INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Orion

Intelligent Environments for Problem Solving by Autonomous Systems

Sophia Antipolis

THEME COG

Activity Report

2005

Table of contents

1. Team

Head

Monique Thonnat [DR Inria]

Administrative Assistant

Catherine Martin

Research Scientists

François Brémond [CR1 Inria]
Sabine Moisan [CR1 Inria, long-term medical leave since April 2005]
Annie Ressouche [CR1 Inria]

Post-doctoral Fellow

Florence Duchêne [INRIA Post Doc grant, up to November 2005]

Technical Staff

Alberto Avanzi [Bull Engineer]
Julien Canet [Fellow Engineer, up to August 2005]
Nicolas Chleq [IR2 Inria ,April-December 2005]
Frédéric Cupillard [Videa/Vigitec Brussels, up to November 2005]
Gabriele Davini [SAMSIT project]
Florent Fusier [European project AVITRACK]
Ruihua Ma [SERKET project, since November 2005]
Magali Mazièree [Cassiopée Project/Crédit Agricole]
Christophe Tornieri [RASV, SNCF Paris, up to May 2005]
Valery Valentin [European project AVITRACK]
Thinh Van Vu [SERKET project, since October 2005]

Phd Students

Bernard Boulay [Paca Lab Grant, Nice Sophia-Antipolis University]
Binh Bui [CIFFRE RATP Grant, Nice Sophia-Antipolis University, since February 2005]
Benoit Georis [Louvain Catholic University (UCL) Belgium]
Céline Hudelot [Regional Grant, Nice Sophia-Antipolis University, up to September 2005]
Naoufel Khayati [STIC-Tunisie grant, ENSI Tunis]
Lan Le Thi [Hanoi University and Nice Sophia-Antipolis University, since October 2005]
Nicolas Maillot [Nice Sophia-Antipolis University]
Vincent Martin [Regional Grant, Nice Sophia-Antipolis University]
Jean-Philippe Vidal [CEMAGREF Grant, Toulouse University, up to February 2005]
Marcos Zuniga [CONYCIT Grant,Nice Sophia-Antipolis University]

Student Intern

Mohammed Becha Kaaniche [AVF grant, ENIT Tunis, since October 2005]
Alexander Toshev [Karlsruhe University, February-September 2005]

Invited Professors

Bob Fisher [Edinburgh University, May-September 2005]
Paolo Remagnini [Kingston University, October-November 2005]
Jean-Paul Rigault [Professor, Nice Sophia-Antipolis University, since September 2005]

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 the representation of knowledge about the 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 facilitate the construction of **cognitive systems**.

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

2.1.2. International and Industrial Cooperation

- Participation to an industrial project, CASSIOPEE, which aims at developing an automatic video surveillance platform in a bank context. This project involves a bank, a video system integrator, a telesurveillance operator (Crédit Agricole Bank Corporation, Eurotelis and Ciel), and INRIA;
- Participation to the European project AVITRACK in collaboration with Silogic S.A. Toulouse, CCI Airport Toulouse Blagnac (FR), University of Reading (UK) and ARC Seibersdorf research GMBH, Wien (Austria). The main objective of this project is to recognize the activities around parked aircrafts in apron areas;
- Participation to projects with ALSTOM, SNCF, CEA and INRETS, related to visual surveillance inside a train;
- Contract with Vigitec (Belgium) concerning videosurveillance technology transfer;
- Cooperation with Bull (Paris) on Intelligent Visual Surveillance;
- Cooperation with ENSI Tunis (Tunisia) in the framework of Franco-Tunisian cooperations. The topic of this collaboration is the distributed supervision of medical imagery program supervision;
- Cooperation with National Cheng Kung University (Taiwan), National Taiwan University (Taiwan), MICA (Vietnam), IPAL (Singapore), I2R (Singapore), NUS (Singapore), CLIPS-IMAG (CNRS-INPG-UJF Grenoble) and IRIT (Toulouse). This cooperation concerns both research on semantics analysis, reasoning and multimedia data, and applications of these results in the domains of e-learning, automatic surveillance and medical issues;
- Cooperation with ST-Microelectronics (Le Rousset Aix en Provence) on Intelligent Cameras.
- Cooperation with CSTB (Centre Scientifique et Technique du Batiment) to build a laboratoty aimed at long-term monitoring of a person at home.

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 (for instance 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 up 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 a suitable chain 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 lack of certain type of knowledge has to be compensated by control mechanisms, like powerful repair 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 [5] 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. [49]

3.3. Software Platform for Cognitive Systems

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

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

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

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

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

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

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

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

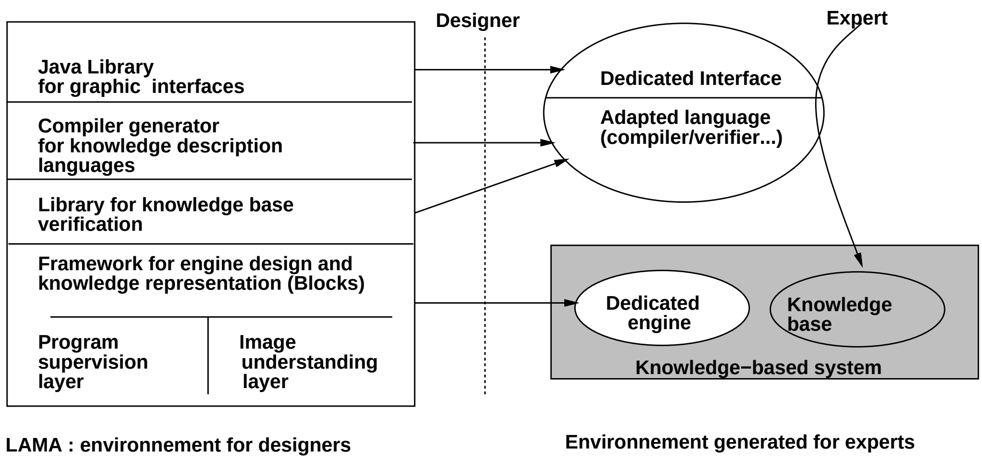


Figure 1. LAMA *architecture and tools for engine design, knowledge base description, verification, and visualization*

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

Framework for Engine Design

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

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

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

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

Engine Verification Toolkit

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

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

Knowledge Base Verification Toolkit

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

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

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

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

3.4. Automatic Interpretation of Image Sequences

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

Participants: Francois Brémond, Monique Thonnat.

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

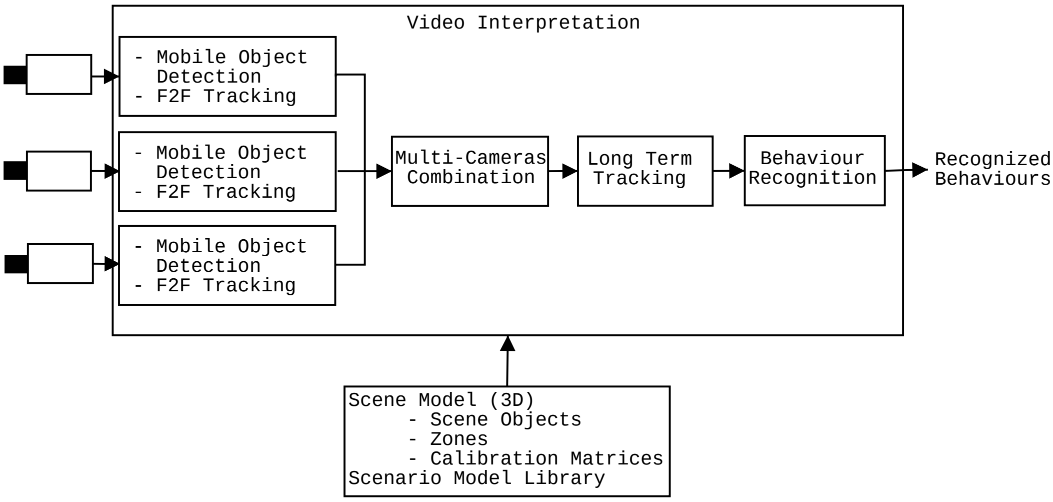


Figure 2. Overview of the interpretation of image sequences.

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*.

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 the 4 tasks (see figure2): 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) which is developed at the research group ORION. 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.

