trajectories. We will also work to improve the semantic distances we have implemented such that better relations can be extracted.

6.2.12. ETISEO, performance evaluation for video surveillance systems

Participants: Anh Tuan Nghiem, Valéry Valentin, François Brémond, Monique Thonnat.

ETISEO was a two year program on performance evaluation for video surveillance systems which ended in December 2006. After the final stage of ETISEO, we have collected the evaluation results of the 16 participants. Based on these results we have highlighted ETISEO advantages as well as its limitations to improve future evaluation projects. In particular, the evaluation metrics in ETISEO have been analyzed to clarify their characteristics, to define their conditions of use, one of the features missed by many other evaluation projects.

The main advantages of ETISEO are the two ontologies, the two-phasis evaluation process, the selection of challenging video sequences associated with ground truth and the definition of priority level, for representative video sequences.

First, in ETISEO, two ontologies have been defined to facilitate the communication among the participants. The first one, used by developers, describes technical concepts (e.g. blobs, individual trajectory) as well as concepts for evaluation. The second one, used by end-users, describes concepts of application domains (e.g. event of opening a door). Two phases of evaluation have been set up. During the first phase, the participants were able to try their algorithms on the data, to try the evaluation tools, and to give feedback to improve the evaluation process. The second phase was the final evaluation consisting in the same tasks with new data and a shorter duration.

In ETISEO, the video processing algorithms of the 16 participants have been evaluated in challenging situations for the five video processing tasks (up to the recognition of events of interest). To have comparable results, we have defined priority levels for representative video sequences. As a result, each representative video sequence has been processed by at least five algorithms.

One of the important contributions of ETISEO has been the metric analysis explaining how to get the best evaluation from the metrics. Thus, we have classified the metrics into 3 main types:

- Main metrics: These metrics measure the general trend of algorithm performance. In other words, if algorithm A is better than algorithm B according to most criteria, then the evaluation of algorithm A with the main metrics is usually higher or at least equal to the evaluation of algorithm B.
- Complementary metrics: These metrics can be used to reveal minor characteristics of the different algorithms. For example, if we use only the main metrics, the performances of two algorithms for object detection can be similar. In this case, the complementary metrics could help us to find out which one is more precise.
- Noisy metrics: These metrics provide little or no information about algorithm performance. Reliable algorithms may have worst evaluation results and vice versa.

Therefore, to evaluate an algorithm, firstly we should use the main metrics to discover the general trend of that algorithm performance. Then, we will use complementary metrics to detect minor problems of that algorithms.

For each video processing task (mobile object detection, tracking, classification etc) we have defined a main metric and several complementary metrics. We have explained how to use these metrics to evaluate the algorithm performance. The metric analysis has been published in [28]

6.2.13. SERKET – Crowd Behavior Analysis

Participants: Ruihua Ma, Van Thinh Vu, François Brémond, Monique Thonnat, Masaki Yoshimura.

SERKET is a European ITEA project that tackles the issue of security of places and public events by developing an innovative software approach where dispersed data coming from discrete devices are automatically correlated and analysed so as to provide security personnel with the right information at the right time [5]. The project will reached its end by June 2008. The Orion Team of INRIA are involved in various themes such as the architecture specification, audio/video event specification (XML), among others. In particular, we focused on crowd behavior analysis.

Our approach to crowd behavior analysis is based on reliable motion feature extraction and tracking. In crowd images, it is unlikely to distinguish individuals. Flow information is more relevant about the behavior of a crowd. The KLT (Kanade-Lucas-Tomasi) tracker is an algorithm that combines feature point extraction and tracking. Only feature points good to track are extracted, resulting in reliable motion information. To further increase reliability, we apply KLT on a temporal window containing multiple video frames and perform motion vector filtering on the results, obtaining thus reliable motion vectors of feature points. The direction and strength (length) of these motion vectors are used to analyze crowd behavior. Normal crowd movement is the motion along the road direction. Thus the abnormal movements we want to detect are first those not in the case, namely, lateral and opposite movements. Naturally, the main attribute used is direction. Other cases include, for example, panic, fighting, vandalism.

[5] http://www.research.thalesgroup.com/software/cognitive_solutions/Serket/index.html

Last year we obtained preliminary results with the multi-frame KLT tracker. This year, we have further tested the algorithm on new data, particularly on the mockup data collected at the ENP (Ecole Nationale de Police). Numerous scenarios in public places were performed by the students of ENP, including Identity Control (IC), riots and public demonstrations. These data are meant to be used to test the various algorithms developed previously by the SERKET partners as well as to be used in the Demonstrator. Since our focus is on crowd behavior analysis, we have worked on the demonstration mockup data, the goal being detection of abnormal crowd behavior.

6.2.14. The Demonstrator

- **Abnormal Crowd Behaviors**

  In the demonstration mockup video, we need to detect abnormal crowd movements, in particular, those not along the road direction. Specifically, these include: *Left Crowd Movement, Right Crowd Movement, Stopped Crowd Movement* for the dominant crowd motion, *Opposite Movement in Crowd, Strong Lateral Movement (Left/Right)* for secondary motion inside the crowd. Our implementation is based on the VSIP platform and is integrated as a smart sensor. The context knowledge, namely the normal crowd movement direction, as well as Left and Right directions, are defined as constants relative to the road in the scenario model. In order to reduce false alarms, we accumulate temporal evidence on events over several frames: only when an abnormal event is detected consecutively for N frames an alarm is sent. This mechanism is provided by the event recognition engine. Figure 20 shows some results of the crowd movement analysis.

Figure 20. Crowd movement analysis results: an example.

We also used the motion direction information for the detection of abnormal behavior of a group of people in a IC control sequence (figure 21). The bar terrace is defined as a zone in the VSIP context and is used as the zone of interest in the scenario model. We detect two types of abnormal events: *Erratic movement* and *Falling*.

- **Complex Event Processing (CEP) and Vehicle Movement Determination during Demonstrations**

  In addition to the specification of video/audio events in XML, we have also contributed to the SERKET demonstrator that accommodates Complex Event Processing (CEP). As video information is concerned, the event of vehicles moving toward the demonstration crowd at different times should be detected. Note that since the crowd is moving, the *to_crowd* direction depends on the current crowd position and it does not stay static. Therefore the constant “dir_to_crowd” is defined dynamically in the scenario model. In total, we have processed seven video clips, the relevant events are effectively detected and alarms sent via XML to other SERKET components such as the CEP layer.

