INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Orion

Intelligent Environments for Problem Solving by Autonomous Systems

Sophia Antipolis

THEME COG

Table of contents

1. Team

Head

Monique Thonnat [DR1 Inria, HdR]

Team Assistant

Catherine Martin

Research Scientists

François Brémond [CR1 Inria]
Sabine Moisan [CR1 Inria (long-term medical leave from April 2005 to April 2006), HdR]
Annie Ressouche [CR1 Inria]

Long Term Invited Professor

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

Technical Staff

Alberto Avanzi [Bull Engineer, up to 31 March 2006]
Nicolas Chleq [IR2 Inria, up to March 2006]
Etienne Corvée [European Project CARETAKER, since April 2006]
Gabriele Davini [European Project SAMSIT project, up to April 2006]
Ruihua Ma [European project SERKET]
Magali Mazière [Cassiopée Project/Crédit Agricole up to April 2006]
José Luis Patino Vilchis [European Project CARETAKER, since September 2006]
Valery Valentin [European project AVITRACK, up to September 2006]
Thinh Van Vu [European project SERKET]

PhD Students

Bernard Boulay [Paca Lab Grant, Nice Sophia-Antipolis University]
Binh Bui [CIFFRE RATP Grant, Nice Sophia-Antipolis University]
Benoit Georis [Louvain Catholic University (UCL) Belgium, up to January 2006]
Mohamed Bécha Kaâniche [Paca Lab Grant, Nice Sophia-Antipolis University, since May 2006]
Naoufel Khayati [STIC-Tunisie grant, 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, from November 2006]
Marcos Zúñiga [CONYCIT Grant, Nice Sophia-Antipolis University]

Intern Students

Patrice Jacq [Paris 6 University, from April to September 2006]
Mohamed Bécha Kaâniche [AVF grant, ENIT Tunis, from October 2005 to March 2006]
Anh Tuan Nghiem [DEPA grant, IFI Hanoi, from April to October 2006]
Shobhit Saxena [ADRET Intership, Indian Institute of Technilogy (New Delhi), from May to August 2006]
Nadia Zouba [Paris 8 University, from April to September 2006]

Short Term Technical Staff

Patrice Jacq [Internal grant, from October to December 2006]

Visitors

Bernd Neumann [FB Informatik dpt, Hamburg University, from September to October 2006]
Claudio Piciarelli [Mathematics and Computer Science dept, Udine University, from June to October 2006]

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 Batiment (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.

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. [44]

3.3. Software Platform for Cognitive Systems

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

Participants: Sabine Moisan, Nicolas Chleq, 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.

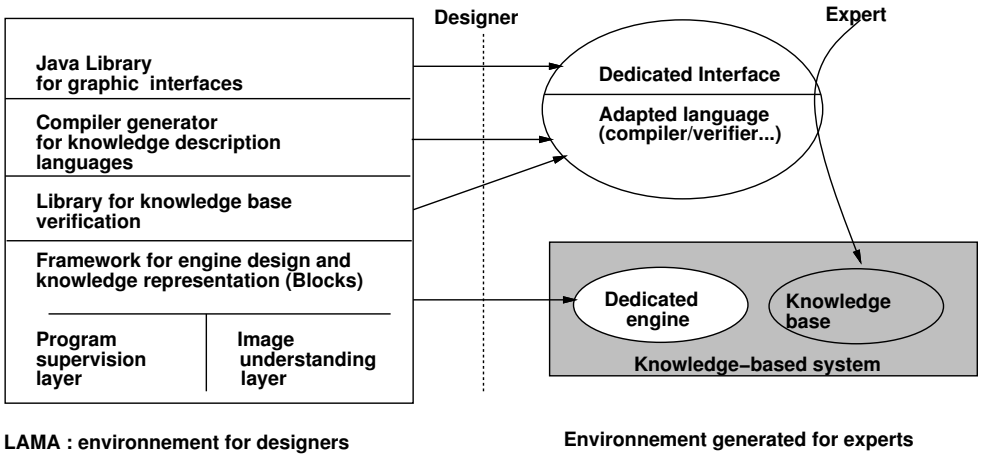


Figure 1. LAMA *architecture and tools for engine design, knowledge base description, verification, and visualization*

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*.

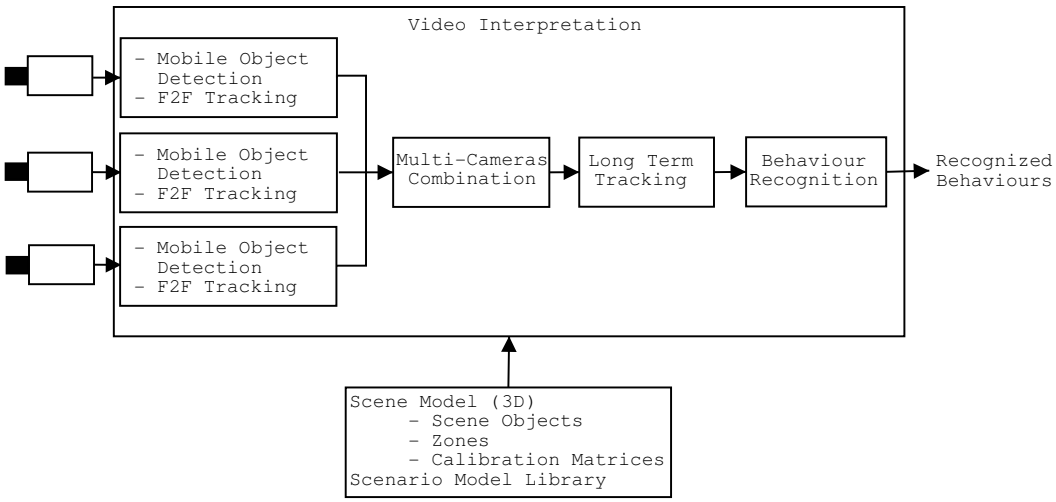


Figure 2. Overview of the interpretation of image sequences.

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 restricted our works to 2D object recognition and 2D static scene understanding. By *cognitive vision*, we refer,

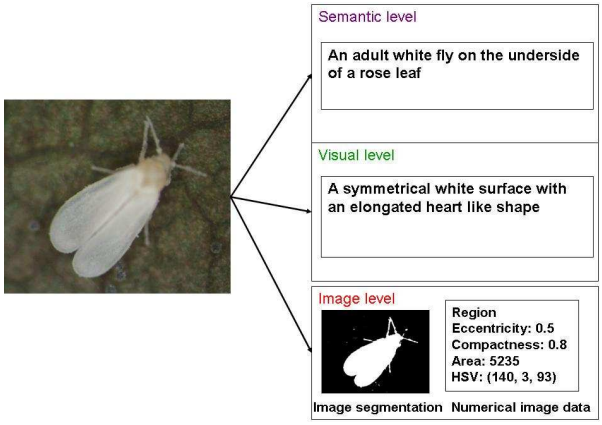


Figure 3. Illustration of the three abstraction levels of data corresponding to the sub-problems of semantic image interpretation. The image is a microscopic biological image.

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

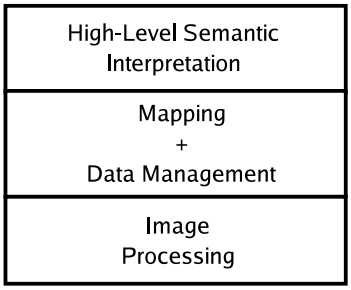


Figure 4. The problem of image interpretation is divided into three sub-tasks.

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] [47]. 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 [48], [49].

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 and AVS-RTPW and more recently, European projects ADVISOR and AVITRACK, 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 (occultation, 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). The new challenge in bank agency monitoring is to handle a cluttered environment and in train car surveillance is to take into account the motion of the cameras.

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

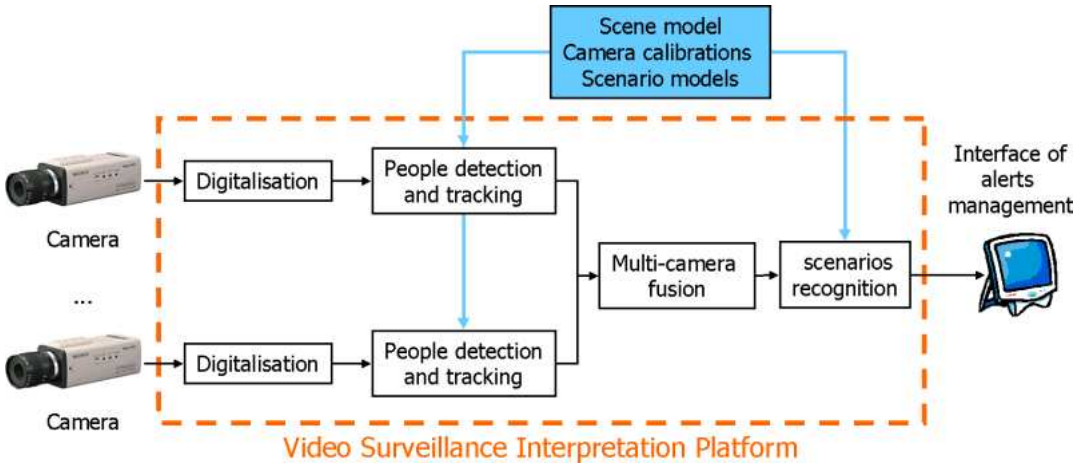


Figure 5. Components of the Video Surveillance Interpretation Platform (VSIP).

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: Nicolas Chleq, Naoufel Khayati, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Monique Thonnat.

This year we have continued to work on the program supervision task for distributed applications, on the classification task for object recognition, and on the verification of framework components. To handle distribution, we have adopted an approach using mobile agents. This year we have tested the Aglets [41] platform for implementing a distributed program supervision scenario. we took the opportunity to embed an inference engine into Java agents. We have also finalized our previous work on model calibration [30]. Finally, a completed PhD work [16] has applied our program supervision engine PEGASE+ *to video surveillance and object tracking.*

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 (e.g., classification together with program supervision).

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 system for medical imaging. We have continued to develop the engine verification toolkit. We rely on formal methods to improve a hybrid language we began to develop last year and devoted to express the code generated by the engine verification toolkit. Finally, we begin to study the modeling of a scenario recognition engine.

6.1.2. Distributed Program Supervision

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

A Program Supervision system consists of three main components: a set of (independent) programs (written in various languages: C, C++, Java, Matlab...); a knowledge base, containing knowledge (description, inputs and outputs, rules of use...) about the use of these programs; a Program Supervision engine which uses the knowledge information to establish a plan to chain the execution of the various programs and evaluate their results (in our case, the engine is called PEGASE+). When distributing such a system over a network, any of these three components may be dispatched onto several nodes.

Collaborating with ENSI Tunis, we are developing a prototype of a multi-agent distributed system [21] in medical imaging, more precisely for osteoporosis detection. This year we have improved the architecture and the scenario, tested the whole system, and speeded up performances (parallelization, replacement of some Matlab procedures by C code...). The communication between the various elements of the distributed system is performed owing to (possibly mobile) agents using the Aglets platform [41]. In the prototype, there are several sorts of agents. Permanent agents have a life-time which is the whole application. Some of them are stationary (remaining in the same node): for instance, they allow the agent system to communicate with various (non-agent) programs, in particular with the PEGASE+ executable and Matlab. Others are mobile, visiting several nodes allowing to exchange data and results between agents and programs. Transient agents have a life time which is shorter than the application one. They are just created on the fly, to execute a particular task, then they disappear. In our case, all transient agents are mobile.