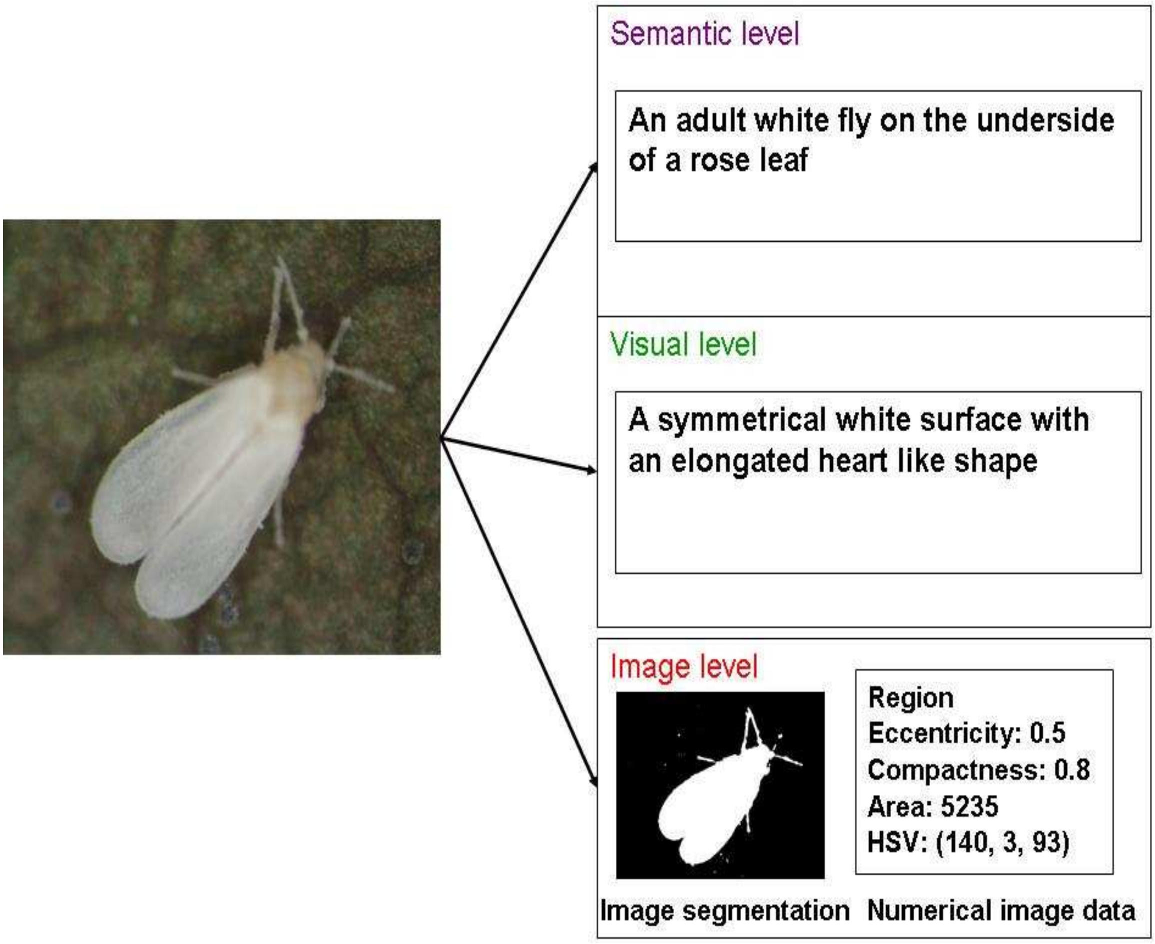


Figure 3. Illustration of the three abstraction levels of data corresponding to the sub-problems of semantic image interpretation. The image is a microscopic biological image.

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 semantic 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, 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 a priori knowledge and current context, 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).

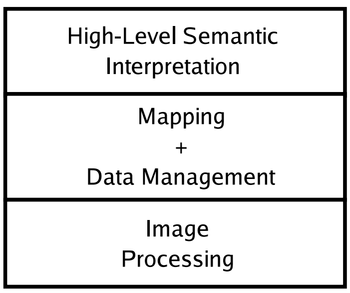


Figure 4. The problem of image interpretation is divided into three sub-tasks.

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

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*, *hydraulics*, *multimedia*, *pollen*, *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 two fold. 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 numerical calculus domain such as program supervision for numerical simulation. For instance, in order to model physical processes related to river hydraulics people use simulation codes which are based on the sampling of the simplified fluid mechanics equations (de Saint-Venant equations) that model stream-flows. This task, called model calibration, is close to program supervision and it has a predominant role in good modeling practice in hydraulics and in water-related domains.

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 [55], [52]. We are expert in this domain both concerning the image processing of galaxies field and theoretical models for morphological classification. This application is a reference to validate our models and software related to interpretation for complex objects understanding and to program supervision [53], [58].

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). More recently, we work 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. Pollen Grain Recognition

A part of Orion activities related to health care and environment are dedicated to automatic pollen grain recognition. We aim at providing palynologist tools for processing large amounts of data in a short time. For that purpose we use complex object recognition techniques which rely on image processing, knowledge based systems and pattern recognition.

The aim is to quantify the correlation between the environmental stress (so-called envi-contamination factor that is a combination of the concentration of allergens, the concentration of atmospheric pollutants including ozone and black dusts), and some indicators of the population health (medical data, hospitalization statistics, school and work absenteeism, medicine consumption). The task of the palynologist is to recognize the pollen particles present on a microscope slide, to give every pollen a name (family, genus, specie, group) and to finally produce a pollen spectrum for the given day. Not only because of the time required to obtain the pollen measurements from the sensor samples but also because possible human errors of counting and identifying the pollen grains can occur, it is of major interest to develop a system capable to recognize the pollen grains and to count them per types, this means to make possible an automatic evaluation of the atmospheric pollen concentration.

In this context, two main directions are taken : global counting of the number of pollen grains found on a slice and individual recognition of each pollen grain found on a slice. The second approach gives the accurate quantity of each type of pollen grain. Automatic global counting has been studied by using image processing techniques ([40] and [56]). Nevertheless, individual recognition remains of great interest. That is why the Orion team has been studying this problem since 1996 [48].

Due to the complexity of the different types of pollen grains, palynologist knowledge is taken into account. That is why a cooperative approach between image processing techniques and artificial intelligence techniques is proposed. Image processing techniques allow to extract and analyze relevant regions in images. Knowledge based systems use taxonomic knowledge to perform interpretation. They are also used for the supervision of image processing operators. It can be noted that similar approaches have been used in the past for galaxy, zooplancton and foraminifer recognition [54], [51], [43], [44].

The European project ASTHMA started in 1998 and finished in 2001. One of the goals of this project was to provide near real time accurate information on aeroallergens and air quality to the sensitive users. During ASTHMA, the Orion team was in charge of the conception and study of a platform dedicated to the recognition of 3D pollen grain images

4.5. 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 pesticial 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 mains 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.

4.6. 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 crisis. The ultimate goal is to cluster and/ or classify crisis of patients 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 a new 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

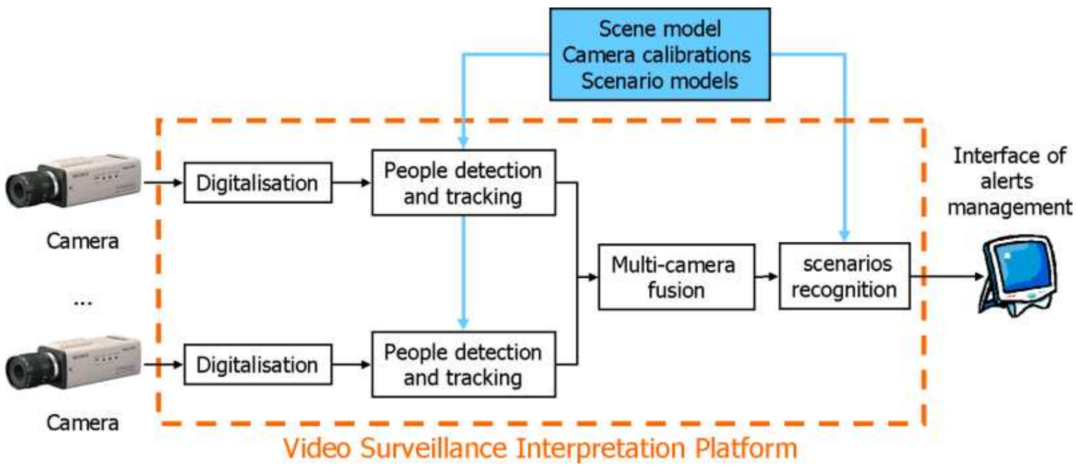


Figure 5. Components of the Video Surveillance Interpretation Platform (VSIP).

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 will be exploited by **Keeneo** the Start-up created in 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 a SME in videosurveillance **Timeat, Rennes** and to the Paris subway **RATP**.

6. New Results

6.1. Software Platform for Cognitive Systems

Participants: Nicolas Chleq, Celine Hudelot, Naoufel Khayati, Nicolas Maillot, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Monique Thonnat, Jean-Philippe Vidal.

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. We have also completed and tested our new engine for model calibration. Our major applications are medical imaging and river hydraulics. Our program supervision engine PEGASE+ *has also been applied to video surveillance and object tracking by a PhD candidate in Orion (see section 6.2.9). On the more practical side,* LAMA *which has first been developed under Linux, has been ported to the Windows system this year. We have continued the elaboration of an hybrid language to provide a front-end to the verification task. We also designed a language to express scenarios.*

6.1.1. Introduction

The efficiency in the design of knowledge based system is a major topic of Orion research. To this aim, the devoted platform LAMA provides a unified environment to design knowledge bases, inference engines and additional tools. 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.

The core of the platform is a *framework* of re-usable components, called BLOCKS (Basic Library Of Components for Knowledge-based Systems), 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. 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. Associated tools may ensure correct and safe reuse of components, as well as automatic simulation and verification, code generation, and run-time checks.

This year, we have addressed the problem of distribution of knowledge based system for medical imagery. We have extended the LAMA platform capabilities to cope with a classification task concerning the recognition of objects of interest in the complex images related to both pest detection on plant leaves and military scenes. We have completed the design of a knowledge language and an engine dedicated to model calibration task. We have continued to develop the engine verification toolkit. We design a hybrid language to express the generated code from the engine verification toolkit. Finally, we design a scenario language that allows to design scenarios in a modular and reusable way.

6.1.2. Distributed Program Supervision

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

In the framework of a co-directed PhD with Tunisia (ENSI Tunis) we worked on distributed program supervision for medical imagery. The objective is to study various distribution methods of program supervision knowledge-based systems for medical applications. Given distributed data, programs, and knowledge, the aim is to propose convenient and efficient models of distributed program supervision, first, to execute distant physician queries using existing distributed knowledge-based systems and, second, to favor collaborative consulting and constructing of knowledge-based systems.

Last year, we proposed a distributed architecture based on mobile agents for a program supervision system. This architecture includes a set of program servers, a set of knowledge servers, a supervision engine (PEGASE+), and a supervision server which plays an intermediary role between the physicians and the system. Moreover, a metadata warehouse allows to localize the various resources (programs, data, knowledge files, etc.) and to establish agent execution contexts (authentication, security, user interface).

As a second step, this year, we elaborated a scenario of use [30], [31], involving several kinds of mobile agents: a Supervisor agent, several Solver agents, and possibly several Evaluator agents. The Supervisor agent is tightly associated with the PEGASE+ engine. It triggers the creation of other agents, and facilitates communication among agents or between an agent and the engine. Under the control of the Supervisor agent, Solver agents migrate to distant program servers, execute needed programs there, collect the results, and bring back these results to the user or to the engine. If a program happens to need a local evaluation of its results, a Solver agent may create an Evaluator agent that will forward the evaluation result to the user or to the engine.

The scenario itself solves a request in medical imagery, namely osteoporosis detection. It involves processing bone radiographies by supervising the execution of several remote Matlab image analysis programs, developed in Tunisia.

This year, we designed and implemented a suitable mobile agent model. Our implementation relies on the simple, yet powerful Aglets [42] mobile agent platform, which is written in Java.

We are also in the process of extending the knowledge description language for program supervision (YAKL) to cope with distribution needs, particularly to handle the aforementioned metadata. We also plan to extend our agent model to increase agent autonomy by embedding local knowledge and reasoning facilities.