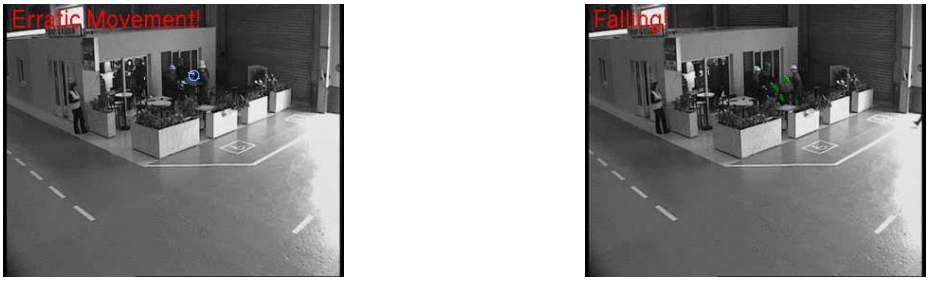


Figure 21. Erratic movement and falling on a bar terrace.

6.2.15. Content-based Video Indexing and Retrieval

Participants: Lan Le Thi, François Brémond, Monique Thonnat.

The increasing number of cameras provides a huge amount of video data. Associating to these video data retrieval facilities become very useful for many purposes and many kinds of staff. When working with video surveillance indexing and retrieval, we are facing five main challenges: (1) the lack of means for finding data from the indexed databases, (2) the lack of approaches working at different abstraction levels, (3) imprecise indexing, (4) incomplete indexing (5) the lack of user-centered search. The objective of our work is to answer these challenges.

Figure 22 gives a general architecture of our approach. This approach is based on an external **Video Analysis module** and on two internal phases: an **indexing phase** and a **retrieval phase**. The external Video Analysis module performs tasks such as mobile object detection, mobile object tracking and event recognition. The results of this module are some Recognized Video Contents. These Recognized Video Contents can be physical objects, trajectories, events, scenarios, etc. So far, we are only using the physical objects and the events but the approach can be extended to other types of Recognized Video Content. The indexing phase takes results from the Video Analysis module as input data. The indexing phase has two main tasks: **feature extraction** and **data indexing**. The feature extraction task performs feature extraction to complete the input data by computing missing features and data indexing using a data model. At present, we use the affine covariant regions and SIFT descriptors to complete the description of the detected mobile objects. The retrieval phase is divided into five main tasks: **query formulation**, **query parsing**, **query matching**, **result ranking** and **result browsing**. In the query formulation task, in order to make the users feel familiar with the query language, we propose a SVSQL (Surveillance Video Structured Query Language) language. In the query, the users can select,in addition to physical objects and events, a global image as example or a region in an image from the database (by the image selection task). In this case, the feature extraction task computes some features in the image example which are used by the query matching task. In the query parsing task, queries built with the proposed language are transmitted to a parser. This parser checks the vocabulary, analyzes the syntax and separates the query into several parts. The query matching task searches in the database the elements that satisfy the query. The obtained results are ranked and returned to the users.

The following example shows a query formulated in the proposed language:

Query: **SELECT** COUNT(o) **FROM** CARE 3 **WHERE**((o:Person) AND(o's Duration > 50))

Description: Count the number of indexed persons that appear in CARE 3 video in more than 50 frames

Output: Number of indexed persons appearing in more than 50 frames

This query counts the number of the indexed persons that appear in the CARE_3 video in more than 50 frames. Four videos from two projects (the CARETAKER project and the AVITRACK project) which have been partially indexed are used to validate the proposed approach. We have analyzed both the query language usage

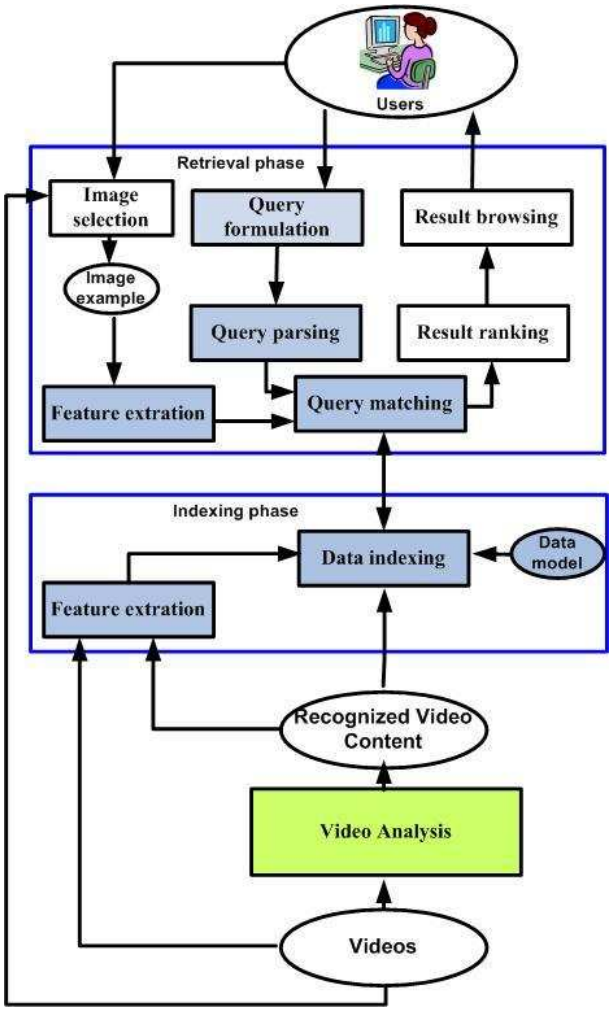


Figure 22. Global architecture of our approach.

and the retrieval results. The richness of the proposed language is demonstrated through more than ten types of queries. The proposed approach enables to answer the most frequent queries users have. The obtained retrieval results are analyzed by the average normalized ranks. The obtained results from both the query language usage and the retrieval results prove that this approach is able to answer the previous challenges. A part of this has been published in [34].

6.2.16. Area of Interest (AoI) system

Participants: Tomi Raty, François Brémond.

The Area of Interest (AoI) system has been specified, designed, implemented and tested in collaboration with INRIA and VTT.

The Area of Interest (AoI) is a distributed scalable video transmission subsystem, for a surveillance system, which concentrates on decrementing the amount of video information transmitted to the end-user equipped with a mobile device. The video information is processed by the Video Surveillance Intelligent Platform (VSIP) to discriminate the essential images of the indoor area under stationary video surveillance. The AoI system analyzes the output of the VSIP images and eXtended Markup Language (XML) image information. The AoI system is able to define and extract the essential information, e.g., a tracked individual, and it transmits only this image to the end-user. First, the AoI transmits the entire image of the indoor area to the mobile device of the end-user. Then, the AoI system transmits only the secluded tracked objects images to the mobile device. The end-user device portrays the scaled portrait images of the targeted object on top of the background image. The AoI system endeavors to decrease the size of the video images transmitted to a smart phone over a wireless network and to retain the comprehension of a tracking situation. The operability of the constructed prototype indicates that this endeavor is attained. The research is based on the constructive method of the related publications and technologies and the results are derived by the implemented AoI system.