Figure 6 presents an overview of the corresponding agent architecture to implement a simple scenario. The scenario starts by deploying several agents: a Supervisor agent, which will be in charge of the whole session; a Pegase agent to communicate with the supervision engine; and as many execution agents as there are programs to interface with (in this simple scenario, there is just one such agent which interfaces with Matlab, since our application programs are written for this latter system). All these agents are permanent and can be located on different hosts. To execute a request a Solver (mobile) agent is created; it will goes from the Pegase server –where it obtains pieces of the supervision plan– to an execution server to achieve the requested computation, then back again to Pegase... A computation may require remote resources: this is the role of the (transient) Transporter agents to fetch them. When the computation ends, an Evaluator agent may be necessary to assess the result quality either by transporting results and communicating them to the Supervisor or even the User.

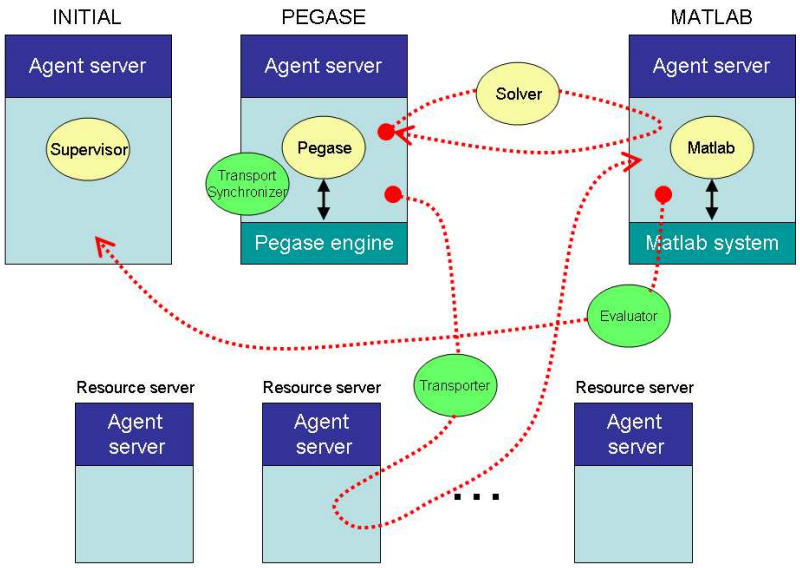


Figure 6. Distributed Multi-Agent Architecture.

Finally, we are in the process of integrating results of the various student projects that were conducted last year. The issues include security and privacy of the distributed system (particularly important for medical applications) as well as the possibility of concurrent and incremental update of knowledge bases.

6.1.3. 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 [45]. 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. Last year, 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.

This year, we have started the development of a graphical interface to express BDL programs and to simulate them with respect to their synchronous semantics. Instead of developing a specific purpose tool, we plan to develop 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. Then, we intend to specialize this generic graphical interface to design BDL programs. Such an approach will allow to have several specializations and to get new graphical interfaces for almost free.

6.1.4. Hybrid Event-driven System Description

Participant: Annie Ressouche.

The aim of this research axis is the specification of controllers listening to external events, mainly resulting from sensors. *Signal processing and sampling of continuous values must take into account sensors values in controllers.* Such controllers are reactive: the evolution of a reactive system is a sequence of reactions to external stimuli, at a speed defined by the environment. It is well recognized that general purpose programming languages are not suited to reactive critical systems, hence the development of dedicated languages. Following this point of view, in collaboration with V. Roy (CMA Ecole des Mines) and D. Gaffé (CNRS and UNSA) we defined a specific purpose language which allows a modular description of control-dominated reactive synchronous systems. The language semantics allows us to compile a program into a boolean equation system. We found a method to compose two sorted equation systems without sorting the whole system again. These results clearly improve compilation efficiency..

This year we have defined a graphical tool to express reactive programs and translate them into internal formats. These formats make it possible to interface both with a built-in simulation tool and with external verification tools. We are also continuing to integrate sensor values into our models.

A strong point of our approach is the definition of a semantics that allows both an efficient modular compilation and the computation of a model for programs as labeled transition systems. A first consequence of such an approach is that program debugging, testing, and validating is made easier. In particular, formal verification is possible with techniques like model checking. Another consequence is that we are able to generate code avoiding the tedious and error-prone task of implementing the code corresponding to a specification. To deal with sensor values and keep program modeling as labeled transition systems, we rely on static analysis and abstract interpretation methods [39]. Abstraction 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 hybrid programs. This last point has been particularly studied in a master thesis by Lionel Daniel [40].

One of the initial motivations for this work was to generate code for components of knowledge-based systems from their behavioral description in BDL (see section 6.1.3). Another issue could be to use this language to generate code dedicated to activity recognition.

6.1.5. Scenario Description

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

Scenarios representation and recognition is a research topic studied for a long time in the Orion project, in the domain of Automatic Video Interpretation. Particularly, in the last years Van-Thinh Vu has proposed a description language to describe human behaviors and an algorithm to recognize temporal scenarios from scenes of a video sequence [15]. Relying on this background, in collaboration with V. Roy (CMA Ecole des Mines), we defined a language to express scenarios in a modular way as composition of automata. Basic scenarios are composed of events, while general scenarios support hierarchy, i.e calls to sub-scenarios, either in a parallel way or in a sequential one. Temporal constraints between scenarios and events can be expressed. This approach was a first attempt to study scenario modeling.

For behavior recognition, as for all automated systems, validation is really a crucial phase and an exhaustive approach of validation is clearly needed. To be trusted behavior recognition must rely on formal methods from the very beginning of its design. Formal methods help to produce a sound code the size and frequency of which can be estimated. Hence, we defined a synchronous model for our scenario language, following the approach of André et al. [35]. This approach leads to automatic recognition of scenarios, based on the "observer" model-checking technique. Moreover, we can use this approach to generate a set of observers to formally verify the temporal constraints between scenarios.

On the other hand, the scenario recognition algorithm defined by V.T Vu is real time, a strong and important characteristic in the domain. Indeed, we need models 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. Therefore, models of scenarios must also take into account the uncertainty of sensor results. This year, we begin to extend our model to address these needs (logic time, real time and uncertainty). Lionel Daniel has started a PhD at CMA (Ecole des Mines) in October on this topic: "Formal methods to model and recognize scenarios", co-directed by V. Roy (Ecole des Mines) and A. Ressouche. This thesis is related to the SECMAR project in the "pole de compétitivité" Mer. The project concerns the security of a sea area and will be an opportunity to apply our scenario modeling.

6.2. Automatic Interpretation of Image Sequences

Participants: Alberto Avanzi, François Brémond, Bernard Boulay, Binh Bui, Etienne Corvée, Gabriele Davani, Florent Fusier, Patrice Jacq, Mohamed Bécha Kaâniche, Ruihua Ma, Magali Mazière, Anh Tuan Nghiem, José Luis Patino Vilchis, Shobit Saxena, Lan Le Thi, Monique Thonnat, Valéry Valentin, 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. The techniques developed have been applied in nine projects: two transfer actions – Keeneo and Telescope3, the European IST project AVITRACK, the European ITEA project SERKET, the SAMSIT Predit project (department of research) and GERHOME project (PACA region), and the three following industrial projects: Intelligent Cameras (STmicroelectonics), PFC (RATP), and CASSIOPEE (FNCA). In addition, we have lead the evaluation project ETISEO which enables to evaluate these techniques and to compare them with other video surveillance algorithms and systems.

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 in order to recognize crowd behaviors in the future. We have extended two methods for object categorization, firstly by combining a high density of sensors, and secondly by taking advantage of the 3D parallelepiped volume of the observed objects. We have also extended classical categorization methods to recognize the shape and behavior of milimetric wasps in a laboratory slide observed by a microscope. We have improved our approach for the

recognition of human postures by optimizing the combination of 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: AVITRACK and CASSIOPEE. In the framework of the ETISEO project, we have a proposed a new 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. Crowd Tracking in Video Sequences

Participants: Shobhit Saxena, François Brémond, Ruihua Ma, Monique Thonnat.

The objective of this work is to perform crowd tracking in video sequences to be able to determine crowd motion parameters which can help in detecting interesting events of crowds in public places.

By performing crowd tracking, we are interested in estimating the speed and direction of crowd motion in a video sequence. This information can be used for building higher level models of crowd behavior. These models can be used to analyze the crowd behavior and detect anomalous events. Crowd tracking generally is often a harder problem to solve as compared to individuals tracking. There are challenges like handling a greater degree of occlusion and inhomogeneous backgrounds.

To estimate crowd velocity in a video sequence, we perform KLT (Kanade Lucas Tomasi) tracking [43] [46] on video frames, looking for interesting feature points in the scene and tracking them over time. From the tracking results we extract meaningful crowd motion indicators which tell us the existence of feature points motion, their speed and direction. We then quantifie these crowd motion indicators based on their directions. This provides us with classes of similar crowd motion indicators (similarity of direction). We then cluster the indicators in each of these classes based on spatial proximity. This gives us the regions where a particular direction of motion is prominent. We also compute the statistical values related to speed for each of these clusters. Thus we are able to obtain the areas where a crowd motion exists, its direction and speed along with its strength. We could observe the robustness of the tracking method and the crowd motion indicators in the results we obtained on various video sequences.

The KLT tracker offers numerous advantages for low-level tracking over other methods. It is independent of the degree of crowd density or camera position. KLT is also easier to use for subsequent analysis as compared to block matching or optical flow since we deal with only the good features and not all the features. The biggest advantage of the KLT tracker is its robustness and stability across multiple frames.

While the results obtained under different scenarios of varying crowd density, varying camera perspective, etc., seem to be satisfactory, this approach can be fine-tuned with better algorithms for direction similarity determination and spatial clustering. We can find out learning relations between crowd motion density and actual crowd density using some knowledge of the context. Future work will include building primitive crowd-events which can then be utilized for behavior analysis.

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

Participants: Binh Bui, François Brémond, Monique Thonnat.

Last year, we have presented a real-time system for shape recognition. This system is a video and multi-sensor platform with a fixed camera observing the mobile objects from the top and lateral sensors observing the side of mobile objects. This system is able to classify the mobile objects evolving in the scene into several expected categories. The key of the recognition method is to compute mobile object properties (i.e. width, height, density of occluded/non-occluded sensors) by making use of the top camera and lateral sensors and then to apply Bayesian classifiers to these properties. A learning phase based on ground truth data is used to train the Bayesian classifiers. Our recognition method has been integrated into an existing access control device used in public transportation (subway) at RATP to improve safety and comfort, to prevent fraud and to count people for statistical matters. The expected categories in this case are mainly "`adult`", "`child`", and "`baggage`" (i.e. suitcases, bag, and backpack).

This year, we have studied a degraded operating mode of the system, i.e., we do not use the top camera for categorizing mobile objects. We have also tested other classification methods with our actual database. Our goal is, on one hand, to compare the performance of both systems (with and without the top camera) and, on the other hand, to make a comparison between our classification method and other methods such as those based on support vector machines (SVM) and neural networks (NN).

We have developed a software component to evaluate the performance of the system in an automatic way. This module takes as input the output of the classification system and the ground truth file and outputs automatically the evaluation file with the true positive rate, the false positive rate and the false negative rate for the whole sequence. Using this component, we have evaluated the performance of both systems (with and without the top camera) and have compared the classification techniques (cf. table 1). About 3500 frames of different expected objects (adult, child, suitcase, bag and backpack) have been used for this evaluation.

Table 1. A comparison between two operating modes: with and without the top camera.