During this year, Naoufel Khayati has supervised two student projects in direct relation with his thesis at ENSI Tunis. The first project dealt with security of mobile agents as supported by the Aglets platform; this topic is particularly important for distributed applications in medical environment since major confidentiality problems may arise. The second project aimed at evaluating protocols for concurrent update of distributed knowledge bases. It should be noted that both projects are to be continued, extended and focused on our application.

We have also discussed with physicians taking part in the project. These discussions highlighted a number of suitable extensions, such as, for example, the possibility of interfacing with a statistical analysis tool (e.g., R [37]), the definition of a mechanism keeping the history of analyses and treatments to study the long term behavior of physicians, or the extension to multiple images, multiple collaborating physicians and multiple patients. Some of these features indeed constitute long-term goals.

6.1.3. Classification Task

Participants: Nicolas Chleq, Celine Hudelot, Nicolas Maillot, Sabine Moisan, Monique Thonnat.

Two projects in Orion required a classification approach to recognize objects of interest in complex images, one dealing with pest detection on plant leaves and the other with recognition of transport vehicles. Therefore, we decided to extend our LAMA platform capabilities to cope with the classification task. This year we identified the minimal common kernel both for expressing classification knowledge and for basic classification reasoning. We thus had to define, first, a knowledge description language and its parser, and, second, a classification layer in LAMA in order to build classification engines. The classification layer benefits from our previous experience with the LAMA implementation of a former simple classification engine (CLASSIC).

Based on the work of Nicolas Maillot 6.3.2 and Celine Hudelot 6.3.3, a classification algorithm has been designed and implemented in the LAMA platform. This algorithm solves the problem of classifying multiple objects with subparts constrained by spatial relationships and constraints due to the identification of the numerical descriptors extracted from the image and the symbolic objects to be found in the scene. It is based on a strict use of the ontology defined in [18].

A language called KRIL has been defined to implement this ontology and a parser dedicated to this language has been developed and integrated in the LAMA platform. This parser is able to generate executable code in C++ that realizes the intended object categorization.

An object categorization engine has been built using the LAMA platform. This engine is part of the cognitive vision platform described in section 6.3. It uses a priori knowledge and also machine learning techniques (e.g. Support Vector Machines). Examples of LAMA components used for building this engine are taxonomy, class, slot, knowledge base... This engine has been used for semantic image indexing and retrieval purposes [32]. Its full specification can be found in [18].

6.1.4. Model Calibration Task

Participants: Sabine Moisan, Jean-Philippe Vidal.

Numerical models can be used to simulate physical processes in order to understand past phenomena or forecast future behavior. In order to tune the models, computed results from past events have to be checked against field data measured during the corresponding events. Agreement between field data and computed results depends mainly on the values of some model parameters. The process of adjusting these values in order to reproduce field data is called *model calibration*.

We work on this task in the framework of Jean-Philippe Vidal PhD thesis, in collaboration with CEMAGREF Lyon and INPT Toulouse, dealing with river hydraulics and flood prevention. This thesis has been defended in March [19].

Until last year we used a program supervision approach to develop an automatic knowledge-based system dedicated to hydraulic model calibration. We have used our previous curve evaluation module to fully automate the evaluation process.

This year, we have completed the design of a knowledge language (OVAL) and an engine (HYDRE) dedicated to model calibration tasks. We thus obtained a calibration support system coupling an existing simulation code (namely MAGE, developed at CEMAGREF) with knowledge-based techniques relying on LAMA [22]. First, the system assigns roles to available data: either inputs to the simulator or reference data to be checked against the simulation results (this step usually requires user interaction). Second, the system triggers the simulation using MAGE, tries to match simulation results with reference data, adjusts the simulation parameters if necessary, and then loops automatically.

Developing the engine and its associated knowledge language have been quickly taken over owing to our previous formalizing of expert knowledge and methodology for model calibration and more particularly for river hydraulics calibration.

We have applied the calibration support system to calibrate the numerical models of two French rivers in various hydraulic situations, demonstrating the accuracy of the approach compared to traditional techniques. Moreover, compared to the later techniques, the system allows to capitalize expertise about good calibration practice and to provide reusable techniques to domains other than hydraulics.

6.1.5. Component Framework Verification

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

Two years ago, we defined both a dedicated language (BDL) to express BLOCKS component behavior and a *synchronous* mathematical model to give a semantics to BDL programs [50]. 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 concentrated on the verification of safety properties related to BDL programs and their preservation when applying BDL operators. The model-checking context we have chosen considers Kripke structures as models and a specific temporal logic ($\forall CTL*$) powerful enough to express safety properties. In this context, we know that there exists verification algorithms that take advantage of the structure of models. Practically, we have already interfaced with the NuSMV model-checker to verify behavioral substitutability properties and this particular model-checker fits our use of Kripke structures and $\forall CTL*$ temporal logic. An interesting point of our approach is the ability to support incremental design of engines from components. We show that bottom-up and top-down verification mechanisms hold and thus we can apply both compositional and modular mechanisms to achieve automatic proofs.

This year, we have completed our synchronous formalism [33] and interfaced with the latest version of the NuSMV model checker. However, efficiency in the verification activity requires to apply decomposition method automatically and we are looking for an algorithm for decomposing global properties into local ones. Finally, the implementation of a substitutability analyzer for knowledge based system engines is still underway.

6.1.6. Cognitive System Modeling

Participant: Annie Ressouche.

Cognitive System Design and verification

We have already adopted a mathematical based approach to get a safe re-usability of the LAMA platform libraries. In this approach we rely on a synchronous model to give a formal semantics to the library components (see 6.1.5) to ensure a safe usage of this library. To obtain this correctness result, we rely on results about synchronous models verification. Such an approach also allows to generate code for the applications with the help of LAMA libraries, and a devoted language to express the generated code would be useful. On another hand, many realistic applications accept different levels of description involving entities of various kinds which will be designed differently : some via a controller (synchronous), some via continuous dynamical systems. To cope with this feature, the language required to express generated code must be hybrid and must offer both a synchronous formalism to express the control part of applications and a formalism aimed at dynamical systems handling. Moreover, the hybrid language we are looking for must guarantee scaling up feasibility in order to be efficient in application specification and verification.

Thus, in collaboration with V. Roy (CMA Ecole des Mines de Paris) and D. Gaffé(CNRS and UNSA), we have begun to study the problem of synchronous language modular compilation. These two last years, we defined a synchronous language with modularity facilities and sound behavioral semantics in terms of process algebra model. We defined a circuit semantics that allows us to compile a program into a boolean equation system. The circuit semantics gives us a practical means to compile programs. Moreover, we defined a new method to sort equation systems. Our method is able to merge two sorted equation systems into a global sorted one without sorting again the whole equations. As a result, we can compile programs in a modular way. Practically, we have implemented a (prototype) tool to recognize and compile programs of our synchronous language. Moreover, program simulations are available through a dedicated graphic tool. This year, we improved both the sorting algorithm and the circuit semantics to reduce the size of the equation systems computed by the compiler. We also begun to prove that the circuit semantics is equivalent to the expected behavioral semantics with the COQ theorem prover (http://coq.inria.fr). The second point we begin to focus on is the extension of the synchronous language to the hybrid language we need to design real applications. To achieve this extension in a correct way, we rely on static analysis and abstract interpretation methods [38]. We search for an abstraction interpretation function that map an hybrid program to a synchronous one where data are abstracted into boolean values, to apply verification methods to the control part of hybrid programs.

Scenario Description

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, these last years Van-Thinh Vu has proposed a description language to express the knowledge of human behaviors and he proposed 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 de Paris), we defined a language to express scenarios in a modular way. Basic scenarios are sequences of events while general scenarios support hierarchy i.e call to sub scenarios either in a parallel way or in a sequential one. To illustrate the expressiveness of this language, we consider a scenario previously used by the Orion team to describe a person jumping over a barrier in a subway station. The scenario must recognize a person going from an entrance zone to a validation one, then jumping and finally going from validation zone to the hall. In our language, such a scenario is expressed as follows : first we define two sub scenarios *Jumping* and *ChangeZone* which are basic automata. Then , the scenario *Jumping_over_barrier* is the sequence of calls to these sub scenarios :

```
--- Declaration
type : PhysicalObjects, Person, Group, Crowd;
enum Zone      { entrance, validation, out, ticket_vending_machine, corridor, hall };
enum Equipment { wall, seat, trashcan, validation_machine, door };
function : LegsUp(Person) : boolean;
           SpeedIncrease(Person) : boolean;
           InsideZone(Person, Zone) : boolean;
scenario Jumping (Person p) {
  automaton
    state 0: SpeedIncrease(p) -> 1 ;
    state 1: LegsUp(p) -> 2 ;
    state 2 : empty
    initial:0
    final: 2
    end automaton
}
scenario ChangeZone (Person p, Zone z1, Zone z2) {
  automaton
    state 0 : InsideZone(p,z1) -> 1;
    state 1:  InsideZone(p,z2) -> 2;
    state 2 : empty
    initial : 0
    final: 2
   end automaton
}
main scenario Jumping_over_barrier (Person p, Zone z1, Zone z2, Zone z3) {
  ChangeZone(p,z1,z2) => Jumping(p) => ChangeZone(p, z2, z3)
constraint:
  is_entrance(z1) and is_validation(z2) and is_hall(z3);
}
```

Moreover, we give a semantics for scenarios in a synchronous model (synchronous models are naturally automata). By the time, we interfaced the EsterelStudio tool (http://www.esterel-technologies.com) to get easily scenario simulation and verification by model-checking technique. We also got generated C code but in the future we plan to generate code in our hybrid language. First of all, we plan to interface Van-Thinh Vu scenario recognition tool to benefit from an efficient recognition means for the sub class of temporal scenarios.

6.2. Automatic Interpretation of Image Sequences

Participants: Alberto Avanzi, François Brémond, Bernard Boulay, Frédéric Cupillard, Gabriele Davani, Florence Duchêne, Benoit Georis, Florent Fusier, Magali Mazière, Monique Thonnat, Alexander Toshev, Christophe Tornieri, Valéry Valentin, Marcos Zúniga.