The Area of Interest (AoI) system comprises the AoI server and the AoI client. The AoI server resides in a desktop and the AoI client resides in a surveillance personnel mobile device. The AoI server utilizes images and eXtended Markup Language (XML) files, containing image information, that are received from Video Surveillance Intelligent Platform (VSIP). The images are snapshots from a stationary camera of a surveyed indoor area. The XML files contain information about the tracked entities of the surveyed indoor area. This information includes the location of the tracked entity on the snapshot image. During the initial transmission from the AoI server to the AoI client, one entire image of the surveyed indoor area is transmitted. This image is utilized as a background image by the AoI client and it is displayed on the security personnel end-device. After the first image transmission, the AoI server distinguishes the tracked image from each image received from the VSIP. The AoI server extracts the tracked object from each image and the extracted image is transmitted to the AoI client. Upon reception of an extracted image, the AoI client displays the extracted image on the correct location, i.e., where the tracked object actually resides, of the background image. This procedure of extracting the tracked object by the AoI server, transmitting it to the AoI client, and displaying the extracted image at the correct location of the background image is conducted until the AoI system is shut down. By extracting the tracked object from the image and forming an extracted image decreases the amount of information to be transmitted to the end-device.

The intent of the AoI system is to ultimately abate the quantity of information required to be transmitted to the security personnel while retaining all the required information for the security personnel to be fully cognizant about the surveyed indoor area. The operability of the constructed AoI system prototype indicates that this endeavor is attained.

Video surveillance is an important branch in the field of surveillance. With the utilization of advanced video surveillance tools, such as VSIP, it is possible to distinguish images of tracked objects. By abating the amount of image information to be transferred, the images can be transmitted faster to the end users, e.g., surveillance personnel. We have illustrated the implemented design with the AoI system.

6.2.17. Smart Camera

Participants: Valéry Valentin, François Brémond, Monique Thonnat.

Nowadays, sensors are no more passive and they process data according to their environment and/or user requests. We have developed our own smart sensor (called Smart Camera) to match these needs.
The Smart Camera consists of three main components (see Fig. 23):

- Acquisition (Video Server and cameras).
- Processing (VSIP platform).
- Communication (Communication Layer and Web Interface).

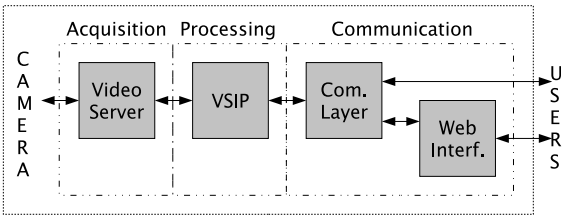


Figure 23. Overview of the Smart Camera

In order to increase the modularity of the Smart Camera, we have built a distributed architecture and separated the Acquisition component from the two others, which can be seen as clients. This means the first one can run on a computer, and send the video stream through a TCP connection to the Processing component, running on the same computer or on another one.

Acquisition Component
Smart Camera Acquisition's component's main task is to help the processing component to access to video data. This data can be a live video stream (currently from an IP Camera), or recorded ones from a local disk (jpeg pictures). It can be accessed by opening a TCP connection on a specific port.
Once the connection to the server is effective, there are two methods to get images: the Push method (used to broadcast live or recorded video streams to the client) and the Pull method (used to send a snapshot -one jpeg image- when requested by the client).

Processing Component
This component is the heart of our Smart Camera, as it performs all the video processing tasks. It connects to the Acquisition component in order to get the video data, and processes them according to the parameter set predefined via the Interface component. It then sends the requested results to the Communication component to make them available to third party components or end users. The Processing component is built on the VSIP platform, in which we added communication features to output result data.

Communication Component
There are two ways to send requests to our Smart Camera and show results obtained from the Smart Camera. The first one, called Communication Layer, is the core of this component. It allows a third party process to connect to and control the Smart Camera. The Communication Layer consists of a set of scripts and processes that can modify parameters used by the Processing component, manage its execution and broadcast the results. On top of it, we have built a Web Interface so that human users can directly access and manage the Smart Camera. The technologies used for this part are Apache for the Web Server, PHP and Java-script for dynamic content generation and CGI scripts (in C/C++ and Perl) for advanced tasks management.

So far, the Smart Camera has been successfully tested in two projects.
In the GERHOME project (shown in section 6.2.9), we use it to process recorded videos and to output analysis results in the same format than other previously installed non video sensors in order to access it and process its results as we do with the other sensors.

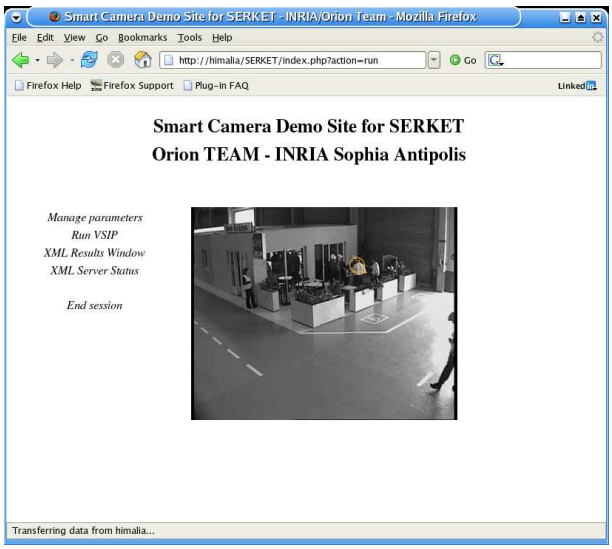


Figure 24. Website Interface of the Smart Camera

In the SERKET project (shown in section 6.2.13), we have used the Smart Camera as a module of the whole architecture. Its main task is to process recorded videos on request, and send the analysis results to connected third party components. The Web Interface was used during the final review of the project (November, 6th 2007) to display and share results obtained in real time by the ORION team during the project.
In the SIC project, we are using the Smart Camera in the same manner than in the SERKET project. It helps us to validate the Smart Sensor concept and enhance the Communication component in order to seamlessly interact with new third party components.
We are working on further improvements including multi camera support and integration of new results and algorithms developped in the VSIP plateform.