		Full mode (with the top camera)	Mode without the top camera
Mobile objects completely inside the lateral sensor zone	TP	97.6%	95.5%
	FP	3.4%	4.6%
	FN	2.4%	4.5%
Mobile objects partially detected by the lateral sensor zone	TP	95.6%	89.3%
	FP	2.3%	5.3%
	FN	4.4%	10.7%

As we can see in Table 1, inside the lateral sensor zone, the system without the top camera has almost the same performance as the system with the top camera. Thus with only the lateral sensors, our approach is robust enough to distinguish different expected categories. The majority of True Positive difference between two operating modes is due to backpacks carried too high. In our system, the height of lateral sensor zone is limited to about 110 cm. So, the backpacks which are above 110 cm cannot be detected by lateral sensors. Only the top camera is able to detect and classify the backpacks. The top camera also helps to better detect and classify mobile objects that are only partially detected by the lateral sensors. Moreover, this camera allows to detect the events needed for behavior interpretation such as gestures of transport ticket validation of passengers in the scene.

We have also compared our classification method with two other approaches: support vector machine (SVM) and neural network (NN) (cf. table 2). About 4000 frames of different categories have been used as training data set and 3500 other frames have been used as testing data set. The details on how to use the functions and parameters for the SVM and the neural network used for this comparison can be found in [51].

Table 2. A comparison with two well-know classification techniques: support vector machine (SVM) and neural network.

	Bayesian classifiers (used in our system)	Neural network	SVM
Correctly classified instances	97.6%	97.3%	98%
Incorrectly classified instances	2.4%	2.7%	2%

As we can see in Table 2, SVM gives the highest rate of correctly classified instances followed by our method and neural network. However, the difference is minor. We can say that with our actual database, these methods have almost the same performance. Future work will consist in continuing the performance evaluation on more data to be significant as well as to study the failure cases of the system to obtain an operational prototype working in all the situations.

6.2.4. Generic 3D Object Categorization Method

Participants: Marcos Zúñiga, François Brémond, Monique Thonnat.

Binocular visual perception allows human beings to perceive depth of their environment. At the same time, a person can shut one of his/her eyes and still preserve the depth sensation, without loosing too much of precision on depth estimation of the focused object. This capability is a consequence of the interpretation that the brain performs about the new visual information, by associating it to similar environments or objects previously observed, and then concluding on its nature and 3D shape. This means that the brain uses a priori knowledge to conclude about the attributes (e.g. position, dimensions) of an observed object.

Following this idea, we have proposed a new 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, considering a pinhole camera model. For this purpose, we have proposed a simple 3D object model that represents an object as a parallelepiped. The model is described by the parallelepiped dimensions (width, length and height) and orientation in the ground plane of the scene. Also, visual reliability measures of the three estimated dimensions have been proposed to relate to their visibility. These measures have been mainly used to aid posterior phases of the video understanding process, such as dimensional estimation of tracked objects, multi-camera object fusion, and discrimination between visually reliable data from purely estimated data on event detection and learning.

Our approach also tries to cope with several limitations imposed by 2D representations, but keeping their capability of being general models able to describe different objects and still being able to work in real-time.

The proposed classification method uses the 2D moving regions obtained from an image segmentation phase, the perspective matrix of the scene, and predefined 3D parallelepiped models of expected objects in the scene, to find the most likely 3D model of the objects. Then, a merging step is performed to improve the classification performance by assembling 2D moving regions showing a better 3D object likelihood when they are put together. The perspective matrix of the scene is previously obtained from an off-line camera calibration phase, considering the pinhole camera model.

This classification method has shown promising results in object classification. First, the proposed approach has been able to cope with the problems of object position relative to the camera position, object orientation and perspective deformation, with high classification rates. Second, the method has shown its capability of performing at video frame rate (516 objects/sec).

Future work comprises the integration of this method with object tracking techniques and the capability of coping with occlusion situations. The problem of static occlusion (moving object occluded by a static object or image borders) can be treated with the information obtained from the proposed classification approach, but dynamic occlusion (a moving object occluded by another moving object) will require more information that can be obtained from the tracking process.

This approach has been published in the VIE conference [34].

6.2.5. Object shape recognition for Trichogramma activity monitoring

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

We are interested in extending previous work on object shape recognition and to apply activity monitoring techniques to the biology field. *Trichogramma* species are recognized as important biological control agents to substitute pesticides in field crops, forests and fruits. As a parasite of caterpillar pests, it protects several vegetables such as corn, rice and sugarcane. Current studies are focused on analyzing the variations of handling-time and on understanding their foraging mechanisms for screening better agents for biological control and improving their efficiency to control their hosts when they are released in the field. To conduct this work, it is essential to understand the behavior of parasites. Currently, the video sequences of laboratory experiments (see Fig. 8 for sample) are analyzed manually by experts. To handle the hugeness of the video sequences to explore, we are interested in automating this task by using scenario recognition techniques.

Figure 7. Resulting frames for the proposed classification approach. Top frames show the results for people counting application in a locked chamber: an instance of the class PERSON (figure (a)), of the class TWO-PERSONS (figure (b)), and of the class THREE-PERSONS (figure (c)) have been detected. Bottom frames show the results for a parking lot application: in figure (d) an instance of the class VEHICLE have been detected, in figure (e) the same PERSON at two different frames is correctly classified, and in figure (f) an instance of the class PERSON and other of the class VEHICLE is correctly classified at the same frame.

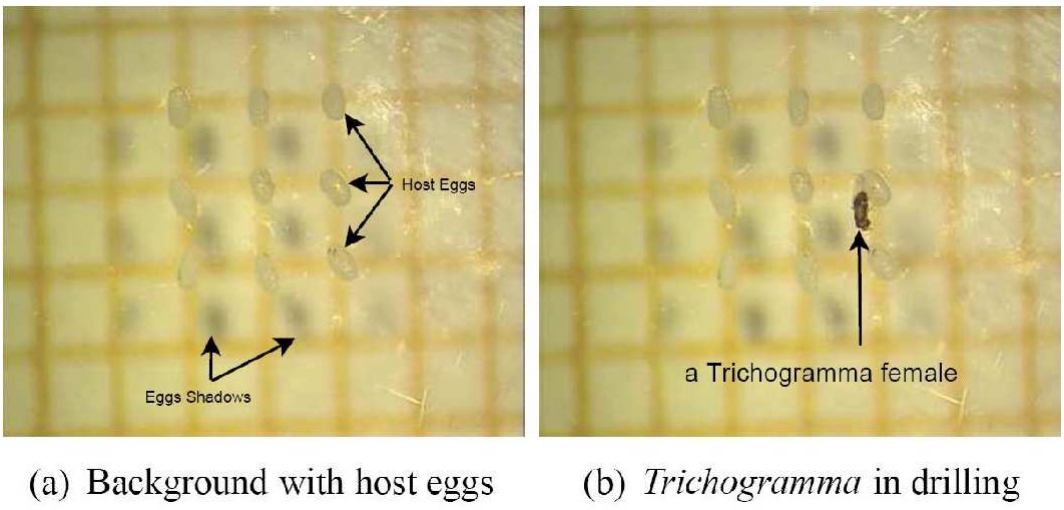


Figure 8. The figure shows the background of a sample image sequence and a sample frame from the video. Image (a) illustrates the nine host eggs and their shadows. Image (b) illustrates a Trichogramma female in drilling phase.

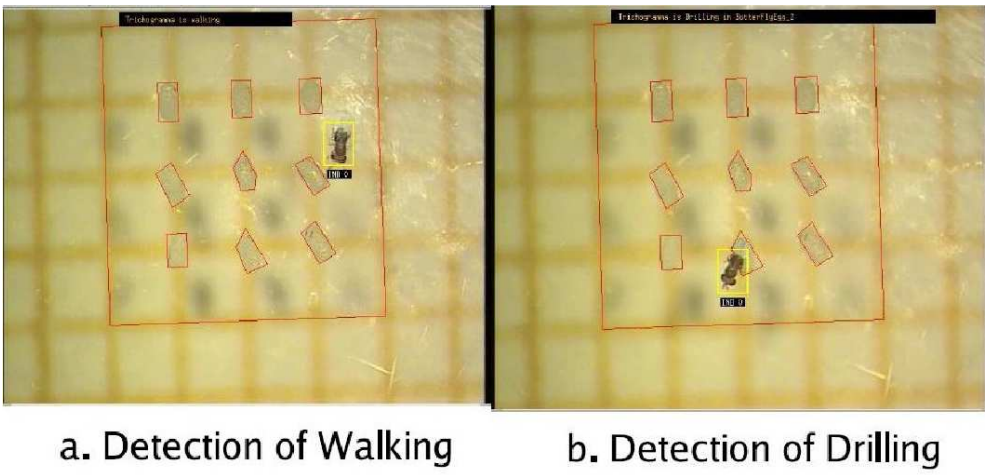


Figure 9. Output of the scenario recognition module.

Traditionally, video understanding is focused on recognizing human activities. Furthermore, behaviors are usually recognized through the study of trajectories and positions of studied objects and using *a priori* knowledge about the scene. This is quite sufficient when we deal with scenes having a large field of view and coarse human activities. However, we often need to compute visual features characterizing the shape of the mobile object (i.e. *Trichogramma*) in order to identify its behavior. In this research, our aim is to adapt an automatic video interpretation system to recognize *Trichogramma* behaviors. The system is composed of a vision module and a scenario recognition module. It takes two types of input: (1) a video stream acquired by camera(s) and (2) *a priori* knowledge concerning scenario models predefined by experts and the 3D geometric and semantic information of the observed environment. The output of the system is a set of recognized scenarios at each instant.

In general words, the system performs through three steps. First, a low-level image processing algorithm subtracts the current frame with the background frame and detects moving regions. Then a tracking algorithm tracks the detected regions and computes their trajectory. Finally, the scenario recognition module identifies the tracked moving regions as mobile objects and interprets the scenarios that are relative to their behaviors.

Our goal is to build the history of six *Trichogramma* activities. First, we detect every entry (“Enter” event) and exit (“Exit” event) in an experimental zone surrounding the host eggs; if the *Trichogramma* female exits from this zone and stays out more than sixty seconds, the experimentation is stopped. Then, we recognize the “Walk” event which corresponds to the walking of the *Trichogramma* females in the experimental zone between host eggs. Finally, we focus on the duration and time-bound of the three phases of egg laying behavior. The three phases are: (1) antennal drumming, (2) ovipositor drilling and (3) oviposition.

To reach this goal, we have used VSIP (Video Surveillance Intelligent Platform), described in [1], including a scenario recognition module based on [15]. We have also integrated a module to handle *Trichogramma* specific behaviors. Our approach computes the visual features characterizing mobile object shape to distinguish between the three phases of the egg laying behavior and allow the recognition of the global activities. After extending the vision module of this system, we have conceived a module which extracts these visual features and bind them to a scenario recognition module. As a feasibility proof, we have obtained acceptable results according to the complexity of the egg laying behavior. Figure 9 shows the output of the system at two instants: when it recognizes a Walk scenario (a) and when it recognizes a Drilling scenario (b).

This work can be improved by developing the following aspects. First, we plan to improve the segmentation algorithm in order to have a better determination of the three egg laying behavior phases. Currently, we focus on applying program supervision techniques to dynamically tune the parameters of the segmentation algorithm. A second task consists in refining the definition of the walk activity by a better understanding of the knowledge of the expert. Third, in order to develop an operational system, we plan to automate the description of the context of the scene (i.e. host eggs). For instance, currently we have to define the context for each new experiment due to the variations of the background. We are also planning to develop a learning module for automating the acquisition of the experimental context which will define the context of the scene without expert help. Finally, in the long term, we also plan to add a new front-end tool which will collect the output of the scenario recognition module for all experiments and mine these output results to deduce the frequent activities and their probabilistic law: it is claimed that the behavior of the *Trichogramma* while selecting a host egg can be described by the dynamic game theory.