The goal of this work is to automate the understanding of the activities happening in a scene. Sensors are mainly one or several fixed and monocular video cameras in indoor or outdoor scenes ; the observed mobile objects are mainly humans and vehicles. Our objective is the modeling of the interpretation process of image sequences and the validation of this model through the development of a generic interpretation platform. These techniques are applied in the framework of nine projects : three transfer actions Keeneo, Telescope3 and VIDEA, the European IST project AVITRACK, the European ITEA project SERKET, the SAMSIT Predit project (ministery of research) and the four following industrial projects: Intelligent Cameras (STmicroelectonics), PFC, CASSIOPEE, and RASV(SNCF). Moreover, an evaluation project ETISEO enables to evaluate these techniques and to compare them with others video surveillance results.

6.2.1. Introduction

The problem is the interpretation of the behavior of people acting in a scene; i.e. to find a meaning to their evolution in the scene. This scene is observed by one or several fixed video cameras. 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 “perceptual” module. The second sub-problem is to understand the scene content. To accomplish this process, we try to recognize predefined scenarios based on visual 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 modeling of the observed environment.

This year, we have refined our work on mobile object detection to remove false detection by modeling the background image using segments. We have improved our approach for the recognition of human postures to take into account the ambiguity and the uncertainty in the detection. We have also improved the tracking process by analyzing globally the 3D information on long-term periods. We have also focused on the evaluation of the video understanding platform at two levels, using both recorded videos and pilot sites for different applications. We have continued to use the program supervision framework to control the interpretation platform to easily build a dedicated interpretation system for new applications. We have also started designing new unsupervised learning techniques to help users to define scenarios models.

6.2.2. Advances in Person Detection on Transportation Vehicles

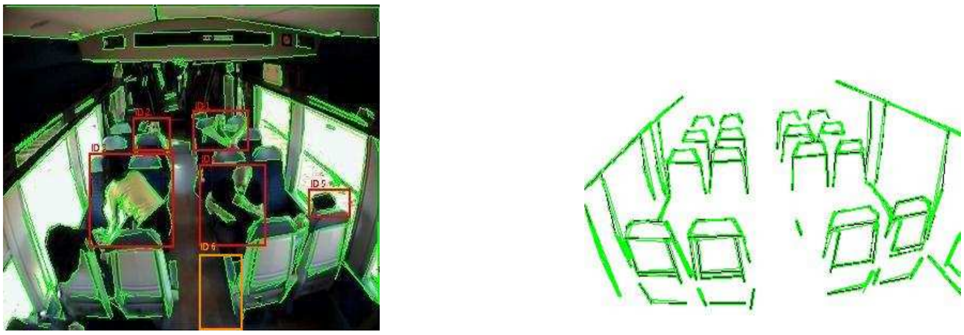
Participants: Gabriele Davini, François Brémond, Monique Thonnat.

Surface transportation vehicles, such as trains or buses, are subjected to relevant and abrupt changes of the internal illumination. In fact, they are not only affected by external light changes, as in outdoor scenes, but also their internal illumination may vary by passing beside external occluding objects, with parts of the interior of the vehicles becoming dark, while some other remain normally illuminated. In addition, internal objects or people may generate unpredictable shadows and when vehicles pass through tunnels, a transient phase occurs, during which internal lights replaces the external one.

In order to cope with these situations, person detection has to be strengthen with enhancement techniques able to limit the impact of illumination changes, especially when the detection is obtained by comparing the current image with a continuously-updated reference image.

We propose a gradient-based technique capable of improving person detection when dealing with such critical cases.

First, we compute gradient vectors on the current image. Then, the points whose gradient vectors have a relevant magnitude and are pointing almost in the same direction are selected to form line segments. We obtain a set of segments describing the main sharp edges of the current image. The same procedure is applied over



(a) Current image segments (b) Mask image segments

Figure 6. Segments on the current image and on the mask image.

a so-called mask image, which consists in a capture of the observed scene under standard and homogeneous illumination conditions and with no external object or person. As a consequence, the segments computed on the mask image are only generated by the static context objects of the scene, such as the seats or the windows.

By comparing the two sets of segments computed over the two images, we are able to recognize false detections due, for example, to shadows. We call *visible* segment, a segment which is found both in the current image and in the mask image, thus meaning that the region surrounding the segment is correctly detected. We can recognize false detections, because the segments falling into one of these regions are only visible segments, i.e. all segments belong to context objects. The same holds for *ghosts*, which are those false detections generated by people who start moving again after a long period during which they hold still.

Segment analysis is also useful when merging moving regions. This process consists in reconstructing the complete outline of an object or a person that a poor detection has split up into multiple, not connected moving regions. By using segment analysis, it is possible to avoid the fusion of two moving regions which actually do not belong to the same person, because the final region given by the merge of two regions must not contain new visible segments.

Segment analysis has been applied over multiple video sequences with different scenarios; an example of its application is provided in figure 6. We can notice in current image that the woman surrounded by blob 3 occludes parts of the chairs in front of her; thus, the segments of these chairs are also occluded and since other segments are found in the blob, this confirms that the detection was correct. On the contrary, blob 6 is discarded since its segments are the same in the mask and no other segment is found in it. Finally, we notice that, due to segment analysis, blobs number 1 and 4 have not been merged, because a visible segment is found between the two.

In conclusion, we can claim that our gradient-based technique has significantly improved the detection by preventing a number of errors deriving from the effects of shadows or moderate illumination changes. This implies that a sufficient level of contrast is needed in order to detect sharp edges and thus segments. As a consequence, the extreme cases where illumination changes lead to a complete saturation of the image, with very low contrast, may not be treated with this technique.

6.2.3. Classification of Lateral Forms for Control Access Systems

Participants: Binh Bui, François Brémond, Monique Thonnat.

We present in this work a real-time system for shape recognition. The proposed system is a video and multi-sensor platform that is able to classify the mobile objects evolving in the scene into several expected categories. Our goal is to design a system with very high recognition rate complying with real-time constraint.

To achieve this goal, we have conceived a device combining a passive static camera and a set of lateral active sensors. Cameras are often static in visual surveillance network to get a robust low-level detection of mobile objects. The lateral active sensors are very useful to separate people entering the access control site. The real-time constraint is very challenging as it implies that the solutions should be kept with a maximal computing time. However, few existing systems have been successfully applied to real world applications due to a large variety of video interpretation issues (e.g. motion detection and tracking are often uncertain and incomplete) and due to strong requirements to obtain a real-time, efficient and robust system. Moreover, most of existing systems focus only on mono camera processing therefore they cannot take advantages of other information sources.

Thus, we propose a system that is able to detect and classify people and objects with very high recognition rate and with real-time constraint. Our approach consists in applying Bayesian classifiers for shape recognition to handle the uncertainty accurately. The key of the recognition method is to compute mobile object properties thanks to the active and passive sensors and then to use Bayesian classifiers. 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", "two adults close to each other" and “suitcases”. The recognition module sometimes misclassifies a child with a small suitcase and vice versa due to the similar appearance.

The recognition module has been tested in two stages: a stand-alone experimentation on recorded image sequences (i.e. test offline) and an live experimentation in interaction with the kernel of an existing access control device used in subways at RATP. In the live experimentation, the recognition module runs on a PC (Pentium IV 2.8 GHz, 1GB memory, Linux) and receives a video stream at 25 images per second. The maximal time for processing one image is inferior to 35ms. The real-time constraint is then satisfied. To train the Bayesian classifiers, for each class "adult", "child", "suitcase", we used about 300 frames as training data and about 1000 frames for testing. For the class "two adults close to each other", at the present time, we have only 32 frames in total to represent this class. For this class, we used 15 frames as training data and 17 frames for testing. In the stand-alone stage, the results are very promising. A large majority of mobile objects have been correctly recognized with a high degree of membership (table 1). More than 94% of adults, children and suitcases are correctly recognized. More precisely, for the adult class, the true positive is 98%, the false positive is 1% and the false negative is 2%. The true positive for "two adults close to each other" is about 73% due to the lack of training data in the learning phase.

Table 1. More than 94% of adults, children and suitcases are correctly recognised.

Mobile Object	True Positive	False Positive	False Negative	Frames used for testing	Frames used as training data
Adult	98%	1%	2%	1102	327
Child	94%	3%	6%	1050	295
Two adults close to each other	73%	0%	27%	17	15
Suitcase	95%	2%	5%	1008	305

Almost all the potential errors in the frame to frame classification have been corrected by the frame to frame tracking. The recognition result depends on training videos used in the learning phase. To obtain better results, we should enrich the training data for each class. For example, for the class "adult", the training data should include large variety of persons (fat, thin, tall, short, adult in summer/winter clothes, etc). Since realistic training data cannot include all varieties of mobile object classes and shapes, the first next step will consist in studying supervised and non-supervised machine learning techniques in order to learn dynamically new classes of mobile objects. Finally, for the system to be more robust, the second next step will consist in studying the system autonomy. For example, the system should be able to detect failures (sensors or camera breakdown, change of light) and set up a degraded operation mode according to the resource available. The objective will be to have a system that can reconfigure itself dynamically and autonomously. This work has been presented at [27].

6.2.4. Human Posture Recognition

Participants: Bernard Boulay, François Brémond, Monique Thonnat.

This year, we have continued previous work on human posture recognition in video sequences to deepen the analysis of the algorithm performances. We have completed the set of studied postures and have studied new silhouette representations.
We have added lying postures in our set of postures of interest. This set is now composed of four general posture categories and eight detailed posture subcategories:

- standing postures : standing with one arm up, standing with arms along the body and T-shape posture,
- seated postures : seated on a chair and sitting on the floor,
- bending posture,
- lying postures : lying with spread legs and lying with curled up legs.

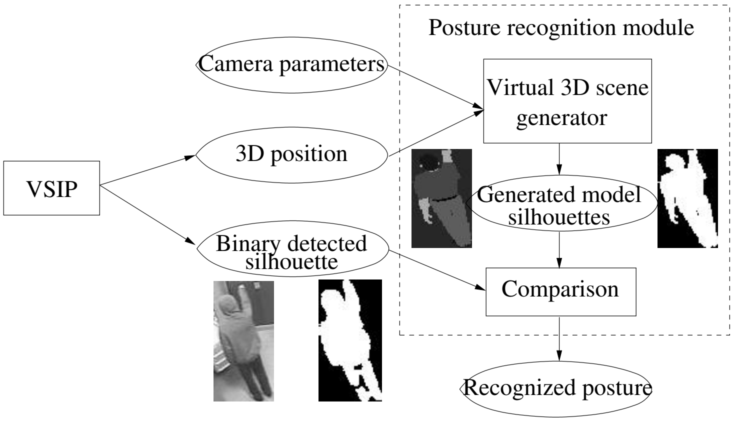


Figure 7. Simplified scheme showing the posture recognition approach

A simplified scheme of the approach is given in figure 7. The first step in posture recognition is to detect persons in videos. The detected persons are represented by their binary silhouettes. The 3D position of the persons are computed by the VSIP platform using camera calibration.