6.3. Cognitive Vision Platform

Participants: Vincent Martin, Sabine Moisan, Monique Thonnat.

This year, we have continued our research on semantic interpretation of images with a cognitive vision platform. The platform is based on reasoning (by means of knowledge-based systems), learning and image processing mechanisms as well as ontology-based representation techniques. The platform is used for the detection of plant diseases and for image indexing and retrieval purposes.

6.3.1. Introduction

Image interpretation depends on *a priori* semantic and contextual knowledge. To address the problem of semantic interpretation of images, we rely on some aspects of cognitive vision: knowledge acquisition and representation, reasoning, machine learning and program supervision. We aim at designing a generic and reusable cognitive vision platform dedicated to semantic image understanding. Object recognition and scene understanding are difficult problems; they require a high-level semantic interpretation, a mapping between high level representations of physical objects and image numerical data (i.e. symbol grounding problem), and image processing (i.e. segmentation and feature extraction). To separate the different types of knowledge and the different reasoning strategies involved in the object recognition and scene understanding processes, we propose an architecture based on specialized modules (see Figure 25). It consists of two knowledge-based

modules: (1) one for processing raw images, in order to extract interesting features, (2) another classification module for interpreting these features into higher level terms of objects (or parts of objects) of interest.

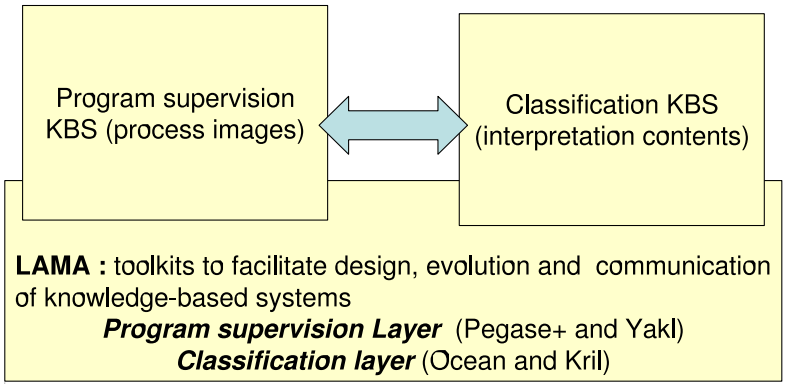


Figure 25. Architecture of the cognitive vision platform

This year, the cognitive vision platform has been enriched by enabling learning faculties at the segmentation level. Previously, segmentation algorithms were manually tuned by an expert in image processing and the dynamic selection relied on a knowledge base written by hand. The same algorithms are now automatically tuned and thus allows an adaptive segmentation of different image, thanks to a training stage. The gain obtained at the segmentation level benefits to the higher level modules of the platform.

6.3.2. Rose Disease Application

Participants: Vincent Martin, Sabine Moisan, Monique Thonnat.

In [19], we propose an original approach for *in situ* early detection of bioagressors, which has been applied to detect mature white flies on rose leaves. To detect biological objects on a complex background, we combined scanner image acquisition, sampling optimization, and advanced cognitive vision. Our cognitive system is composed of extensible knowledge-based systems for image analysis and natural object recognition, coupled with image processing programs and machine learning.

The system has been tested on a database composed of a representative sample of 200 images of scanned rose leaves provided by INRA-URIH (Sophia Antipolis, France). From the data set, 20 images have been used for the learning of the segmentation task (see section 6.3.3) and for the classification task (learning of the possible values for all the numerical descriptors necessary to recognize the visual concepts). From the 180 images composing the test set, 162 contains between zero and five white flies. To assess the quality of our cognitive system, the results have been rated w.r.t. ground truth (manual counts of white flies on the 180 test images). We have also evaluated our system configured with an *ad hoc* segmentation, i.e. using a segmentation algorithm manually chosen and tuned. Globally, the performance of our cognitive system is satisfactory. To fully make use of the results, we can separate the test samples into two classes depicting the most relevant situations. The first class (C_1) represents images with no mature white fly at all and the second class (C_2) represents images with at least one white fly detected. We define the False Positive Rate (FPR) as the rate of over-detection (i.e. images for which the number of detected white flies is greater than the ground truth error bar) and the False Negative Rate (FNR) as the rate of under-detection (i.e. images for which the number of detected white flies is less than the ground truth error bar). The following table summarizes the detection results. The figures represent the mean values of FNR and FPR for class C_1, C_2, and for the whole image test set. We detail False Negative Rate (FNR) and False Positive Rate (FPR) for test images with no white flies (class C_1), at least one white fly (class C_2) and for the whole test set.

	Results for early detection of mature white flies			
	With ad hoc segmentation		With learned segmentation	
Samples	FNR (%)	FPR (%)	FNR (%)	FPR (%)
C_1 (102)	-	9.6	-	3.1
C_2 (60)	24.7	9.0	29.6	2.0
Whole set (162)	9.1	9.4	11.0	2.7

Despite appearances, the robust segmentation of mature white flies on rose leaves is not a trivial task. The variability of the leaves color and texture combined with the semi-transparent nature of the white fly wings and the presence of number of other objects (e.g., white fly eggs, larvae, chemical treatments traces, water drop, etc.) makes the segmentation not so easy. Compared to an *ad hoc* segmentation tuned by hand, our adaptive segmentation achieves better results and thus leads to a better counting precision. Moreover, the semantic segmentation drastically reduces the number of regions by merging the subparts of objects. This technique decreases the computational cost of the system since less regions have to be processed at the interpretation level.

Currently, the platform is able to manage the detection and the counting of only one biological object. Other bioagressors (e.g., greenfly, aphids, etc.) or others stages of development of the white flies (e.g., larvae, eggs) should be treated in order to assess both our adaptive segmentation approach and the platform to a multi-class problem (more than two). To this end, we need to acquire new data (i.e. images with manual segmentation and region annotations) as well as high level knowledge (i.e. description of the objects in terms of visual concepts) for the correct description of new objects.

Finally, the platform is limited by its acquisition system (a flatbed scanner). We plan to overcome such a limitation by using video cameras. Another advantage of video cameras is that they provide temporal information which is of great interest to disambiguate occlusion situations for instance.

6.3.3. Supervised Learning for Adaptive Segmentation

Participants: Vincent Martin, Monique Thonnat.

We have proposed a supervised learning-based methodology for off-line configuration and on-line adaptation of the segmentation (see [18]). The off-line configuration stage requires minimal segmentation knowledge to learn the optimal selection and tuning of segmentation algorithms. In an on-line stage, the learned segmentation knowledge is used to perform an adaptive segmentation of images or videos. This cognitive vision approach to the segmentation issue is thus a contribution for the research in cognitive vision. Indeed, it enables the robustness, adaptation, and re-usability faculties to be fullfiled.