The work accomplished up to now has been published in the International Cognitive Vision Workshop, (ICVW'2006)[25].

6.2.6. Human Posture Recognition

Participants: Bernard Boulay, François Brémond, Monique Thonnat.

We have proposed a real-time human posture recognition algorithm to be part of the automatic interpretation of monocular image sequences. This algorithm takes as input the silhouette (a pixel bitmap coding the detected person) provided by vision algorithms. The proposed approach is composed of four main tasks:

1. the silhouettes of the 3D posture avatars are generated for all possible orientations according to the estimated 3D position of the detected person and a virtual camera;
2. the generated silhouettes and the 2D silhouette of the detected person are compared using 2D techniques;
3. the posture of the detected person is selected according to the task 2;
4. the posture filter task uses postures computed on several frames to repair posture recognition errors from task 3.

The silhouettes of the 3D posture avatars are generated for all possible postures of interest. The generated silhouettes are obtained by projecting the corresponding 3D human model on the image plane, using the estimated 3D position of the person and a virtual camera which has the same characteristics (position, orientation and field of view) than the real camera.
The generated silhouettes are compared with the detected silhouette depending on the chosen silhouette representation. Several 2D silhouette representations have been studied and 4 representations have been chosen according to requirements in terms of computation time and silhouette quality: geometric features, Hu moments, skeletonisation and horizontal and vertical projections. The posture of the detected person is chosen as the posture which maximizes the similarity in term of silhouette.

This year, we have also focused on Task 4 by adding temporal filtering in the posture recognition approach. The recognized postures of a detected person on the previous frames are compared with the posture computed on the current frame to provide the most probable posture. This task recognizes stable postures to analyze the actions of the people observed in the scene.
The approach has been tested on both synthetic and real data. Table 3 gives the recognition rate for the four general postures (standing, sitting, bending and lying) on real data illustrating the 4 general postures in all orientations.

Table 3. General posture recognition rates (%) for the different silhouette representations.

	Standing	Sitting	Bending	Lying
Geometric features	94	82	77	83
Hu moments	68	73	27	35
Skeletonisation	93	68	82	65
H. & V. projections	100	89	78	93

This experiment shows that the approach accommodates satisfactorily the following problems, showing its robustness:

- Segmentation. The human posture recognition approach has been tested for two different segmentations. One segmentation provides silhouettes with some holes and the second one provides silhouettes with pixels which do not correspond to the detected person. By representing the silhouettes with the horizontal and vertical projections these differences are smoothed;
- Intermediate posture. The approach recognizes the most similar posture of interest for an intermediate posture which is slightly different from the posture of interest;
- Ambiguity. Postures can still be recognized in some ambiguous cases when two different postures have similar projections in the image;
- Human model. The generated silhouettes depend on the 3D posture avatar involved in the generation process. The 3D posture avatar is adapted to the detected person by considering her/his height.

The posture recognition approach have been successfully exploited to recognize two simple actions of a person: the *falling* action and the *walking* action. The actions are represented with a finite state machine where each state is represented with any type of posture (which can correspond to a combination of one or several postures of interest). Each state is also characterized by a minimal and maximal threshold values representing the authorized occurring number of successive postures.

However, this approach has some limitations. The first one is the number of postures of interest that can be tested. Indeed, the computing time increases when more postures are considered limiting the number of postures of interest. Moreover, when more postures avatars are considered, the number of ambiguity cases increases. A second limitation is the computation time of the generated silhouettes. By only computing the generated silhouettes when the detected person has a sufficient displacement in the scene, the frame rate is about 5 to 6 frames per second.

A part of this work has been published in the Pattern Recognition Letters [17].

6.2.7. Multi-sensors Analysis for Everyday Elderly Activity Monitoring

Participants: Nadia Zouba, François Brémond, Van Thinh Vu, Monique Thonnat.

In the framework of the GER'HOME project, we have worked on multi-sensors analysis for everyday elderly activity monitoring in order to improve elderly life conditions at home and to reduce the costs of long hospitalizations. The techniques will allow elderly people to stay safely at home, to benefit from an automated medical supervision and will delay their entrance in nursing homes. Thus, the objective of this work is the early detection of deteriorated health status and early diagnosis of illness. It consists in identifying a profile of a person - its usual and average behavior - and then to detect any deviation from this profile based on multi-sensor analysis and human activity recognition.

In this work we propose a video monitoring platform fed by a network of cameras and contact sensors. The platform performs 3 main tasks: (1) People detection, tracking and video event recognition; (2) Sensor stream filtering and contact event recognition; (3) Multimodal event recognition. The detection and tracking task detects and tracks mobile objects (mostly people) evolving in the scene. For each tracked mobile object the primitive event recognition task recognizes the events relative to the objects based on their visual features. Similarly, the contact event task recognizes the events characterized by contact information associated to the

tracked objects. Finally the multimodal event recognition task consists in combining the previous video and contact events in order to recognize more complex events. These complex events are specified by medical experts thanks to a user friendly language.

Our main goal is to improve the techniques of automatic data interpretation using complementary sensors installed in the apartment such as video cameras and contact sensors (installed on the doors, on the windows, in the kitchen cabinets and pressure sensors installed on the chairs). The proposed monitoring platform takes three types of input: (1) video stream(s) acquired by video camera(s), (2) data resulting from contact sensors embedded in the home infrastructure, and (3) *a priori* knowledge concerning event models and the 3D geometric and semantic information of the observed environment. The output of the platform is the set of recognized events at each instant.

Figure 10 illustrates the result of detection, classification and tracking of a person in GER'HOME laboratory.

Figure 10. Detection, classification and tracking of a person (from left to right: (a), (b) and (c)). (a) Detection output (b) Classification output (c) Tracking output.

In Figure 10(a), the detected moving pixels are highlighted in white and clustered into a mobile object enclosed in an orange bounding box. In Figure 10(b), the mobile object is classified as a person and a 3D parallelepiped matching the person indicates the position and orientation of the person. Figure 10(c) shows the individual identifier (IND 0) and a colored box associated to the tracked person. We illustrate on Figure 11 the recognition of a primitive state "Inside_zone" in the GER'HOME laboratory.

Figure 11. Recognition of the primitive state "Inside_zone" in GER'HOME laboratory. A text message "Person is in the Livingroom" is displayed on the screen when the event is recognized.

To model the activities of interest specified by medical experts, we have defined 3 composite events: `Use_food`, `Use_dishes` and `Prepare_meal`. As an example, meal preparation entails at least the detection of

a person in motion in the kitchen and use of cabinets where food, plates and/ or utensils are stored. Presence in the kitchen can be indicated by the detection of a person (video camera) in the kitchen lasting for a minimum duration of time, whereas the use of meal ingredients can be indicated by the use of a food storage cupboard or the refrigerator (contact sensors), etc. The multimodal (contact-video) event recognition algorithm recognizes which complex events occur combining primitive video events detected by the video detection module and the contact events detected by the contact detection module.

Preliminary results. The primitive state (Inside zone) has been well recognized by the video sensors and the primitive events (Open /close cupboard) are correctly recognized by the contact sensors. These primitive events define the "use food" and "use dishes" composite events, which define the "prepare meal" model. This model is recognized using both the video and the cupboard sensors. The preliminary results are encouraging to recognize more composite events for homecare applications using multi-sensors analysis.

This work has been published in the International Conference: Sciences of Electronic, Technologies of Information and Telecommunications (SETIT 2007) in TUNISIA [33]

6.2.8. Audio-Video Event Recognition For Scene Understanding

Participants: Van Thinh Vu, Gabriele Davini, François Brémond, Monique Thonnat.

Our goal is to extend existing video Event Description Language (EDL) and video event recognition algorithms ([50], [36]) for **audio-video event representation and recognition** for automatic scene understanding in the framework of the SAMSIT and SERKET projects.

For the first objective – the representation of audio-video events, we have added the notion of **audio event** in our video EDL for audio events that are directly detected by audio processing algorithms. In the extended EDL, a primitive audio event is represented as a state computed from audio processing and is related to the location of the associated microphone. To link this audio event to human activities, we suppose that this audio event has been performed by the people surrounding the microphone. Then, an audio-video event is described as a composite event [50]. Figure 12 shows an example of a composite event combining both audio and video information for describing a real situation for inside train surveillance.

composite-event(**vandalism_against_window**,
physical-objects(($vandal$: **Person**))
components(
($vandalism\_against\_window\_VIDEO$: ***composite-event***
vandal_close_to_window(vandal))
($vandalism\_against\_window\_AUDIO$: ***composite-event***
spray_detected_close_to_person(vand al)))
constraints(
($vandalism\_against\_window\_AUDIO$ ***during***
$vandalism\_against\_window\_VIDEO$))
alert("**Vandalism against window!**")
)

Figure 12. ***vandalism_against_window*** *event describes a situation where a passenger is using a spray to paint graffiti on a window of a train. This event model combines both audio and video information to describe more precisely the situation.*

For the second objective – the real-time recognition of audio-video events, there are two main issues to be focused on: (1) synchronization of audio and video information to fuse both events and (2) fast recognition of complex events. To cope with the synchronization issue, we currently use different configurations of transmission delays between components composing a complex audio-video event for the recognition algorithm to process temporal constraints with time *tolerances*. More precisely, we define different event models corresponding to variations of reception order between audio and video processing for modeling one audio-video event. In our experimentation, audio events are always detected with an important delay and consequently they

are modeled to be at the end of the complex audio-video events. Besides, the recognition of audio-video events remains as the recognition of video events.

Our audio-video EDL and the recognition algorithm have been tested and evaluated on different data of real-word applications. The experimental results show that the proposed algorithm can recognize robustly in real-time activities of interests specified for different applications (e.g. the French SAMSIT project of on-board train surveillance and the European SERKET project on security and threat assessment).

Our research results have contributed to a publication in the *Imaging for Crime Detection and Prevention 2006 conference* [31].

6.2.9. Unsupervised Behavior Learning and Recognition

Participants: Patrice Jacq, José Luis Patino Vilchis, François Brémond, Monique Thonnat.

We have developed an algorithm allowing to identify frequent behaviors on video data. In the context of this work, frequent behavior is considered as the composition of simple events and their association depends on the frequency of their spatio-temporal relationship. The input data for the algorithm are those simple events, previously defined in a generic library, and recognized on the video. The proposed approach builds a graph that relates simple events with each other following their topological distance and the existence of a temporal hierarchy between the events. Two main points are taken into account to build this graph, the existence and the denomination of a relationship between two simple events. The former indicates that two simple events cannot be related if their topological distance is not small enough and that a temporal hierarchy can be established between the two events. The latter denotes the kind of temporal hierarchy. The topology itself is application-dependent and can be defined from *a priori* knowledge of the data. The distance threshold and the possible temporal hierarchies are parameters of the algorithm. The main contribution of this work is the possibility to visualize in the graph, by means of a computer interface, the set of relationships that can link different simple events in a given topological space. Furthermore, we applied Data Mining techniques, currently employed on graphical data, to extract those relational patterns which appear more regularly and can thus be considered as frequent behavior. This algorithm has been applied to video data coming from a parking car surveillance system. The results were compared to those obtained on the same data by employing an *a priori* method previously developed in our group. The main inconvenient of the proposed approach is the algorithmic complexity associated to the graph structure and thus the high computing time required by our method. The future work of the project involves three main objectives. (i) Make the user interact earlier with the interface in order to decrease the complexity of the algorithm. (ii) Build a better representation of the geographical structure of the events by assembling them into coherent trajectories. (iii) Associate the generated graphs to similar semantical structures in order to automate the interpretation. This work will also be studied for the European research project CARETAKER.