Then the proposed human posture recognition algorithm compares the detected silhouette with generated model silhouettes and their 3D positions to determine the posture of the detected person. The generated model silhouettes are obtained by projecting the corresponding 3D human model on the image plan, using the 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 3D human model silhouettes are compared with the detected silhouette depending on the chosen silhouette representation. Three different representations are studied: Hu moments, Skeletonisation, and Horizontal and Vertical (H.& V.) projections.
Table 2 shows the validation of our approach on real videos for more than 600 frames with associated ground truth by treating each frame independently. A prototype has been setup at STMicroelectronics (Le Rousset) for demonstration.

Table 2. General posture recognition rates (%) for the different approaches tested on real data

	Standing	Seated	Bending	Lying
(H. & V.) projections	96	86	87	92
Hu moments	68	73	27	35
Skeletonisation	93	68	88	65

The (H. & V.) projections gives the best recognition. The algorithm gives satisfying results using only one camera. An accurate analysis has been made to determine the quality of the algorithm performances in cases where videos were taken in ambiguous/optimal view points and in cases where silhouettes are good/badly segmented. We have shown that the approach is efficient to discriminate the 4 general postures : standing, sitting, bending and lying from any view point. This approach manages also to recognize detailed postures except in the cases where postures are visually ambiguous. The algorithm is able to deal with badly segmented silhouettes where some body parts are missed. The algorithm is relatively fast (5 frames per second). This frame rate is sufficient since we only need to recognize postures on few key frames to help the behavior analysis.
Now, we are working on using temporal information to improve the approach. First it will solve partially ambiguity problem. Second it will enable us to detect representative postures in the video. And finally posture tracking will help human behavior understanding.
This work has been published in the ICDP symposium [26].

6.2.5. *Global Tracking*

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

(a) Before global tracker (b) After global tracker

Figure 8. On the left, a Loader vehicle is detected as several mobile objects (over detection) while on the right, the Loader vehicle is correctly tracked as one mobile object.

The VSIP platform developed in Orion team is based on two main modules: one dedicated to vision and tracking, and one dedicated to scene understanding. While the scene understanding module needs accurate

input data to recognize behaviors, some incorrectly tracked objects are sometimes present in the vision and tracking output. We thus have developed a Global Tracking module which takes as input the a priori knowledge of the observed environment (static and dynamic context) and the tracked objects and sends corrected tracked objects to the scene understanding module. The correction of the tracking objects is guided by user defined rules.
The Global Tracker is a generic prototype based on rules which allows to quickly and easily adapt it to a new application. Each incorrect tracking result is analyzed and a rule is used to correct it. This rule can be a completely new rule, or a default rule (defined for another situation or application) overloaded with a new set of parameters.
For the Airport Apron Monitoring European project (AVITRACK), we have so far solved two main problems adding new rules in the Global Tracking:

- The first is the loss of tracked objects which occurs when a vehicle stays for a long time at the same place. The defined rule manages the lost vehicle when it is parked in a specific zone and ends when the vehicle restarts (the newly detected vehicle is linked with the stopped one) or after a predefined period of time (the stopped vehicle is removed);
- The second problem appears when a vehicle is detected as several smaller objects (over detection case, as shown on Figure 8). It occurs when a vehicle is slowly moving in a small area. The correcting rule merges the mobile objects (as shown in Figure 8) into the real vehicle. To achieve that, the rule takes into account the predefined vehicle models, and the 3D position and motion of the mobile objects.

The developed Global Tracker is currently used in the AVITRACK european project and will be integrated to others projects which use the VSIP platform.

6.2.6. *Complex Activity Recognition*

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

In the previous period, the VSIP platform has been adapted to Airport Apron Monitoring. The focus was on Simple Activity Recognition (operations involving a single vehicle and/or a person) occurring around parked aircraft. This initial work provided good and encouraging results. Thus, the new challenges for this year were to extend the set of basic and simple video events and deal with Complex Activity Recognition (lasting operations involving several vehicles and/or people).
Concerning basic video events, we have now modeled and recognized 21 events occurring on the airport : 10 primitive states (e.g. a vehicle is located inside a zone of interest), 5 composite states (e.g. a vehicle is stopped inside a zone of interest), and 6 primitive events (e.g. a vehicle enters a zone of interest). These video events are the basic ones which are used to model complex scenarios.
We have also defined new models of simple video events. Simple video events are dedicated to activities involving a single vehicle and/or a single person. They have been modeled and recognized for the following operations :

- “Aircraft Arrival Preparation” : 8 scenarios, describing the Ground Power Unit vehicle arrival and the deposit by a worker of the chucks before the aircraft arrival. This operation has been recognized on 10 video sequences with duration of about ten minutes;
- “Refueling Operation”: 8 scenarios, involving the Tanker vehicle and one person refueling the aircraft. This operation has been recognized on seven sequences of twenty minutes duration;
- “Aircraft Arrival”: 3 scenarios, describing the Aircraft arriving at its parking position. This operation has been recognized on two sequences of about ten minutes;
- “Tow Tractor Arrival”: 3 scenarios, describing the arrival of the Tow Tractor vehicle before towing the aircraft before its departure. This operation has been recognized on two sequences of about five minutes;

- “Conveyor Arrival”: 3 scenarios, describing the Conveyor Belt vehicle arrival before the back door baggage unloading. This operation has been recognized on one sequence of about ten minutes.

We have also 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. Up to now, we have been able to recognize the following operations :

- “Unloading global operation”, involving the Loader vehicle (3 scenarios), the Transporter vehicle (3 scenarios), and three zones of interest. This operation has been recognized on two sequences of about ten minutes;
- “Unloading detailed operation”, involving the Loader vehicle (3 scenarios), the Transporter vehicle (3 scenarios), one person, one container (2 scenarios) and four zones of interest. This operation has been recognized on two sequences of about ten minutes.

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 “Front Unloading Operation” involves 9 physical objects :

- 2 persons : the Handler of the Loader who drives the vehicle, and a worker who unloads the baggage containers one by one;
- 3 vehicles : the Loader vehicle, the Transporter vehicle and a container (considered as a vehicle due to its motion) ;
- 4 zones of interest : the ERA zone, the Loader area, the Transporter area (a dynamic zone which is added automatically to the static context when the Loader is detected as parked), and the Worker area which is the zone where the containers are unloaded from the Loader.

This complex scenario (composite video event) is composed of 3 components (more basic scenarios) :

- the Loader arrival (also composed of 3 sub-scenarios);
- the Transporter arrival (also composed of 3 sub-scenarios);
- the Manipulation of containers by the Worker in charge of the containers (also composed of 2 sub-scenarios).

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 [23], [25], [34], [29], [24], [36] and also demonstrated in CVPR 2005 (San Diego, CA, USA, June 20-25 2005) and ICCV 2005 (Beijing, CHINA, October 15-21, 2005).

A remaining task is to evaluate on live (directly at the airport) the whole operational system to test the robustness in case of challenging environmental conditions (e.g. night, fog, sunset...). This task is planned to be done at the beginning of 2006.

6.2.7. Evaluations of the VSIP Platform

Participants: Alberto Avanzi, Frédéic Cupillard, Magali Mazière, Christophe Tornieri, François Brémond, Monique Thonnat.

The ORION video interpretation system (called VSIP) has been evaluated within three industrial projects : Cassiopee (video surveillance of bank agencies), VIDEA (recognition of specific human behaviors) and Telescope 3 (car park video-monitoring).

Cassiopee Project

Within Cassiopee, the 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, (see Figure 9). 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 reasons : the room was dark (people did not turn the light 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 next step is to evaluate the system in live conditions for 6 months. This evaluation is going to be performed directly by the company in charge of the remote surveillance.

Figure 9. Two people are detected by the system in the ETS area.

VIDEA Project

Within VIDEA, the system has been designed to deal with two applications : (1) "directional people counting in buildings" and (2) "violence detection in urban areas".

The first application consists in detecting for each person going over a specific area inside a building, his/her origin (where he/she is coming from) and his/her destination (where he/she is going to). The second application aims at detecting suspicious fighting behaviors occurring in urban areas. The first application has been integrated and evaluated within two sites (a corridor and an airlock) located in the building of a Belgium company. Both sites, each equipped with one camera have been evaluated by INRIA and end-users using recorded videos and live acquisition. The recorded videos correspond to 14 hours associated with ground Truth. Concerning the evaluation for the airlock camera, origins/destinations for one single person crossing the airlock were correctly detected 16 times out of 17, origins/destinations for more than two persons crossing the airlock were correctly detected 24 times out 25 and no false alarm (false presence detection) was detected.

Concerning the evaluation for the corridor camera, origins/destinations for one single person crossing the corridor were correctly detected 68 times out of 72, origins/destinations for more than two persons crossing the corridor were correctly detected 27 times out 45 and three false alarms (false presence detection) were detected. The evaluation performed on live videos gave similar results for both cameras. The main errors were due to a lack of robustness when several persons are significantly crossing each other (this situation occurred often in the corridor).

The second application has been integrated and evaluated within 5 cameras located in a Belgium town, three observing a square and the two other ones a pedestrian street mixing pedestrians and vehicles. Two cameras for each site have been evaluated by INRIA using recorded videos. The recorded videos correspond to 8 hours (mixing actor played sequences and normal life situations) associated with ground Truth. Results for both cameras were similar. Violence behaviors were correctly detected 15 times out of 26 for the pedestrian street camera and 16 times over 26 for the square camera. In total, 3 false alarms (false violence detection) were detected. Two methods to recognize two types of violence actions (observed in the video sequences) were developed : "erratic trajectory of a group of people" and "person on the ground close to a group". However, another remaining part of violence actions (observed in the video sequences) were very brief actions without erratic trajectories or persons on the ground. For this type of violence, a method based on motion detection and analysis inside groups of people has been launched but was not finished by the end of the project. False alarms were mainly due to erroneous integration of person(s) in the reference image. In conclusion, performances of both applications were satisfying regarding the ambitious requirements defined in the initial specifications. Moreover, the application (1) "directional people counting in building" has been evaluated very positively by end-users and its deployment has already started within other sites. Concerning the application (2) "violence detection in urban area", the next step is to evaluate the system in live conditions.