The proposed approach has been implemented and validated on two types of real-world applications: adaptive static image segmentation in a biological application and figure-ground segmentation in a video surveillance application. This work is published in [26] and [22]. The following sections details the main parts of our approach.

6.3.4. A Generic Optimization Procedure

Our optimization procedure automatically extracts the optimal parameters of segmentation algorithms based on a quantitative evaluation of the segmented image quality w.r.t. manual segmentations. The method is independent of the application and of the segmentation algorithms. Only the free parameters to tune with their range values are required. This kind of knowledge is usually provided by the algorithms' authors. The criterion used to evaluate the segmentation quality makes no assumptions on the application nor on the algorithm behaviors. It has been applied to assess segmentation tasks in two applications (a biological application and a video surveillance one). It has also been applied to the Berkeley public segmentation database [42]. Two free-derivative optimization algorithms (a direct search method and a genetic algorithm) have been successfully used to minimize the criteria. In this field, my contribution is a comparative study of the two optimization algorithm performances. Thanks to this study, we have identified two situations: when the number of parameters is up to two, the Simplex provides good results in a minimal number of iterations. When the number of parameters is greater than two, the genetic algorithm should be preferred.

The main difficulty of this supervised learning approach is the manual segmentation of images. This task is tedious, subjective, and time-consuming. User-friendly annotation tools should be used to alleviate users' efforts. The strength of this approach is also dependent on the intrisic performance of the segmentation algorithms. As a consequence, this approach supposes that at least one algorithm is able to perform good results for the target segmentation purpose.

6.3.5. A Strategy for the Algorithm Selection

After that several segmentation algorithms have been optimized on a training image set by using the proposed generic optimization procedure, the next issue is algorithm selection. The goal of this step is to answer to the user's question: "which algorithm is the best adapted to segment my image?". The first part of my strategy consists in identifying different situations in the training image set. A situation, also called a *context*, is represented by a sub-set of images sharing the same global characteristics, such as color distributions. First, an unsupervised clustering algorithm is used to identify these contexts. The second part uses the results of the previous optimization stage to perform a local ranking of the optimized algorithms for each context according to their performance values.

This strategy allows a dynamic control of the segmentation task (i.e. algorithm selection plus optimal parameter setting) without the need of explicit *a priori* knowledge of the application domain or the segmentation algorithms themselves.

It should be noted that this strategy makes several assumptions. First, it supposes that all possible contexts are illustrated in the training image set. Second, this strategy argues that for each identified context, a mean parameter set of the best ranked algorithm exists to deliver good segmentation results.

6.3.6. A Semantic Approach to Image Segmentation

Most of the time, segmentation results provided by bottom-up algorithms are semantically meaningless. I propose a semantic approach to image segmentation where high level region labels help to validate region segmentation results. The region labelling algorithm relies on three steps and makes use of the results of the previous stages (parameter optimization and algorithm selection). In a first step, for each training image, the user is invited to assign semantic labels to regions of manual segmentations according to the application needs. Then, an automatic region label matching is achieved between the regions of the manual segmentation and the regions of the optimized segmentation. Finally, a set of classifiers (SVMs) are trained for each label based on numerical features of regions. The originality of the approach is that each step of the learning process, i.e. feature extraction and SVM training, is optimized in a wrapper scheme so as to maximize the classification performance of the algorithm.

Currently, region features are limited to color and texture information. The method could be improved by also taking into account spatial information, such as the relationships between the different semantic classes of regions.

6.3.7. A Software Implementation of the Methodology

A software implementing this methodology for off-line configuration and on-line adaptation of the segmentation is proposed. Starting from a training image set with the corresponding manual segmentations, the system, via a graphical user interface, is able to:

- extract optimal parameters for six segmentation algorithms (four for static image segmentation and two for video segmentation),
- perform the image cluster decomposition,
- select the best performing algorithm for each identified context,
- annotate the regions with respect to predefined class labels,
- train region classifiers,
- control the segmentation of new images with respect to the learned segmentation knowledge,
- visualize segmentation results.

Finally, by addressing the problem of adaptive image segmentation, we have also addressed underlying problems, such as feature extraction and selection, and segmentation evaluation and mapping between low-level and high-level knowledge. Each of these well-known challenging problems are not easily tractable and still demand to be intensively considered. We have designed our approach (and our software) to be modular and upgradable so as to take advantage of new progresses in these topics.

7. Contracts and Grants with Industry

7.1. Industrial Contracts

The Orion team has strong collaborations with industrial partners for a long time through European projects and national grants. In particular with RATP, STMicroelectronics, Bull, Thales, Solid, Metro of Turin (GTT), Metro of Roma (ATAC) and Keeneo.

8. Other Grants and Activities

8.1. European Projects

ORION team has been involved this year in two European projects: a project on crowd behavior analysis (SERKET) and a new project on multimedia information retrieving.

8.1.1. SERKET Project

SERKET is a European ITEA project in collaboration with THALES R&T FR, THALES Security Syst, CEA, EADS and Bull (France); Atos Origin, INDRA and Universidad de Murcia (Spanish); XT-I, Capvidia, Multitel ABSL, FPMs, ACIC, BARCO, VUB-STRO and VUB-ETRO (Belgium). It has begun at the end of November 2005 and will last 2 years and a half. The main objective of this project is to develop techniques to analyze crowd behaviors and to help in terrorist prevention.

8.1.2. CARETAKER Project

CARETAKER is a new STREP European project that began in March 2006. Its duration is planned for thirty months. The main objective of this project is to discover information in multimedia data. The prime partner is Thales Communications (France) and other partners are: Multitel (Belgium), Kingston University (UK), IDIAP (Switzerland), Roma ATAC Transport Agency (Italy), Metro of Turin (GTT), SOLID software editor for multimedia data basis (Finland), and Brno University of Technology (Czechia). Our team has in charge modeling, recognizing and learning scenarios for frequent or unexpected human activities from both video and audio events.

8.2. International Grants and Activities

Orion is involved in two academic collaborations with ENSI in Tunis (a joint PhD is in progress) and with MICA and University of Hanoi in Vietnam (a joint PhD is in progress).

8.2.1. Joint Partnership with Tunisia

Orion team has been cooperating with ENSI Tunis (Tunisia) for several years. A joint PhD thesis (N. Khayati) dedicated to research on distributed program supervision for medical imaging is in progress. The current test application is an image processing supervision system for osteoporosis detection, in collaboration with physicians and image processing researchers from France and from Tunisia.