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

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

The CARETAKER (Content Analysis and REtrieval Technologies to Apply Extraction to massive Recording) project [23] is a 30 months project starting from March 2006. 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. More precisely, CARETAKER will focus on the extraction of structured knowledge from large multimedia collections recorded over networks of cameras and microphones deployed in real sites. The produced audio-visual streams, if stored and automatically analyzed, represent a useful source of information in urban/environment planning, resource optimization, disabled/elderly person monitoring, etc., in addition to surveillance and safety issues.

We will consider two types of content knowledge: a first layer of primitive events that can be extracted from the raw data streams, such as ambient sounds, the degree of crowding present in the scene and the routes taken by individual people. A second layer of higher semantic events is defined based on more complex relationships between the primitive events and detected from longer term analysis.

Real testbed sites inside the metro of Roma and Torino, involving more than 30 sensors each (20 cameras and 10 microphones), will be provided. Additionally, the identification of the real user needs and beneficial use-case scenarios will serve as a reference point for the correct framing of the semantic description scheme i.e. the ontology, the knowledge extraction components and the interface and demonstrator optimization.

Within CARETAKER, the Orion project team is in charge of the long term tracking of objects of interest, the ontology-based event detection and of knowledge discovery.

- Ontology

  An ontology is the set of all concepts and relations between concepts shared by the community of a given domain. The ontology is particularly useful for experts of the application domain to use scene understanding systems in an autonomous way. In ORION, we have extended this year our video ontology toward two directions: (1) the ontology has been adapted for urban monitoring applications; (2) the video ontology has been combined with an ontology for audio events, to enrich the description of scenarios of interest.

- Tracking

  Tracking is one of the most studied topics in dynamic scene analysis. In ORION, as well as with other partners, novel algorithms available at the consortium have first been adapted to the specific environment of metro sites, and properly evaluated for the task. All the available context information is employed (e.g. rough calibration of the ground plane). In particular, one aim of CARETAKER is to increase robustness with respect to knowledge discovery: the detection and tracking of people and activity recognition should be performed on a long-term basis (more than one month).

- Event recognition and knowledge discovery

  The first objective in knowledge modeling and event recognition consists in developing new algorithms for the recognition of higher-level/composite events defined in the ontology from the user requirements, using the low-level primitive streams-of-data events coming from diverse sources of information (trajectories, audio/video activities), ontology-driven methods (i.e. scenario-based), or a mixture of both. For example, we are able to define and recognize when a luggage is being abandoned by a person based on two constraints: when a luggage remains in a predefined zone for a minimum predefined period of time and when it is located far enough to any passing persons, then this luggage is considered as 'abandoned luggage' as shown in figure 13. In this figure, the luggage labeled 'LUGGAGE 10' has been left apart by the person labeled 'PERSON 11' for a too long period of time and both person 11 and person 8 are located far enough from this luggage for it to be detected as abandoned.

  The second objective in knowledge modeling and event recognition is the investigation of unsupervised techniques for the recognition of both common events and unusual events. In all cases, the overall goal is to produce indexing information to feed the user system, for both the online supervision system and the offline retrieval one. Broadly speaking, different types of events exist, which can be characterized by their occurrence frequency, their expectedness, and their relevance. Their recognition poses different challenges that have been addressed using different models and methodologies. In ORION, the first results of primitive video events and object-of-interest trajectories are being analyzed using data mining tools. For instance, trajectories in a scene have been clustered into categories according to their entering and exiting a zone in the scene.

6.2.11. Evaluation of the VSIP platform on the sites

Participants: Magali Mazière, Florent Fusier, Valéry Valentin, François Brémond, Monique Thonnat.

This year we have evaluated our activity recognition approach through the completion of two projects, AVITRACK in January 2006 and CASSIOPEE in March 2006.

AVITRACK Project

The goal of the AVITRACK project was recognition of complex activities in airport apron monitoring.

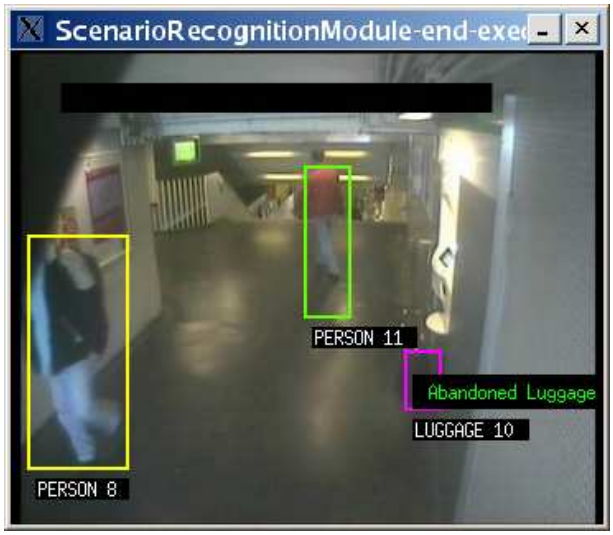


Figure 13. An abandoned luggage detected by the ontology-based event detector

We have modeled complex video events involving several vehicles and people operating around the vehicles interacting with each other. These events refer to activities containing several basic activities corresponding to the different steps of the whole operation. The complexity of video events can be illustrated by the detailed description of the "Front Unloading Operation". This operation consists of unloading the baggage containers of the aircraft through its front right door. The scenario model is complex and involves many physical objects, composite components, and temporal constraints.

The system now recognizes 21 basic video events, 8 video events involving the Ground Power Unit (GPU), 8 video events involving the Tanker, 3 video events involving the Aircraft, 3 video events involving the Tow Tractor, 3 video events involving the Conveyor, and 12 video events involving the Loader and the Transporter, so a total number of 58 video events. These video events and their complexity demonstrate both the system **effectiveness** and **flexibility**. Thanks to this project, we have shown that automatic video system can monitor an airport apron.

This work has been published in [19], [24].

CASSIOPEE Project

We have also completed the CASSIOPEE project on bank security monitoring. The VSIP system has been installed and evaluated in four bank agencies near Paris with different configurations using both recorded videos and live acquisition. The system has been designed to recognize bank attack scenarios. However it has been mainly devoted to count people in the secured technical room of the bank (ETS) since only a maximal number of persons are allowed to be together in this room. Concerning the counting people evaluation, the recorded videos corresponds to twenty days and 4 days were associated with Ground Truth. The situation with one person in the ETS was correctly detected 75 times out of 75 with no false detection. The situation with more than two persons in the ETS was correctly detected 6 times out of 11 with no false detection. The evaluation performed on live videos gave similar results. The main errors were due to two problems: the room was dark (the light was not turned on) and the position of the camera could not permit to observe the people evolving in the room. No false alarm has been observed during the whole evaluation process which was a main requirement from both end-users, the bank and the video security operators. The system has also been evaluated in live conditions for a few weeks. This evaluation has been performed directly by the company in

charge of the remote surveillance for a few weeks and by ourselves for several days. Thus we have been able to verify the effectiveness of our approach: the alarms are sent in real time to the surveillance headquarter. The ORION startup Keeneo is currently in charge of commercialization of the system.

6.2.12. A New Evaluation Methodology for Video Surveillance Algorithms

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

We lead the project ETISEO (Evaluation du Traitement et de l'Interprétation de Séquences Video) campaign on performance evalutation of video surveillance. More than 16 international teams have processed the video dataset and submitted their results. Our goal is to propose a new evaluation methodology for inferring meaningful information on the evaluated algorithms. Specifically we have carried out the following works:

- Refinement of the evaluation tools. Certain evaluation tools devised have some limits so appropriate modification is needed. We found that different participants use different hypotheses. For example, some participants do not detect the objects that do not move for a certain period of time. Therefore, it is impossible to compare the results of two participants that use two different hypotheses. To overcome this problem, filters are implemented to extract the reference data on which all the participants have adopted the same hypothesis;
- Analize the evaluation metrics. There are some metrics that add noise to the evaluation result. For some metrics, we have proposed to use other distance function so that these metrics become more accurate;
- Proposition of a new evaluation methodology. It can generalize the evaluation results performed on selected videos to new video sequences. More precisely, we address each video processing problem separately and estimate the upper bound of algorithm capacity in solving a given problem. If this value is smaller than the difficulty level of new sequences, we can conclude that the algorithm cannot achieve acceptable performance on these sequences. To validate the new evaluation methodology, we have defined two metrics to address the problems of handling weakly contrasted objects and objects mixed with shadow. The preliminary results show that, with this methodology, we can extrapolate the evaluation results for new sequences.

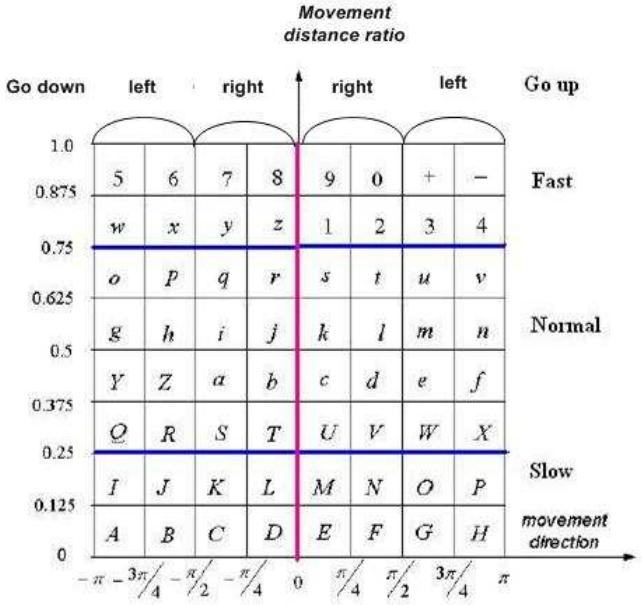
Finally we have organized a seminar on the evaluation campaign gathering more than 50 participants from international organizations. This work has been published in [28].

6.2.13. Content-based Video Indexing and Retrieval

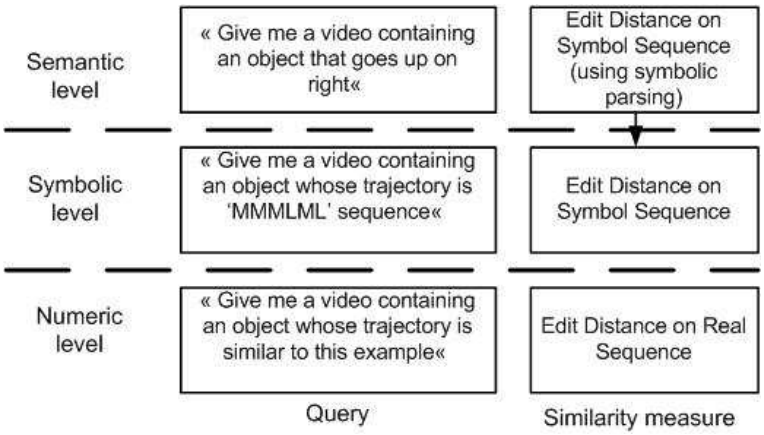
Participants: Lan Le Thi, François Brémond, Monique Thonnat.