Telescope 3 Project

Within Telescope 3, the system has been installed and evaluated in a public car park in Paris using recorded videos in a first time and live acquisition in a second phase. The system is currently non-stop running as part of the surveillance system of the parking, in live conditions. The system is able to detect three scenarios:

- human presence in car access lanes;
- affixing messages to automatic payment machines;
- opening the door of automatic payment machines.

The system is fed with 3 surveillance cameras, two indoor and one outdoor. The evaluation performed on recorded sequences (3 sequences lasting 24 hours each, and containing no true positive, and 6 short sequences containing several acted scenarios) gave the following results:

- for human presence, 97% of good detections, 5% of false positives;
- for affixing messages, 95% of good detections, 10% of false positives;
- for door opening, the evaluation is ongoing.

The high rate of false positives on affixing messages is mainly due to the integration in the background of some parts of the body of people standing for a long period in front of the payment machine (“ghost integration”). Some algorithms for ghost removal are currently being added to the system. These experiments have been detailed in a journal [20] and in a conference paper [28].

6.2.8. Unsupervised Event Learning and Recognition

Participants: François Brémond, Florence Duchêne, Monique Thonnat, Alexander Toshev, Marcos Zúniga.

Unsupervised Primitive Event Learning and Recognition

We have developed a new method for unsupervised frequent primitive events recognition, based on incremental concept formation models. These models allow to incrementally construct hierarchies of concepts based on incomplete or uncertain data, updating the hierarchy structure with each new arrived data instance and allowing, at the same time, the classification of the instance, based on the frequency of previous ones. In this work, a concept corresponds to a primitive event and data correspond to visual attributes of tracked objects. More generally, methods based on incremental concept formation models consist in, given a sequence of instances and their associate attributes, finding:

- clusters that group those instances into categories;
- a prototype for each category that summarizes its instances;
- a hierarchical organization of those categories.

The method under development is based on a method of incremental concepts formation called CLASSIT, which basically consists on updating a hierarchical structure using four operators (merging, splitting, deletion and addition) which are selected according to the operator which gives the best category utility increment. The category utility measure is based on psychological studies, maximizing intra-class similarity and inter-class differences (Gluck & Corter 1985). We use this method for classifying the most frequent states on a video sequence, where each input instance is represented by the detected object class name (e.g. person, vehicle) and a subset of its attributes (e.g. posture, trajectory, location) called group, where a group represents a particular primitive event recognition objective. The idea is that the user will not need to pre-define all the possible primitive states to be found, but to define the groups of attributes of interest for a given primitive event recognition objective (e.g. postural changes of interest) and what is considered as a significant change for each attribute in a group.

This work is at testing stage of the hierarchy generation algorithm and in the acquisition of significant test sets for validating the algorithm performance for finding frequent primitive states on a video sequence. The algorithm is currently able to classify instances representing physical objects into a hierarchical representation of the more frequent attributes for a group representing the resulting primitive events for a given recognition objective (e.g. posture, global location, location relative to other objects). It is also able to recognize the current primitive states (one for each object class group) for a given object instance (e.g. for a person, to recognize the state "crouching" of the group "posture" and to recognize the state "in the parking place" for the group "global location"), according to the information extracted from the previous object instances.

Future work will study the construction of links between the generated primitive states on the hierarchies generated from each group, which correspond to the definition of primitive event in the sense that it represents the passage from a primitive state to another one. This approach will allow to learn a coherent hierarchical representation of states and events and be able to classify new instances to recognize the occurrence of primitive events for a given object instance (e.g. for a person, to recognize the event "stand up" from the change of the state "crouching" to the state "standing", both states of the group "posture"), based on the frequency of the previous ones and in an unsupervised and incremental way, meaning that the method will be able to learn on-line.

Unsupervised Behavior Learning and Recognition We have also developed an algorithm for detecting frequent behaviors (composite events) in videos and generating models of such behaviors. It uses as input a set of simple events recognized in the videos and discovers all reoccurring patterns of such events. The approach is based on a data mining algorithm called Apriori which detects frequent patterns in data sets employing the principle that subpatterns of frequent patterns must also be frequent. Therefore, starting with short patterns this method gradually creates longer patterns by combining frequent shorter ones. As a result, this heuristic leads to an efficient computation. The detected frequent activity patterns are used to generate models of these activities suitable for their recognition. In this way, event models are learned in an unsupervised way.

The main contribution of this work is the application of the Apriori algorithm in the video interpretation domain where, in contrast with the traditional data mining applications, we have to cope with data uncertainty and noise. For this purpose, the different steps of the Apriori algorithm were modified to handle uncertainty. The major issue is the choice of an appropriate similarity measure which evaluates the degree of similarity between activities. Due to the usage of this similarity measure several issues arise. First, the principle of the initial Apriori algorithm does not hold and, therefore, a new principle called Weak-Apriori property was formulated which takes into account the similarity while not decreasing the efficiency of the approach. Second, the simple counting of identical patterns is not possible in the case of behavior analysis where we never have identical patterns. As a solution, an entropy-based agglomerative clustering technique was proposed which forms classes of similar patterns. Additionally, for the video interpretation domain a concrete similarity measure is proposed, based on notions concerning analogy reasoning, developed in the cognitive sciences.

This work shows that the above data mining method can be successful used to process moderate amounts of data. Additionally, besides the domain-dependent similarity, the method does not need further domain apriori knowledge or labeled training data. One drawback of the approach is the necessity to specify for each domain a new similarity measure. Additionally, the conformity of each new similarity with the Apriori-based framework must be shown. The approach depends also on the generic primitive event library which should be also manually defined.

The developed system was tested on both data from the parking lot monitoring domain and synthetic data. In the experiments the algorithm has detected successfully the expected composite events. In certain cases it has detected also events which slightly differ from the expected ones. These problems are caused by noisy and partially incorrect input events. Another reason for the problems is the fact that the similarity measure does not model perfectly all relations in the real world. This work will be presented in [57].

Unsupervised learning of recurrent behaviors from multivariate time-series

We investigate the issue of mining multivariate time-series for learning usual patterns. The idea is to automatically acquire some knowledge to improve decision-making in surveillance systems. Most monitoring purposes indeed require analyzing large sets of temporal data, collected by several types of sensors or information sources. Dealing with complex systems, where all possible critical situations may not be easily described, we then aim at learning recurrent usual behaviors. Any deviation from this behavioral profile is considered as an unexpected situation.

We are particularly interested in learning patterns of human behavior from camera-based tracking of human body. VSIP video interpretation platform especially allows to get, along time, low-level information about the persons observed within a scene, like their 2D position (x,y) and size (height, width). These multidimensional data are used as input for learning recurrent behaviors at several temporal scales: from usual basic states and events, to composite events representative of activities.

A hierarchical learning approach is used for activity monitoring from low-level data, involving the following successive steps:

- Representation/ abstraction of the raw time-series, to highlight basic states and events;
- Feature mining, to select the most meaningful features – that is, subsequences – considering the purpose of frequent pattern identification;
- Clustering, to perform the unsupervised classification of the selected features into frequent patterns.

The representation step is really important because it at least partly governs the level of details of the analysis. We especially focus on proposing robust methods for representing multidimensional time-series into a one-dimensional symbolic sequence, made-up of usual states and events. One promising method is based on the standard deviation of the original time-series observed in a fixed-length sliding window. That allows to discriminate between subsequences representative of states (spatio-temporal property stable on a time interval) and events (change of state). Subsequences of events are further segmented to highlight basic and recurrent ones, and get a synthetic and meaningful symbolic representation of the original time-series.

This approach especially requires to define a similarity measure appropriate to the comparison of multivariate time-series of different lengths, so that events and states are properly identified and clustered. We use the Derivative Dynamic Time Warping (DDTW) method [41] for matching pairs of points between two subsequences, and then computing a distance measure. This matching technique minimizes the cumulative distances between "shapes" of the adjacent elements.
Another crucial point is to define an appropriate cost function for segmenting sequences representative of successive events. The idea is that further clustering of the delimited subsequences provides meaningful classes, especially in terms of their size.

Preferred monitoring applications are related to medical domains. One application is 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 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. The general purpose concerns the issue of monitoring and learning about person activities at home, using autonomous and non-invasive sensors. The aim is to detect the sudden occurrence of worrying situations, such as any slow change in a person frailty. The analyses are based on both data collected from sensors installed at home, and a priori knowledge about the monitored person condition obtained from a standardized gerontological evaluation (EGS). In the context of using the VSIP platform, the video interpretation results are used to define variables like the person trajectory, velocity, or agitation. We are particularly interested in automatically defining frailty indicators from the analysis of their temporal evolution. Combining several types of data (physiology, environment, activity) and sensors (wearable or unobtrusive) consolidates information extraction and decision making.

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 (Pr. P. Chauvel, La Timone). One purpose in terms of activity monitoring is to model behavior changes for a given patient when in crisis. The ultimate goal is to cluster and/ or classify crisis of patients groups, so that the localization of the brain lesion of a given patient is more easily determined.

6.2.9. Techniques for Easy Configuration of Video Understanding Systems

Participants: Benoît Georis, François Brémond, Monique Thonnat.

This work at the frontier of program supervision (3.2) and automatic interpretation of image sequences (3.4) has been done in the context of a thesis [16]. Video understanding is here defined as the task of recognition of predefined event models (e.g., human activities) in a given application domain starting from a pixel analysis up to a symbolic description of what is happening in the scene viewed by cameras. Although various approaches and techniques have already been proposed for understanding video sequences, this task remains a complex problem regarding the lack of performances when facing challenging environments (e.g., a cluttered scene) and the lack of reusability when a system has to be deployed on a large scale with minimum human interventions.