We also have co-directed the Master Theses of two Tunisisan students, Raoudha Chebil (ENSI) and Makrem Djebali (ENIT) on topics related to distributed program supervision.

8.2.2. Joint partnership with Vietnam

Orion has been cooperating with the Multimedia research center in Hanoi MICA for several years within the ISERE STIC-Asie project on semantics extraction from multimedia data. Currently we continue through a joint supervision by A. Boucher and M. Thonnat of Lan Le Thi 's PhD on video retrieval (funded by an AUF grant).

8.3. National Grants and Collaborations

Orion Team has five national grants: the first one concerns the implication of the team in a new "pole de compétitivité". We continue both our collaboration with INRA and our collaboration with STmicroelectronics. A collaboration in homecare domain involves a PhD student. Another grant concerns passengers classification in the framework of a PhD thesis funded by RATP.

8.3.1. SYSTEM@TIC SIC Project

Orion is strongly involved in SYSTEM@TIC "pôle de competivité" which is a strategic initiative in security. More precisely a new project (**SIC**) has been accepted last year for funding for 42 months in perimeter security. The industrial partners include Thales, EADS, BULL, SAGEM, Bertin, Trusted Logic.

8.3.2. Cognitive Vision for Biological Organisms

Orion cooperates with INRA URIH at Sophia Antipolis (Paul Boissard) for the feasability study of early detection of plant desease from images.

8.3.3. Intelligent Cameras

This year Orion has continued his strong cooperation with STmicroelectronics. On one hand, in collaboration with STmicroelectronics and Ecole des Mines de Paris at Fontainebleau we study the design of intelligent cameras including image analysis and interpretation capabilities. In particular a PhD thesis (Bernard Boulay) has been defended this year on new algorithms for 3D human posture recognition in real-time for video cameras . On the other hand, a PhD thesis on this topic is on going (Mohamed Becha).

8.3.4. Long-term Monitoring Person at Home

Until 2005, Orion has started a collaboration with CSTB (Centre Scientifique et Technique du Bâtiment) and the Nice City Hospital (Groupe de Recherche sur la Tophicité et le Viellissement) in the GER'HOME project, funded by the General Council 06. GER'HOME project is devoted to experiment and develop techniques that allow long-term monitoring of persons at home. In this project an experimental home is built in Sophia Antipolis and relying on the research of the Orion team concerning unsupervised event learning and recognition, a platform to provide services and to perform experiments should be devised.

8.3.5. Classification of Lateral Forms for Control Access Systems

In the framework of a collaboration with RATP, a PhD thesis (B. Bui) is ongoing on a real-time system for shape recognition. The aim of this work is the development of a system that is able to detect and classify people and objects with very high recognition rate and with real-time constraint.

8.3.6. Semantic Interpretation of 3D sismic images by cognitive vision techniques

A cooperation is ongoing with IFP (French Petrol Institute) and Ecole des Mines de Paris in the framework of a joint supervision by M Thonnat and M. Perrin of Philippe Verney PhD at IFP. The topic is Semantic Interpretation of 3D sismic images by cognitive vision techniques

8.4. Spin off Partner

Keeneo (http://www.keeneo.com) is a spin off of the Orion team which aims at commercialising video surveillance solutions. This company has been created in July 2005 with six co-founders from the Orion team and one external partner.

9. Dissemination

9.1. Scientific Community

- M. Thonnat is a reviewer for the journals PAMI (IEEE Transactions - Pattern Analysis and Machine Intelligence), IEEE Transaction on Multimedia, Pattern Recognition Letters, Image and Vision Computing (IVC), and Eurasip Journal on Image and Video Processing.
- M. Thonnat is Program Chair of ICVS07 International Conference on Vision Systems.
- M. Thonnat is a Program Committee member for the following conferences: CVPR07, ICVS07, ICCV2007, PETS2007, ICVW and ICVS08.
- M. Thonnat is an expert for ANR and Prix These Gilles Kahn (SPECIF).
- M. Thonnat is reviewer for the following theses: Cina Motamed (HDR Univ. Littoral), Nicolas Thome (Univ. Lyon2), Arnaud Renouf (Univ. Caen).
- M. Thonnat is member of the INRIA Evaluation board since 2003.
- M. Thonnat is member of the scientific board of INRIA Sophia Antipolis (bureau du comité des projets) since September 2005.
- M. Thonnat had an invited talk at the Conference Include, Connect, Keep it Safe - ICT for Safe Digital Cities, (Bologna, Italy, the 28th and 29th June 2007) on Research and future perspectives on Intelligent Videosurveillance Systems and at PEA Action (Toulouse, 24 October 07) on “Analyse et interprétation de vidéo pour la reconnaissance d’activités sémantiques”.
- M. Thonnat is member of the EuCognition network of exellence.
- M. Thonnat and F. Brémond are co-founders and scientific advisors of Keeneo, the videosurveillance start-up created to exploit their research results on the VSIP software.
- F. Brémond is an ANR reviewer for the 2007 edition of Project Call: "Concepts Systèmes et Outils pour la Sécurité Globale".
- F. Brémond is reviewer for the journals: Image and Vision Computing Journal, the Machine Vision and Applications Journal, EURASIP Journal on Image and Video Processing, Artificial Intelligence Journal, Medical Engineering & Physics, Computer Vision and Image Understanding, and IEEE Transactions on Multimedia.
- F. Brémond is Program Committee member of ICVS2008: The 5th International Conference on Computer Vision Systems, VIE-2007 Visual Information Engineering, The Seventh IEEE International Workshop on Visual Surveillance VS2007, and AVSS 2007 IEEE International Conference on Advanced Video and Signal based Surveillance.
- F. Brémond is a reviewer for the conferences and workshops: WMVC08: IEEE Workshop on Motion and Video Computing, BMVC2007: 2007 British Machine Vision Conference, ICCV 2007: Eleventh IEEE International Conference on Computer Vision, IWINAC: International Work-conference on the Interplay between Natural and Artificial Computation, and CVPR 2007: CVPR: Computer Vision and Pattern Recognition.
- F. Brémond was chairman of a session at WMVC087.
- F. Brémond had an invited Talk at PETS2007.
- F. Brémond is a reviewer for the PhD defense of Mr Maxime Cottret Toulouse in LAAS.
- Sabine Moisan is a member of the Scientific Council of INRA for Applied Computing and Mathematics (MIA Department).
- Jean-Paul Rigault is a member of AITO, the steering committee for several international conferences including in particular ECOOP. He is also a member of the Administration Board of the Polytechnic Institute of Nice University.

- A. Ressouche is a member of the Inria Cooperation Locales de Recherches (Colors) committee.