Advances in computer technologies and the World Wide Web have made an explosion of multimedia data in particular video being generated, stored, and transmitted. For managing this amount of information, one needs to develop efficient content-based retrieval approaches that enable users to search information directly via its content. Motion is an important feature in video. When browsing a video, people are more interested in the actions of a car or an actor than in the background. Among the extracted features from object motion, we have firstly used the trajectory because of its significance for describing the object motion. In order to answer the user query at different levels, we have represented object trajectory at three levels: numeric, symbolic and semantic levels. At the numeric level, a trajectory that is represented by a sequence of object positions is transformed into a sequence of pairs of direction and relative distance ratio by using the method proposed in [38]. At the symbolic level, we have used the symbolic method in [37] to convert the sequence of pairs of direction and distance ratio in the numeric level into a sequence of symbols according to the reference map (figure 14.a). This map is created by quantizing the (direction, distance ratio) space and representing each sub region by a distinct symbol. At the semantic level, some heuristic rules that link the semantic description to the symbols at the symbolic level are introduced based on some semantic analysis on the reference map. As one see on the reference map, if one object ***goes up*** on the ***right***, whose trajectory at the symbolic level will contain 'M', 'N', 'E' or 'F' symbols. The similarity measures (Edit Distance on Real Sequence, Edit Distance on Symbolic Sequence) based on the Edit Distance on Strings [38] are used to compute the distance between trajectory query and trajectories in the database. Figure 14.b gives three examples of trajectory query at the

three levels. At the symbolic level, user gives 'MMMLML' sequence by choosing symbols from the reference map in figure 14.a corresponding to trajectory description that he/she wants. This sequence means that the user is interested in an object trajectory whose direction is between $-\pi/4$ and $\pi/4$ and whose distance ratio is between 0.125 and 0.25.



(a)

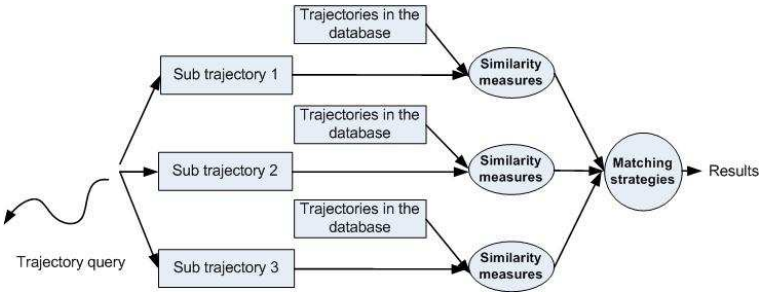


(b)

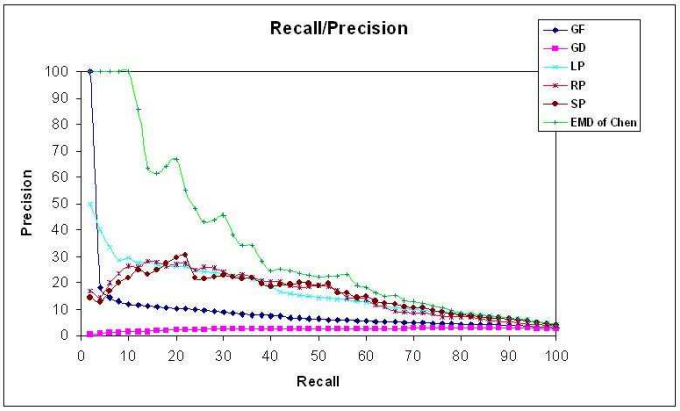
Figure 14. (a) Reference map for three levels of trajectory representation (b) Three levels for trajectory query

Basing on the three levels of trajectory representation and the similarity measures, we can match the trajectory query to trajectories in the database. We call it ***global*** matching strategy to distinguish from others matching

strategies that will be presented. When working with complex trajectories and using global matching, we may loose the local matching between trajectories, so we segment a trajectory into some sub-trajectories. In order to have a similarity measure between two segmented trajectories based on the similarities of their sub-trajectories (Figure 15.a), we have proposed various matching strategies including ***hybrid*** (Dominant Segment (GD), Full Trajectory (GF)) and ***partial*** (Strict Partial (SP), Relative Partial (RP), Loose Partial (LP)) matching. More details about these strategies can be found in [26]. These matching strategies take into account a large of variety of the user needs. We have tested this approach with a trajectory database containing 2500 handmade trajectories illustrated 50 categories. Figure 15.b shows the recall/precision curves for global, hybrid and partial trajectory matching.



(a)



(b)

Figure 15. (a) The role of matching strategies when working with segmented trajectory (b) Recall/precision curves for global (EDM of Chen [37]), hybrid (GF, GD) and partial (LP, RP, SP) trajectory matching

The trajectory represents only one aspect of object motion. Therefore, it is not sufficient to describe object motion. For example, one person could *walk* or *run* with the same **trajectory**. In this case, the speed information could be used to distinguish *walking* from *running* action. Therefore, we are planning to use other information of object motion such as speed, context clues, interested zone, object type and so on. This work is published in [26].

6.3. Cognitive Vision Platform

Participants: Nicolas Chleq, Lan Le Thi, 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 16). 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.

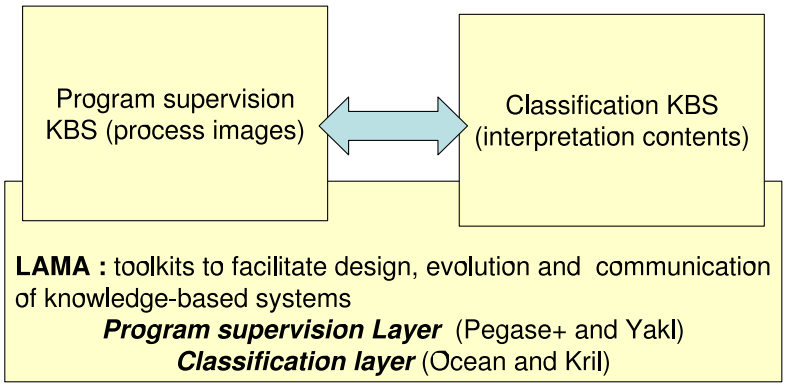


Figure 16. Architecture of the cognitive vision platform

6.3.2. Knowledge-based Semantic Interpretation

Participants: Nicolas Chleq, Vincent Martin, Sabine Moisan, Monique Thonnat.

The cognitive vision platform relies of course on the LAMA software platform toolkits, more precisely on the program supervision layer (PEGASE+ engine and YAKL language) and on the new interpretation layer, that we have started last year, motivated by classification/categorization applications. A prototype cognitive system for early detection of plant diseases is in use at INRA-URIH (Sophia Antipolis). The prototype system integrates the following parts:

- The visual concept ontology, proposed by Nicolas Maillot and Céline Hudelot during their PhDs, is used to formalize the domain knowledge from experts via the KRIL dedicated language (see an example in table 4);
- A classification KBS, based on the new OCEAN engine is used for the interpretation task (in connection with the visual concept ontology);

- A program supervision KBS, based on the PEGASE+ engine is used to supervise image processing requests, such as image segmentation and features extraction. The knowledge on such image processing operators are formally described using the existing YAKL language;
- Both KBSs communicate and exchange information through a (currently simple) interface.

Table 4. Example of a Kril file describing knowledge of objects with respect to their visual concept attributes. It descrides the shape concept for an Aleurode object with the fuzzy range values of its attributes.

VisualConcept	{	name	AleurodeShapeConcept
		Superclass	ShapeConcept
		Constraints	
		$shape.circularity$	[0.05 0.2 0.5 0.6]
		$shape.excentricity$	[0.1 0.2 0.4 0.5]
		$shape.rectangularity$	[0.5 0.6 0.8 0.85]
		$shape.elongation$	[0.3 0.35 0.7 0.8]
		$shape.convexity$	[0.7 0.75 1 1.1]
		$shape.compacity$	[0.1 0.25 0.9 1]
	}		

The classification layer is still under construction. This year, we have pursued our work on the new classification engine (OCEAN) and its knowledge representation language (KRIL) for classification tasks. The classification layer has thus been enriched to improve the genericity of the classification algorithm, the communications with the program supervision layer and the connection with ontology management. A lot of work remains to be done to design generic components in LAMA to accommodate the needs of the classification task in general and to manage communication between classification and other tasks. Extensions of the KRIL language have also been specified to better take into account constraints on objects to be classified and relationships between objects.

6.3.3. Rose Disease Application

Participants: Nicolas Chleq, Vincent Martin, Sabine Moisan, Monique Thonnat.

An application of the cognitive vision platform has been delivered to INRA. It achieves the detection of mature white flies (Aleurodes) and theirs eggs on rose leaves (see figure 17).

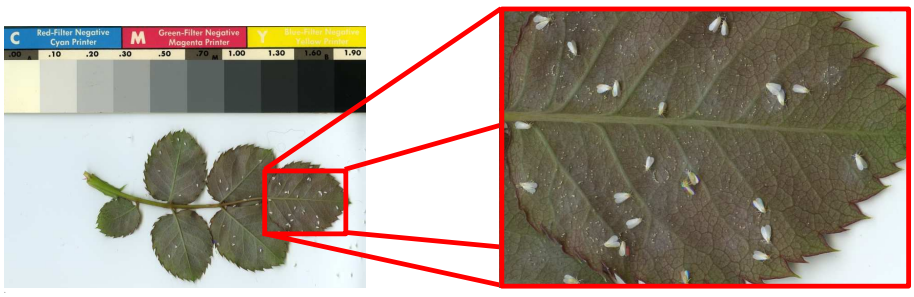


Figure 17. A scanned rose leaf with aleurodes and theirs eggs

The interpretation of rose disease is performed by a classification KBS. Its knowledge base contains some biological knowledge of rose disease leaves useful for white fly recognition (5 KRIL classses using 5 visual concepts).

The image processing tasks are performed by a program supervision KBS. Its knowledge base contains 11 image processing YAKL operators (7 primitive operators and 4 composite operators) and 12 YAKL rules.

In particular, the segmentation operator is based on a smart algorithm able to automatically (1) extract the leaf from the background, (2) denoise the resulting image in an adaptive manner and (3) segment the image into regions. Then, the feature extraction operator is called to compute multiple region attributes, according to the domain feature concepts (e.g. color, shape and geometric descriptors). These attributes are then used to classify each region of the segmented image. At this time, the system is able to detect and count two different objects: mature white flies and their eggs.

The system has been tested on a database composed of a representative sample of 200 images of scanned rose leaves (equivalent to a greenhouse of 200 square meters) provided by INRA-URIH (Sophia Antipolis). Visual descriptor (e.g. size, color, shape) values of objects to detect have been learned manually on 20 images. The global computing time for an image of 2495x4056 pixels is around 45 seconds. The result of a classification request on an input image is presented in figure 18.

The counting results have been compared with the groundtruth. The detection rate of white flies is satisfactory: in average, 6.33 flies have been counted per leaf and 5.62 have been detected by the platform. However, a deeper analysis of the results shows some limitations. Indeed, ambiguous cases such as occlusions or overlapping objects cannot be treated without explicit *a priori* knowledge of object subparts. More specialized operators must be integrated to extract such information.