In this context, we propose a complete framework to conceive supervised video understanding platforms. Such platforms ease the creation and configuration of video understanding systems. Several issues need to be addressed to provide a correct configuration:

1. the ability to choose, among a library of programs, those which are best satisfying a given request;
2. the ability to dynamically adapt programs and parameters to environment changes;
3. the ability to evaluate performances and to continuously guarantee a maximum performance rate which is satisfactory regarding end-user requirements.

In this framework, a knowledge-based approach for the supervision of video processing programs enables to externalize both the control and the knowledge of programs. We propose a model for program control which is generic in the sense that it is independent of any application. The need of having a formalism for knowledge representation is demonstrated for each type of knowledge: knowledge of the application domain, knowledge of the scene environment and knowledge of video processing programs. In addition, the framework provides a methodology for the evaluation of system performances. This methodology proposes a video sequence characterization which guides the selection of video sequences which are used for testing a system. In order to perform the evaluation, the methodology also proposes a set of metrics to compare results with reference data. Thanks to this evaluation methodology, video processing experts are able to acquire expertise of the use of programs. Finally, the framework allows the use of learning techniques where knowledge is hardly available. The proposed framework has been used to conceive the supervised video understanding platform illustrated in figure 10. This figure shows the way the platform is used to build particular video surveillance systems. First, the experts fill in the knowledge bases by using the appropriate formalisms. A learning tool enables to complete the video processing program knowledge base in case the video processing experts encounter difficulties to express the knowledge. Then, these knowledge bases are assembled together with the control component and the library of video processing programs in order to obtain a particular system, which can be evaluated. Finally, experts can improve the system by iterating the design (i.e., analysis of evaluation results).

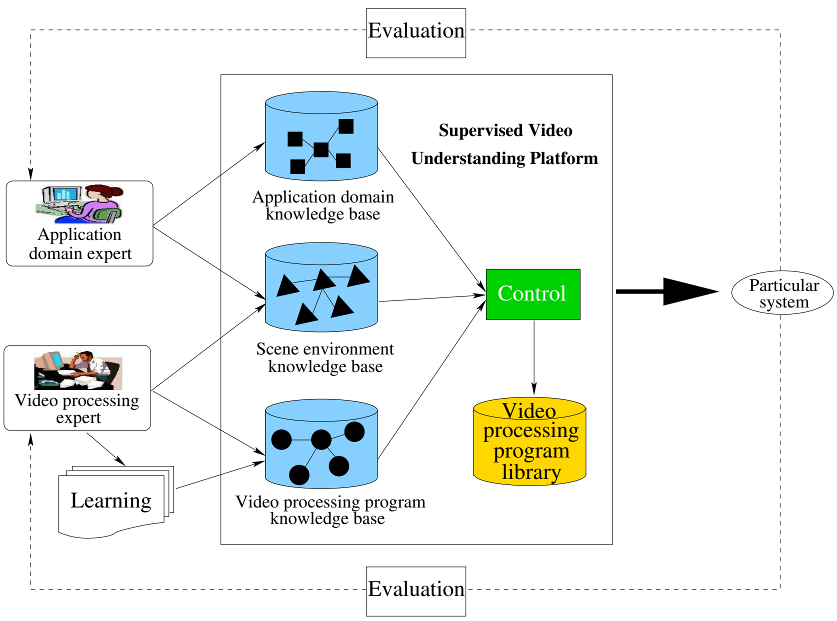


Figure 10. A supervised video understanding platform

The first component is a library of video processing programs. The next three components are three knowledge bases containing respectively the knowledge of the application domain (i.e., what the system is expected to do, the end-user goal, such as counting people), the knowledge of the scene environment and the knowledge of video processing programs. Each of these knowledge bases is created by different persons (e.g., application domain expert, video processing expert) using a particular formalism. The last component is a reasoning engine which is intended to control the programs by using the information contained in the different knowledge bases. The control can be viewed as an intelligent assistant for a system developer when building a system or as a process which guarantees the reactivity of the system and the property of adaptability. To use this supervised platform, the different domain experts (e.g., end-users, application domain experts or specialists in video analysis) first create three knowledge bases which are called *minimal knowledge bases*. Then, all the platform components are assembled together by a compilation operation. The output of this operation is a particular system. The evaluation component can then be used to evaluate the system performances. In case the video understanding performances are not sufficient, the experts are able to improve the system by either developing and adding new video processing programs or by augmenting the *minimal knowledge bases*.

Thanks to this supervised platform, six systems have been built and configured with the PEGASE+ engine of the LAMA platform (see 6.1). These systems are characterized by four properties: adaptability, reusability, efficiency, real-time processing. The current knowledge base contains 25 primitive operators, 13 composite operators, 104 initialization criteria, 16 choice criteria, 9 optional criteria, 25 assessment criteria and 7 repair criteria.

Future works will consist in obtaining a complete and flexible evaluation tool (in the context of the ETISEO evaluation project [35] and in formalizing the learning model of the framework.

This work has been accepted for presentation at the International Conference on Computer Vision Systems 2006 (ICVS'06) [39].

6.3. Cognitive Vision Platform

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

This year, research activities on semantic image interpretation have been continued. A cognitive vision platform is now partially operational. The platform is based on reasoning, learning and image processing mechanisms as well as knowledge 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 image interpretation, we rely on some aspects of cognitive vision: knowledge acquisition and representation, reasoning, machine learning and program supervision. We aimed 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 a distributed architecture based on three highly specialized modules : a semantic interpretation module, a visual data management module and an image processing module. The cognitive vision platform is described in more details in 3.5.

This year, the cognitive vision platform has been improved and is now operational: first, we focus on knowledge acquisition and learning. We propose an ontology based acquisition process to guide knowledge acquisition and we use machine learning techniques to learn visual concept detectors. Moreover, to improve the image processing module, we develop a generic and reusable system for image segmentation based on a supervised learning scheme. Second, we focus on semantic image interpretation. We rely on the ontology based knowledge acquisition phase mentioned previously to achieve a high-level semantic interpretation. We also design a module in the cognitive vision platform dedicated to visual data management. Moreover, relying on program supervision approach, we manage different image high-level contexts in a generic way. Finally, we apply these new improvements both to the detection of plant diseases and to solve the problem of semantic image indexing and retrieval.

Four PhD theses related to this topic are currently on the way or have been defended in 2005:

- A PhD started in October 2005 by Lan Le Thi in cooperation with the international research center MICA in Hanoi, Vietnam. The title of this PhD is "Image Indexing and Retrieval by Semantic Content".
- A PhD was started in September 2004 on the problem of adaptive image segmentation by Vincent Martin. The title of this thesis is "Cognitive Vision: Supervised Learning for Adaptive Image Segmentation".
- Another thesis will be defended on the 14th of December by Nicolas Maillot on the topic of "Ontology Based Object Learning and Recognition".
- A thesis has been defended on the 30th of April 2005 by Céline Hudelot [17] on the following topic: "Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms".

6.3.2. Knowledge Acquisition and Learning

Participants: Nicolas Maillot, Vincent Martin, Monique Thonnat.

Ontology Based Knowledge Acquisition

Experts often use a well-defined vocabulary to describe the objects of their domain (i.e. palynology, astrophysics). Our goal is to capture this knowledge so as to use it in the cognitive vision architecture presented in the next subsection. We propose an ontology based acquisition process to guide knowledge acquisition. A visual concept ontology has been designed for that purpose. This ontology is structured in several parts : spatio-temporal concepts, texture concepts, color concepts, relations between concepts (e.g. spatio-temporal relations). An acquisition context ontology is also proposed (e.g. point of view, acquisition device).

The knowledge base resulting from the acquisition process is used by the cognitive vision platform. A knowledge acquisition tool called OntoVis (see fig. 11) has been developed and used for the description of pollen grain images. This tool allows domain knowledge acquisition (i.e. domain objects and their subparts) and visual description guided by the ontology. This tool also provides a module to manage image samples. For instance, figure 11 shows the screen shot results from a knowledge acquisition session in the domain of palynology. The tool is composed of four main parts: (1) the domain class hierarchy; (2) the subparts associated with each domain class; (3) visual concepts used for the visual description; (4) image samples of the objects of interest used for visual concept learning purposes. This new approach for designing knowledge bases for vision systems is described in [45].

Learning for visual concept detection

One difficult task when using knowledge based approaches is the *symbol grounding* problem. During 2005, our efforts have been focused on the use of machine learning techniques to solve the symbol grounding problem.

Three approaches can be used to obtain segmented image samples from an image training set:

- **Manual segmentation**. This approach consists of manual segmentation of the images within the image training set. This approach has been used in [46];
- **Use of 3-D models**. The use of 3-D models is especially useful for learning the shape of manufactured objects (e.g. cars, aircrafts). Examples of regions obtained from the 2-D projection of 3-D model of an aircraft is given in fig. 13;