9.2. Teaching

- Orion is a hosting team for the master of Computer Science of UNSA.
- Teaching at Master EURECOM on Video Understanding (3h F. Bremond);
- Contribution to a MIG (Module d'Intégration Générale, Ecole des Mines de Paris) Seminar on Formal Method application and managing of student projects (15h A. Ressouche).
- Teaching at Master of Computer Science at EPU (UNSA), Usage of Synchronous languages dedicated tools TP (12h A. Ressouche).
- Jean-Paul Rigault resumed his teaching as a full professor at the Polytechnic School of Nice Sophia Antipolis University (Computer Science Department).

9.3. Thesis

9.3.1. Thesis in progress

- Binh Bui : Conception de techniques d'interprétation 4D et d'apprentissage pour un système autonome de classification et de comptage de personnes, Nice Sophia-Antipolis University.
- Mohamed Bécha Kaâniche : Reconnaissance de gestes à partir de séquences videos, Nice Sophia-Antipolis University.
- Naoufel Khayati : Etude des différentes modalités de distribution d'un système de pilotage de programmes d'imagerie médicale, Nice-Sophia University and Tunis University.
- Lan Le Thi : Semantic-based Approach for Image Indexing and Retrieval, Nice-Sophia University and Hanoi University (Vietnam).
- Anh Tuan Nghiem : Techniques d'apprentissage pour la configuration du processus d'interprétation de scènes, Nice Sophia-Antipolis University.
- Nadia Zouba : Analyse multicapteurs du comportement d'une personne pour la téléassistance médicale à domicile, Nice Sophia-Antipolis University.
- Marcos Zúñiga : Unsupervised Primitive Event Learning and Recognition in Video, Nice-Sophia Antipolis University.

9.3.2. Thesis defended

- Bernard Boulay : Human Posture Recognition for Behaviour Understanding, Nice Sophia Antipolis University (23 January 2007).
- Vincent Martin : Vision cognitive: apprentissage supervisé pour la segmentation d'images et de vidéos, Nice-Sophia Antipolis University (19 December 2007).

10. Bibliography

Major publications by the team in recent years

[1] A. AVANZI, F. BRÉMOND, C. TORNIERI, M. THONNAT. *Design and Assessement of an Intelligent Activity Monitoring Platform*, in "EURASIP Journal on Applied Signal Processing, Special Issue on "Advances in Intelligent Vision Systems: Methods and Applications"", vol. 2005:14, 08 2005, p. 2359-2374.

[2] F. BRÉMOND, M. THONNAT. *Issues of representing context illustrated by video-surveillance applications*, in "International Journal of Human-Computer Studies, Special Issue on Context", vol. 48, 1998, p. 375-391.

[3] N. CHLEQ, F. BRÉMOND, M. THONNAT. *Image Understanding for Prevention of Vandalism in Metro Stations*, in "Advanced Video-based Surveillance Systems", Kluwer A.P. , Hangham, MA, USA, November 1998, p. 108-118.

[4] V. CLÉMENT, M. THONNAT. *A Knowledge-Based Approach to Integration of Image Procedures Processing*, in "CVGIP: Image Understanding", vol. 57, n^{o} 2, March 1993, p. 166–184.

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

[6] S. LIU, P. SAINT-MARC, M. THONNAT, M. BERTHOD. *Feasibility Study of Automatic Identification of Planktonic Foraminifera by Computer Vision*, in "Journal of Foramineferal Research", vol. 26, n^{o} 2, April 1996, p. 113–123.

[7] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", vol. 16, n^{o} 1, December 2004, p. 33-40.

[8] S. 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, April 1998.

[9] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Blocks, a Component Framework with Checking Facilities for Knowledge-Based Systems*, in "Informatica, Special Issue on Component Based Software Development", vol. 25, n^{o} 4, November 2001, p. 501-507.

[10] M. THONNAT, M. GANDELIN. *Un système expert pour la description et le classement automatique de zooplanctons à partir d'images monoculaires*, in "Traitement du signal, spécial I.A.", vol. 9, n^{o} 5, November 1992, p. 373–387.

[11] M. THONNAT, S. MOISAN. *What can Program Supervision do for Software Re-use?*, in "IEE Proceedings - Software Special Issue on Knowledge Modelling for software components reuse", vol. 147, n^{o} 5, 2000.

[12] M. THONNAT. *Vers une vision cognitive: mise en oeuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images.*, Habilitation à diriger les recherches, Université de Nice-Sophia Antipolis, October 2003.

[13] M. THONNAT. *Toward an automatic classification of galaxies*, in "The World of Galaxies", Springer Verlag, 1989, p. 53-74.

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

[15] V. T. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)", 9-15 September 2003.

Year Publications

Doctoral dissertations and Habilitation theses

[16] B. BOULAY. *Human posture recognition for behaviour understanding*, Ph. D. Thesis, Université de Nice-Sophia Antipolis, January 2007.

[17] F. BRÉMOND. *Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition*, Ph. D. Thesis, HDR Université de Nice-Sophia Antipolis, July 2007.

[18] V. MARTIN. *Cognitive Vision: Supervised Learning for Image and Video Segmentation*, Ph. D. Thesis, Université de Nice-Sophia Antipolis, December 2007.

Articles in refereed journals and book chapters

[19] P. BOISSARD, V. MARTIN, S. MOISAN. *A Cognitive Vision Approach to Early Pest Detection in Greenhouse Crops*, in "Int. Journal of Computer and Electronics in Agriculture", To Appear, 2007.

[20] F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT, M. BORG, D. THIRDE, J. FERRYMAN. *Video Understanding for Complex Activity Recognition*, in "Machine Vision and Applications (MVA)", vol. 18, 2007, p. 167-188.

[21] B. GEORIS, F. BRÉMOND, M. THONNAT. *Real-time Control of Videosurveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications (MVA)", vol. 18, 2007, p. 189-205.

[22] V. MARTIN, M. THONNAT. *A Learning Approach for Adaptive Image Segmentation*, in "Scene Reconstruction, Pose Estimation and Tracking", chap. 23, I-Tech Publication, Vienna, Austria, June 2007, p. 431–545.

[23] S. SURESH, N. SUNDARARAJAN, P. SARATCHANDRAN. *Recent Developments in Multi-Category Neural Classifiers*, in "Machine Learning Research", To appear, Nova Publishers, USA, 2007.

Publications in Conferences and Workshops

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

[25] W. LEJOUAD-CHAARI, S. MOISAN, S. SEVESTRE-GHALILA, J.-P. RIGAULT. *Distributed Intelligent Medical Assistant for Osteoporosis Detection*, in "EMBC, 29th International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France", August 2007.