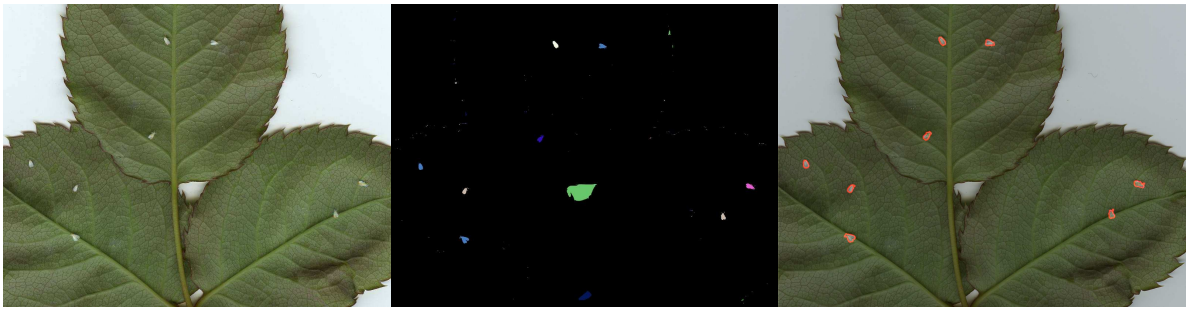


Figure 18. Top left: part of an input image, top right: result of the segmentation task, bottom: result of the classification with 8 correctly detected white flies (surrounded in red)

6.3.4. Supervised Learning for Adaptive Segmentation

Participants: Vincent Martin, Monique Thonnat.

Last year, we proposed an approach for automatic segmentation algorithm selection and parameter tuning by generating a global feature vector out of an image [27]. During the training stage, optimal algorithm parameters of training images were extracted by an optimization procedure. Manual segmentations were used to assess the quality of the segmentation. A neural network was used to learn the selection of segmentation algorithm then case-based reasoning was performed on the global feature vector to select the optimal parameters.

This approach has several limitations. First, only using global image information (i.e. global feature vector) makes the segmentation brittle to non-global variations. Second, the segmentation quality assessment is area-based and does not take into account pixel values information. Third, manual segmentations (used during the learning stage) are subjective by nature and time-consuming.

This year, we have focused our work to deal with these three limitations.

1. The learning stage has been improved so it can take into account local information. Instead of training global feature distribution models on whole images, we train models locally, using image regions information. This allows to enrich feature models with spatial information.
2. A new segmentation quality assessment based on region-class evaluators has been proposed. A region-class evaluator relies on local feature models trained during the learning stage. Thus, the segmentation goal directs the region evaluation. A region-class evaluator takes as input a feature vector and returns a segmentation quality measure as output. The global evaluation of the segmentation is a weighted sum of local region evaluations.
3. The learning stage has also been designed to ease the user effort from manual segmentations (strong supervision) to region annotations (weak supervision). An annotation is a label associated with a region. The labelization procedure is guided by the segmentation goal: one user can label all the regions of an image with the same label (e.g. when all regions refer to the background class), or the user can consider several classes for the labeling (e.g. when a segmented image is composed of several regions referring to different objects). Thus, building large training databases becomes easier than when using manual segmentations and also lead to more robust trained models.

In order to facilitate the task of region annotations, a graphical tool has been designed. The interface gives the possibility to annotate segmented images, train region-based evaluators, optimize the parameters of a selected algorithm and perform automatic segmentation according to the acquired knowledge. It works independently of any segmentation algorithm or application domain.

6.3.5. Towards Automatic Annotations of Videos for Epilepsy Behavior Analysis

Participants: Vincent Martin, Lan Le Thi, Monique Thonnat.

In this work, one purpose in terms of activity monitoring is to model the changes in behavior for a given patient in epileptic ceasure. In order to characterize such changes, the experts in neurophysiology (particularly Prof P. Chauvel, Inserm E9926/CHU La Timone Marseille) have defined a terminology. It aims to standardize the annotation task of epileptic patient videos. A set of terms is used to describe observable events referred to as subjective or objective signs. An objective sign describes the distress of a patient observed by another person (e.g. medical staff) using audio/video information, while a subjective sign is a distress percepted by the patient himself.

The video annotation is performed via two steps. In the first step, a medical staff takes notes in a free language of whatever he/she sees in the video. The second step consists in making correspondences between the notes and the objective signs. Currently, the list is composed of 51 objective signs divided into 8 groups: muscular signs (e.g. tonic posture, head/eyes version), vegetative signs (e.g. rubefaction), emotional signs (e.g. positive, negative), language signs (e.g. verbal), ocular signs (e.g. open wide eyes), alimentary signs (e.g. chewing), conscious signs (e.g. consciousness loss) and undefined signs.

The text hereafter presents an extract of video notes with the corresponding associated objective signs in square brackets:

> *The patient is sleeping on his bed. At 00.02 he opens his eyes and stands up. Then at 00.07 we can see a sudden movement of his hands towards his head and he has a tonic posture of his upper limbs and lower limbs* [***tonic posture***]. *He gets into a rubefaction* [***rubefaction***] *and a small version of eyes towards the left side* [***head/eyes version***]. *At 00.27 he shows agitated movements and even kicks unwillingly the nurse* [***general agitation***].

Currently, such annotations are performed manually by medical staff, previously trained to do this job. However, this task is time consuming as a great amount of videos are accumulated each day. Therefore, our ultimate goal is to automate the annotation task. From the list of 51 objective signs, we have decided to focus on the muscular group signs because they rely on events which are relatively detectable with vision techniques. These signs are related to body parts events (e.g. *head/eyes version*) or whole body events (e.g. *general agitation*, *body rotation*). Our goal is to extract significant information from image sequences to characterize such signs. In order to do so, we have identified two issues:

1. We must be able to extract efficiently body parts from images. In particular, a classical bounding box approach (as used in section 6.2), for motion-based segmentation, is not precise enough. For instance, rotation of the head could not be detected using this method. Hence, we propose to study new approaches (e.g. body part segmentations) to cope with such problems.
2. Temporal information is crucial and should be used to tackle the video annotation problem. Currently, the cognitive vision platform interprets static image contents. The visual concept ontology defined in [7] should be broadened to deal with motion concepts. Motion concepts could be *speed*, *direction*, *frequency* and *spatio-temporal relationship* concepts. The KRIL language could be used to formalize the domain knowledge within the ontology.

At present time, we are working on the segmentation task. A watershed-based motion segmentation algorithm [42] is used to obtain a figure-ground segmentation (i.e. a binary mask). Then, a region-based segmentation algorithm is combined with the binary mask to refine the segmentation of moving objects into homogeneous subparts. Such segmentation may be used to extract regions corresponding to moving body parts. Figure 19 illustrates a preliminary result.

Figure 19. left: image from a video of a patient in ceasure by the courtesy of Prof P. Chauvel, Inserm E9926/CHU La Timone Marseille, right: segmented result where color labels correspond to mean region colors and white is used for background

7. Contracts and Grants with Industry

7.1. Industrial Contracts

In 2006, ORION team has been involved in 4 industrial projects : CASSIOPEE project on bank agency visual surveillance, SAMSIT project on train visual surveillance, TELESCOPE 3 project to improve a toolkit in cognitive video interpretation and SIC project on security.

7.1.1. CASSIOPEE

The **CASSIOPEE** project aimed at developing and testing an automatic visual-surveillance platform for detection of predefined scenarios in a bank agency environment. The project, in collaboration with Le Credit Agricole, Ciel and Securitas Systemes - Eurotelis started in January 2002 and ended March 2006.

7.1.2. SAMSIT

SAMSIT is a project in collaboration with ALSTOM, CEA, SNCF, INRETS. It has begun in January, 2004 and ended in March 2006. The aim of this project was to develop novel techniques to automatically detect human behaviors in trains. Such environments are difficult ones, due to train motion, narrow environment and fast illumination changes.

7.1.3. TELESCOPE 3

TELESCOPE 3 is a project with Bull to complement an initial project (ended in 2001) in which a toolkit in the domain of cognitive video interpretation for video surveillance applications(VIS) has been achieved . The purpose of this project is to improve this toolkit in order to facilitate its usage, to ensure more robustness and to extend its functionalities. The project is funded by BULL.

7.1.4. SYSTEM@TIC SIC Project

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

8. Other Grants and Activities

8.1. European projects

ORION team has been involved this year in three european projects a project on airport surveillance (AVITRACK), a project on crowd behavior analysis (SERKET) and a new project on multimedia information retrieving.

8.1.1. AVITRACK Project

AVITRACK is a European project in collaboration with Silogic S.A. Toulouse (FR) University of Reading (UK), CCI Aeroport Toulouse Blagnac (France), Fedespace (France), Tekever LDA, Lisbon (Portugal), ARC Seibersdorf research GMBH, Wien (Austria), Technische Universitaet, Wien, (Austria) , IKT (Norway) and Euro Inter (Toulouse France). This 2-year project has begun in February 2004 and ended in March 2006. The main objective of this project is to recognize the activities around parked aircrafts in apron areas. Activities may be simple events involving one mobile object like the arrival or the departure of ground vehicles or complex scenarios like refulling or luggage loading.

8.1.2. 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. The main objective of this project is to develop techniques to analyze crowd behaviors and to help in terrorist prevention.

8.1.3. 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 others partners are: Multitel (Belgium), Kingston University (UK), IDIAP (Switzerland), Roma ATAC Transport Agency (Italy), 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 the international program STIC-Asie (ISERE) and in academic collaboration wtih ENSI in Tunis (a joint PhD is in progress).

8.2.1. STIC-Asie:ISERE

Our team is a member of the specific Inter-media Semantic Extraction and Reasoning (ISERE) action. ISERE action gathers four research centers from Asia and three French teams. It concerns both the development of research on semantics analysis, reasoning and multimedia data, and the application of these results in the domains of e-learning, automatic surveillance and medical issues. Besides allowing to share scientific results, this cooperation must increase the exchange of researchers between Asia and France, and more precisely Phd students.

The Asia partners of the ISERE action are: IPAL (Jean-Pierre CHEVALLET, CNRS), I2R A-STAR (Mun Kew Leong) and NUS (CHUA Tat Seng) for Singapor, MICA (Eric Castelli) for Vietnam, the National Institute of Informatics (NII, Shin'ichi Satoh) for Japan and the National Cheng Kung University (Pau-Choo Chung) and the National Taiwan University (Yi-Ping Hung) for Taiwan. The French partners are: INRIA (Monique Thonnat-Equipe Orion), CLIPS-IMAG (CNRS-INPG-UJF, Catherine Berrut-Equipe MRIM) and IRIT (Philippe Joly).

8.2.2. 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 co-direct the Master Thesis of several Tunisisan students on topics related to distributed program supervision.

8.3. National Grants

Orion Team has six national grants: three of them were already established last year and each involves a PhD thesis. A new collaboration in medical domain involves a post-doc fellow. Another new grant concerns passengers classification in the framework of a Phd thesis funded by RATP. The last one is part of the Techno-Vision evaluation network funded by the French ministries of defence and research.

8.3.1. 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.

8.3.2. Intelligent Cameras

Orion also cooperates with STmicroelectronics and Ecole des Mines de Paris at Fontainebleau for the design of intelligent cameras including image analysis and interpretation capabilities. In particular a PhD thesis (Bernard Boulay) is on-going on new algorithms for 3D human posture recognition in real-time for video cameras

8.3.3. Long-term Monitoring Person at Home

Last year, 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 PACA region. 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.

8.3.4. Classification of Lateral Forms for Control Access Systems

In the framework of a collaboration with RATP, B. Bui has started a Phd thesis 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.

8.3.5. Video Understanding Evaluation

Orion is member of the ETISEO project. This project began in 2005 and aims at providing both dataset and evaluation tools which should constitute a reference in "vehicles and pedestrians scene understanding". Project ETISEO focuses on the treatment and interpretation of videos involving pedestrians and (or) vehicles, indoor or outdoor, obtained from fixed cameras. This project is part of the Techno-Vision evaluation network funded by the French ministry of defence and the French ministry of research. The main partners are: Silogic (coordinator, evaluator and data-provider), INRIA (scientific leader), INRETS-LEOST (data-provider) and CEA-List (data-provider). Moreover, it gathers teams that have developed algorithms and want to evaluate their technology on their own. ETISEO project will provide them with free videos and metrics to run evaluations.