[26] V. MARTIN, M. THONNAT. *A Cognitive Vision Approach to Image Segmentation*, in "Proc. of the 19th IEEE International Conference on Tools with Artificial Intellingence (ICTAI'07), Patras, Greece", vol. 1, IEEE, October 2007, p. 480-487.

[27] A. NGHIEM, F. BRÉMOND, M. THONNAT, R. MA. *New Evaluation Approach for Video Processing Algorithms*, in "IEEE Workshop on Motion and Video Computing(WMVC 2007), Austin Texas USA", February 2007.

[28] A. NGHIEM, F. BRÉMOND, M. THONNAT, V. VALENTIN. *ETISEO, Performance Evaluation for Video Surveillance Systems*, in "IEEE International Conference on Advanced Video and Signal based Surveillance", 21 - 25 September 2007.

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

[30] J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Data mining for activity extraction in video data*, in "8ième Journées Extraction et Gestion des Connaissances EGC 2008, Sophia Antipolis, France", To Appear in February 2008, 2007.

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

[32] S. SURESH. *Risk sensitivity hinge loss function for Sparse Multi-Category Neural Classifier*, in "ICSCIS Proceedings, India", 2007.

[33] L. L. THI, A. BOUCHER, M. THONNAT. *Subtrajectory-Based Video Indexing and Retrieval*, in "The International MultiMedia Modeling Conference (MMM'07), Singapore", January 2007, p. 418-427.

[34] L. L. THI, M. THONNAT, A. BOUCHER, F. BRÉMOND. *A Query Language Combining Object Features and Semantic Events*, in "The 14th International MultiMedia Modeling Conference (MMM'08), Kyoto", To appear, 2007.

[35] N. ZOUBA, F. BRÉMOND, M. THONNAT, V. T. VU. *Multi-sensors Analysis for Everyday Elderly Activity Monitoring*, in "Proceedings of the 4th Int. Conf. SETIT'07: Sciences of Electronic, Technologies ofInformation and Telecommunications, Tunis, Tunisia", March 2007.

Internal Reports

[36] D. GAFFÉ, A. RESSOUCHE, V. ROY. *Modular Compilation of a Synchronous Language*, Technical report, INRIA, 2007.

[37] N. PREVOST. *Réalisation d'une interface graphique d'aide à la spécification de comportements de composants*, Technical report, Polytech'Nice-Sophia, February 2007.

[38] WP4 CARETAKER PARTNERS. *First version of audio/video events recognition*, Technical report, n^{o} D4.1.1p, INRIA-MULTITEL-THALES-BUT-KU, 2007.

[39] WP4 CARETAKER PARTNERS. *First version of single/multiple object tracking*, Technical report, n^{o} D4.2.1, INRIA-MULTITEL-THALES-BUT-KU, 2007.

References in notes

[40] N. ANJUM, A. CAVALLARO. *Single camera calibration for trajectory-based behavior analysis*, in "IEEE conference on advanced video and signal based surveillance, AVSS'07, London, UK", 2007.

[41] P. COUSOT. *Abtract Interpretation Based Formal Methods and Future Challenges*, in "Informatics, 10 Years Back-10 Years Ahead", R. WILHEM (editor), LNCS, 2001, p. 138-156.

[42] C. FOWLKES, D. MARTIN. *The Berkeley Segmentation Dataset and Benchmark*, 2007, http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/.

[43] B.-W. HWANG, S. KIM, S.-W. LEE. *A Full-Body Gesture Database for Automatic Gesture Recognition*, in "IEEE Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition - FGR'06, Southampton, UK", IEEE Computer Society, April 2006, p. 243-248.

[44] K. KIM, T. H. CHALIDABHONGSE, D. HARWOOD, L. DAVIS. *Real-time foreground-background segmentation using codebook model*, in "Real-Time Imaging", vol. 11, n^{o} 3, June 2005, p. 172–185.

[45] S. MOISAN. *Program Supervision:* YAKL *and* PEGASE+ *Reference and User Manual*, Tecnichal Report, n^{o} 5066, INRIA, December 2003, http://hal.inria.fr/inria-00071519.

[46] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *A Behavior Model of Component Frameworks*, Technichal Report, n^{o} 5065, INRIA, December 2003.

[47] A. NAFTEL, S. KHALID. *Classifying spatiotemporal object trajectories using unsupervised learning in the coefficient feature space*, vol. 12, 2006, p. 45–52.

[48] F. PORIKLI. *Learning object trajectory patterns by spectral clustering*, in "IEEE International Conference on Multimedia and Expo, ICME '04, Taipei, Taiwan", vol. 2, 2004, p. 1171–1174.

[49] S. SEGVIC, A. REMAZEILLES, F. CHAUMETTE. *Enhancing the Point Feature Tracker by Adaptive Modelling of the Feature Support*, in "Proceedings of the 9th European Conference on Computer Vision Part II - ECCV'06, Graz, Austria", Springer-Verlag, May 2006, p. 112-124.

[50] M. SINGH, M. MANDAL, A. BASU. *Robust KLT Tracking with Gaussian and Laplacian of Gaussian Weighting Functions*, in "IEEE Proceedings of the 17th International Conference on Pattern Recognition - ICPR'04, Cambridge, UK", IEEE Computer Society, August 2004, p. 661- 664.

[51] C. STAUFFER, W. GRIMSON. *Adaptive Background Mixture Models for Real-time Tracking*, in "Proc. of IEEE Conf. on Computer Vision and Pattern Recognition", 1999, p. 246–252.

[52] M. THONNAT, A. BIJAOUI. *Knowledge-based galaxy classification systems*, in "Knowledge-based systems in astronomy", A. HECK, F. MURTAGH (editors), Lecture Notes in Physics, vol. 329, Springer Verlag, 1989.

[53] M. THONNAT, V. CLÉMENT, J. C. OSSOLA. *Automatic Galaxy classification*, in "Astrophysical Letters and Communication", vol. 31, n^{o} 1-6, 1995, p. 65-72.

[54] R. VINCENT, M. THONNAT, J. OSSOLA. *Program Supervision for Automatic Galaxy Classification*, in "Proc. of the International Conference on Imaging Science, Systems, and Technology CISST'97", June 1997.

[55] M. ZÚNIGA, F. BRÉMOND, M. THONNAT. *Fast and Reliable Object Classification in Video Based on a 3D Generic Model*, in "Proceedings of the International Conference on Visual Information Engineering (VIE2006), Bangalore, India", 26-28 September 2006, p. 433-440.