8.4. Spin off Partner

Keeneo 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), CVIU (Journal of Computer Vision and Image Understanding), IEEE Transaction on Multimedia, AEROBIOLOGIA, TS and Signal, 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: ACIVS, ISVC, IEEE WMVC, RFIA06, CVPR07, VISAPP and ICVW and in the editorial board of RFIA06.
- M. Thonnat is a member of the Joint Executive Committee to organize cooperations between the NSC (Taiwan) and French research teams. Franco-Taiwan conferences related to Multimedia and Web Technologies.
- M. Thonnat is an expert for ANVAR the French agency for research valorization and for DGA.
- M. Thonnat is an expert for research proposals in UK Leverhulme Trust and in France for ANR Jeunes chercheurs.
- M. Thonnat is an expert for the French Ministry of Defence.
- M. Thonnat is reviewer for the following theses: Yifan Shi (Georgia Tech Atlanta, USA), Nicolas Gourier (INPG Grenoble), Romain Lerallut (CMM Mines Paris).

- M. Thonnat is member of the INRIA Evaluation board since 2003. She has organized in May the evaluation of CogC INRIA theme.
- 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 BOEMIE (Podebrady, 6 October 06) on Ontology based Object Learning and Recognition.
- M. Thonnat has given a tutorial on Cognitive Vision Techniques for Video Analysis and Understanding at VIE conference at Bangalore India in September 06.
- 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 scientific organizer of Etiseo workshops on the evaluation of techniques for video interpretation.
- F. Brémond is "handling editor" for the international journal Computers and Artificial Intelligence since September 2006.
- F. Brémond is reviewer for the journals: IEEE Transactions on Multimedia, Signal, Image and Video Processing journal, CVIU (Journal of Computer Vision and Image Understanding) and IEEE Transactions PAMI (Patterns Analysis and Machine Intelligence).
- F. Brémond is Program Committee member of the conferences and workshops: VIE06, VS-Pets2006, VS2006, IET ICDP'06 et IEEE ICNSC'06 (International Conference on Networking and Sensing Control), IEEE AVSS'06 (International Conference on Advanced Video and Signal based Surveillance), ICVS'07, VS2007, VIE'07, IWINAC-2007 (International Work-conference on the Interplay between Natural and Artificial Computation).
- F. Brémond is a reviewer for the conferences and workshops: IEEE ECCV'06 (Europeen Conference on Computer Vision), CDPR'06, AVSS'06, BMVC'06, CVPRÂ´07.
- Sabine Moisan is a member of the Scientific Council of INRA for Applied Computing and mathematics (MIA Department).
- Sabine Moisan has presented Orion/NRA joint work at the 60th Anniversary of INRA in Sophia Antipolis.
- 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);
- Teaching at ISIA (Institut d'Informatique et d'Automatique, Ecole des Mines de Paris) grammar analysis lecture and TP (16h A. Ressouche).
- 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 taught the course about Concepts of Programming Languages at ISIA (Ecole des Mines de Paris at Sophia Antipolis).

9.3. Thesis

9.3.1. Thesis in progress

- Bernard Boulay : Human Posture Recognition for Behaviour Understanding, Nice Sophia Antipolis University.
- 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).
- Vincent Martin : Vision cognitive: apprentissage supervisé pour la segmentation d'images, Nice-Sophia Antipolis University.
- 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 téleassistance médicale á domicile, Nice Sophia-Antipolis University.
- Marcos Zúñiga : Unsupervised Primitive Event Learning and Recognition in Video, Nice-Sophia Antipolis University.

9.3.2. Thesis defended

- Benoit Georis: Program Supervision Techniques for Easy Configuration of Video Understanding Systems, Louvain Catholic University (thesis defended in January 2006).

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. *Advanced Video-based Surveillance Systems*, chap. Image Understanding for Prevention of Vandalism in Metro Stations, 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, avril 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, octobre 2003.

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

[14] V. T. VU, F. BRÉMOND, M. THONNAT. *Temporal Constraints for Video Interpretation*, in "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. GEORIS. *Program Supervision Techniques for Easy Configuration of Video Understanding Systems*, Ph. D. Thesis, Louvain Catholic University, January 2006.

Articles in refereed journals and book chapters

[17] B. BOULAY, B. BRÉMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System*, in "Pattern Recognition Letter, Special Issue on vision for Crime Detection and Prevention", vol. 27, n^{o} 15, November 2006, p. 1788-1796.

[18] F. BRÉMOND, M. THONNAT, M. ZÚÑIGA. *Video Understanding Framework For Automatic Behavior Recognition*, in "Behavior Research Methods Journal", vol. 3, n^{o} 38, 2006, p. 416-426.

[19] 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 Journal", To be published, 2006.

[20] B. GEORIS, F. BRÉMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", To be published, 2006.

[21] N. KHAYATI, W. LEJOUAD-CHAARI, S. MOISAN, J.-P. RIGAULT. *Distributing Knowledge-Based Systems Using Mobile Agents*, in "WSEAS Transactions on Computers", vol. 5, n^{o} 1, January 2006, p. 22-29.

[22] N. MAILLOT, M. THONNAT. *Ontology Based Complex Object Learning and Recognition*, in "Image and Vision Computing Journal, Special Issue on Cognitive Computer Vision", To appear, 2006.

Publications in Conferences and Workshops

[23] C. CARINCOTTE, X. DESURMONT, B. RAVERA, F. BRÉMOND, J. ORWELL, J. VELASTIN, J. M. ODOBEZ, B. CORBUCCI, J. PALO, J. CERNOCKY. *Toward generic intelligent knowledge extraction from video and audio: the EU-funded CARETAKER project*, in "IET conference on Imaging for Crime Detection and Prevention (ICDP 2006), London, Great Britain", June 2006, http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[24] B. GEORIS, M. MAZIÈRE, F. BRÉMOND, M. THONNAT. *Evaluation and Knowledge Representation Formalisms to Improve Video Understanding*, in "Proceedings of the International Conference on Computer Vision Systems (ICVS'06), New-York, NY, USA", January 2006.

[25] M. B. KAÂNICHE, F. BRÉMOND, M. THONNAT. *Monitoring Trichogramma Activities from Videos : An Adaptation of Cognitive Vision System to Biology Field*, in "International Cognitive Vision Workshop, (ICVW'2006) in conjunction with the 9th European Conference on Computer Vision (ECCV'2006), Graz, Austria", 2006, http://www-sop.inria.fr/orion/Publications/Articles/ICVW06.html.

[26] T. L. LE, A. BOUCHER, M. THONNAT. *Subtrajectory-based video indexing and retrieval*, in "13th International MultiMedia Modeling Conference (MMM), Singapore", To appear in 2007, 2006.

[27] V. MARTIN, N. MAILLOT, M. THONNAT. *A Learning Approach for Adaptive Image Segmentation*, in "Proceedings of the International Conference on Computer Vision Systems (ICVS'06), New-York, NY, USA", January 2006.

[28] A. T. NGHIEM, F. BRÉMOND, M. THONNAT, R. MA. *A New Evaluation Approach for Video Processing Algorithms*, in "Proceedings of Workshop on Motion and Video Computing (WMVC'07), Austin, Texas, USA", To appear in February 2007, 2006.

[29] A. TOSHEV, F. BRÉMOND, M. THONNAT. *An A priori-based Method for Frequent Composite Event Discovery in Videos*, in "Proceedings of 2006 IEEE International Conference on Computer Vision Systems, New York USA", January 2006.

[30] J. VIDAL, S. MOISAN. *Fuzzy knowledge-based curve evaluation for 1-D river model calibration*, in "Hydroinformatics 2006 – Proceedings of the 7th International Conference on Hydroinformatics, Chennai, India", P. GOURBESVILLE, J. A. CUNGE, V. GUINOT, S.-Y. LIONG (editors). , vol. 2, Research Publishing, 2006, p. 1203-1210.

[31] V. T. VU, F. BRÉMOND, G. DAVINI, M. THONNAT, N. PHAM, P. SAYD, J. L. ROUAS, S. AMBELLOUIS, A. FLANCQUART. *Audio Video Event Recognition System for Public Transport Security*, in "IET conference on Imaging for Crime Detection and Prevention (ICDP 2006), London, Great Britain", June 2006, http://www-sop.inria.fr/orion/Publications/Articles/ICDP06.html.

[32] B. ZHAN, P. REMAGNINO, S. VELASTIN, F. BRÉMOND, M. THONNAT. *Matching gradient descriptors with topological constraints to characterise the crowd dynamics*, in "3rd International Conference on Visual Information Engineering (VIE 2006), Bangalore, India", September 26-28 2006.

[33] N. ZOUBA, F. BRÉMOND, M. THONNAT, V. VU. *Multi-sensors Analysis for Everyday Elderly Activity Monitoring*, in "Proceedings of 4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications (SETIT 2007), Acapulco, Mexico", To appear in 2007, 2006.

[34] M. ZÚÑIGA, 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.

References in notes

[35] C. ANDRÉ, M. PERALDI-FRATI, J. RIGAULT. *Scenario and Property Checking of Real-Time Systems Using a Synchronous Approach*, in "4th IEEE Int. Symp on Object-Oriented Real-Time Distributing Computing, Magdeburg", May 2001, p. 438-444.

[36] F. BRÉMOND, N. MAILLOT, M. THONNAT, V. T. VU. *Ontologies for Video Events*, Technical report, n^{o} 5189, INRIA, April 2004, http://hal.inria.fr/inria-00071397.

[37] L. CHEN, M. ÖZSU, V. ORIA. *Symbolic Representation and Retrieval of Moving Object Trajectories*, in "In Proc of 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR), New York", October 2004, p. 227-234.

[38] L. CHEN, R. T. NG. *On the marriage of Lp norms and Edit Distance*, in "In Proc of International Conference on Very Large Data Bases (VLDB)", 2004, p. 792-803.

[39] 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.

[40] L. DANIEL. *Etude et application des méthodes d'interprétation abstraite pour la vérification des propriétés des systèmes dynamiques hybrides*, Technical report, Nice Unervisity, 2006.

[41] D. LANGE, M. OSHIMA. *Programming and Deploying Java (TM) Mobile Agents with Aglets (TM)*, Addison-Wesley, 1998.

[42] R. LERALLUT. *Modélisation et Interprétation d'Images à l'Aide de Graphes*, Ph. D. Thesis, Centre de Morphologie Mathématiques, Ecole des Mines de Paris, 2006.

[43] B. LUCAS, T. KANADE. *An Iterative Image Registration Technique with an Application to Stereo Vision*, in "Proc. of the International Joint Conference on Artificial Intelligence", 1981, p. 674-679.

[44] S. MOISAN. *Program Supervision:* YAKL *and* PEGASE+ *Reference and User Manual*, Tecnichal Report, n^{o} 5066, INRIA, December 2003, http://hal.inria.fr/inria-00071518.

[45] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *A Behavior Model of Component Frameworks*, Technichal Report, n^{o} 5065, INRIA, December 2003, http://hal.inria.fr/inria-00071519.

[46] J. SHI, C. TOMASI. *Good Features to Track*, in "Proc. of the IEEE Conference on Computer Vision and Pattern Recognition", 1994, p. 593-600.

[47] 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.

[48] 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.

[49] 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.

[50] V. T. VU. *Temporal Scenarios for Automatic Video Interpretation*, Ph. D. Thesis, Nice University, October 2004.

[51] I. H. WITTEN, E. FRANK. *Data Mining: Practical Machine Learning Tools and Techniques (Second Edition)*, Morgan Kaufmann Publishers, 2005.