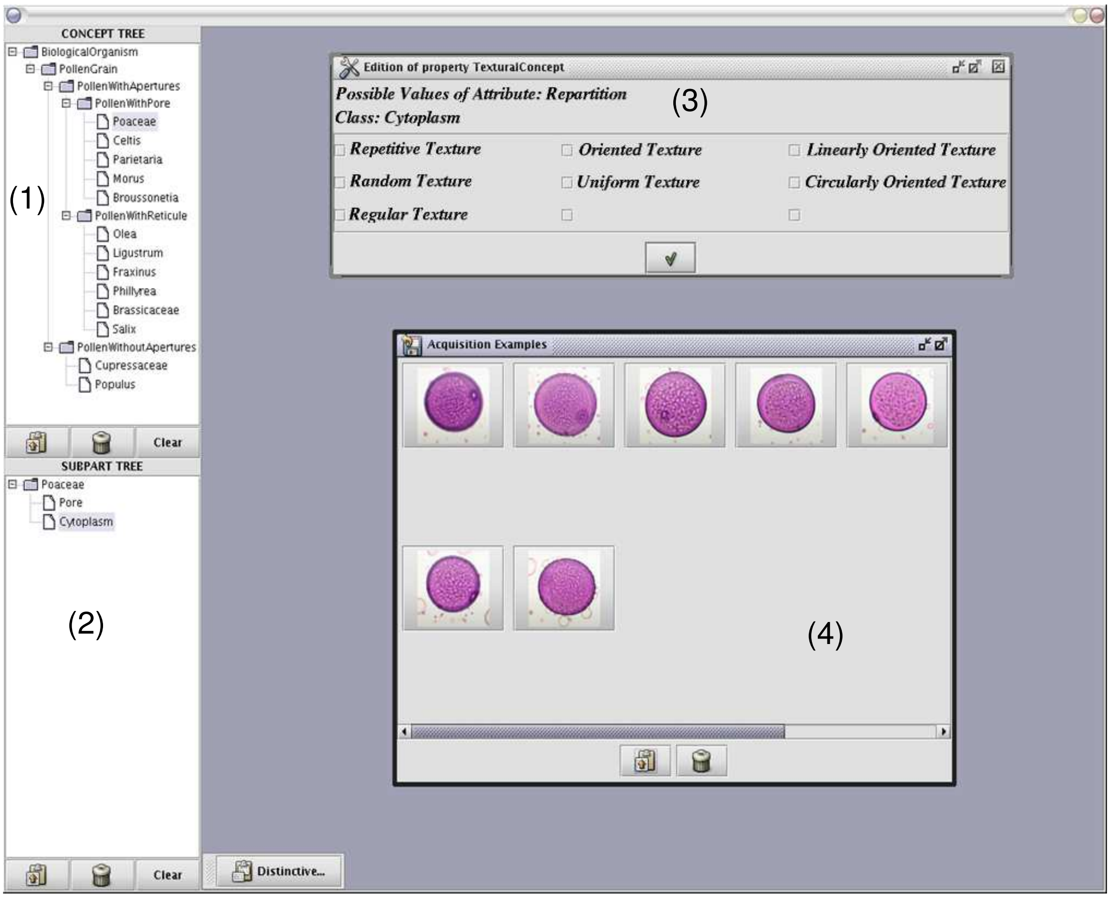


Figure 11. Overview of Ontovis

- **Weakly-supervised approach**. This approach is based on automatic segmentation and feature extraction algorithms and on unsupervised machine learning techniques (i.e. k-means). Regions which have similar properties (i.e. similar color, similar shape, similar texture) are first gathered into clusters by similarity. The resulting clusters only have to be semantically annotated by visual concepts provided by the ontology. This method enables the expert to annotate several regions contained in a cluster at the same time. This weakly-supervised approach is depicted in fig. 14.

In any case, the result is a set of regions annotated by visual concepts. From these *annotated* regions, *supervised machine learning techniques* (i.e. support vector machines, multi layer perceptrons) are used to obtain a set of visual concept detectors.

Once learning phase is over, trained detectors are used to perform automatic visual concept detection. Visual concept detectors detect visual concepts in the image to interpret. An example of visual concept detection can be found in fig. 12.

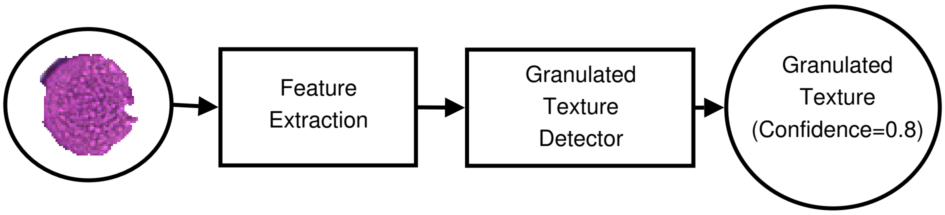


Figure 12. From an image region to a confidence degree associated with a visual concept. In this case, the region is recognized as being $Granulated$ with a confidence value of 0.8. $Granulated\ \ Texture$.

A probabilistic combination of detected visual concepts gives a confidence value about the mapping between high-level hypotheses and image data. This approach is presented in [21].

Adaptive Image Segmentation

Image segmentation is a low-level task that consists in partitioning the image into homogeneous regions distinct from each other, according to some criteria. It is a crucial step in computer vision systems involving image processing (e.g. object recognition, content-based image retrieval) where the challenge is to perform an image segmentation with some semantic meaning.

During this year, we have been focused on the development of a generic and reusable system for image segmentation based on a supervised learning scheme. Our work deals with automatic algorithm parameterization according to image characteristics. We propose a scheme to automatically select segmentation algorithms and tune their key parameters thanks to a preliminary supervised learning stage.

The learning phase (see figure 15) is subdivided into three stages: (1) optimal algorithm parameter extraction, (2) construction of a case base which contains processed cases. Each entry of this base is related to features describing an image with the corresponding optimal algorithm parameters and (3) algorithm selection learning.

The automatic phase (see figure 15) uses this knowledge for automatic and adaptive segmentation. Features are given in input of the algorithm selection predictor trained in previous stage (1). Then, similarity is determined by looking up the case base for similar cases (2). When the closest one is found, image is segmented with corresponding optimal parameters.

This approach is detailed in [47].

Figure 16 presents first results of the automatic phase. The image database is composed of 140 samples images of aircrafts in outdoor scenes. This dataset is randomly divided into 67 training images and 73 testing images. Three candidate image segmentation algorithms compose the library: a mean shift segmentation algorithm, a region growing algorithm, and an inherently parallel hierarchical color segmentation algorithm.

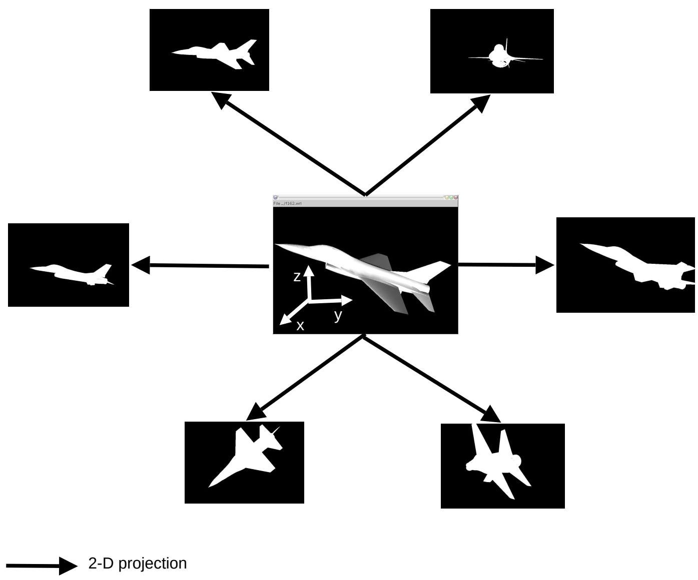


Figure 13. 2-D regions obtained from the projection of a 3-D mesh. After projection, resulting regions have to be annotated.

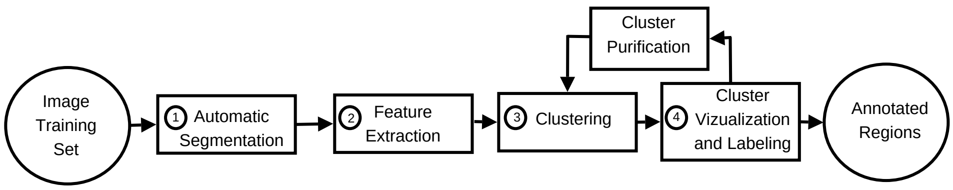
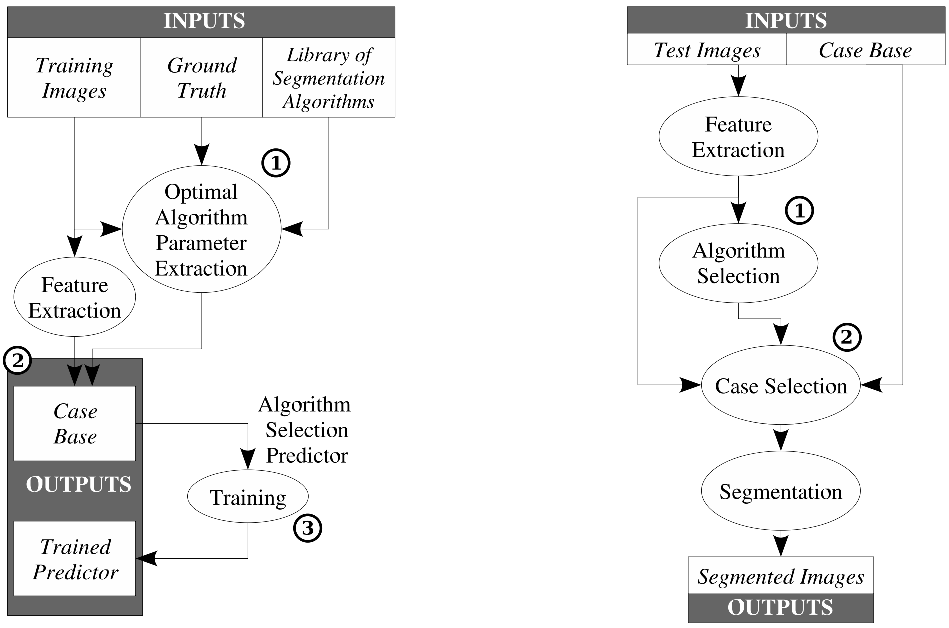


Figure 14. Clustering process.



(a) Overview of the learning phase (b) Overview of the automatic phase

Figure 15.

Our current work is to evaluate this method on large image databases from various domains. Future work is to guide the segmentation task according to a visual concept based description.

6.3.3. Semantic Image Interpretation

Participants: Céline Hudelot, Nicolas Maillot, Sabine Moisan, Monique Thonnat.

As explained in [17], we propose a generic architecture for the problem of semantic image interpretation based on the cooperation of three dedicated modules. Each module is specialized for the corresponding sub-task of semantic image interpretation : semantic interpretation and reasoning at the high level, visual data management for symbol grounding and for spatial reasoning and image processing.

An ontology based communication process is used to enable the inter-operability between the different modules. As described in figure 17, the visual concept ontology described in the previous section is used for the interoperability between the semantic interpretation module and the visual data management module. An image processing ontology described in [17] is used for the interoperability between the visual data management module and the image processing module.

We choose a knowledge based approach to emulate the strategy of the experts performing the corresponding task (application domain expert, cognitive vision expert and image processing expert). To propose re-usable and generic tools, we choose Knowledge Based System (KBS) shells techniques. They enable the design of engines and knowledge base models, independent of any applications but dedicated to a specific task. As described in figure 17, for each sub-task of semantic image interpretation, the cognitive vision platform proposes:

- the task-specific knowledge formalization;
- the specification of a dedicated engine with adapted reasoning.

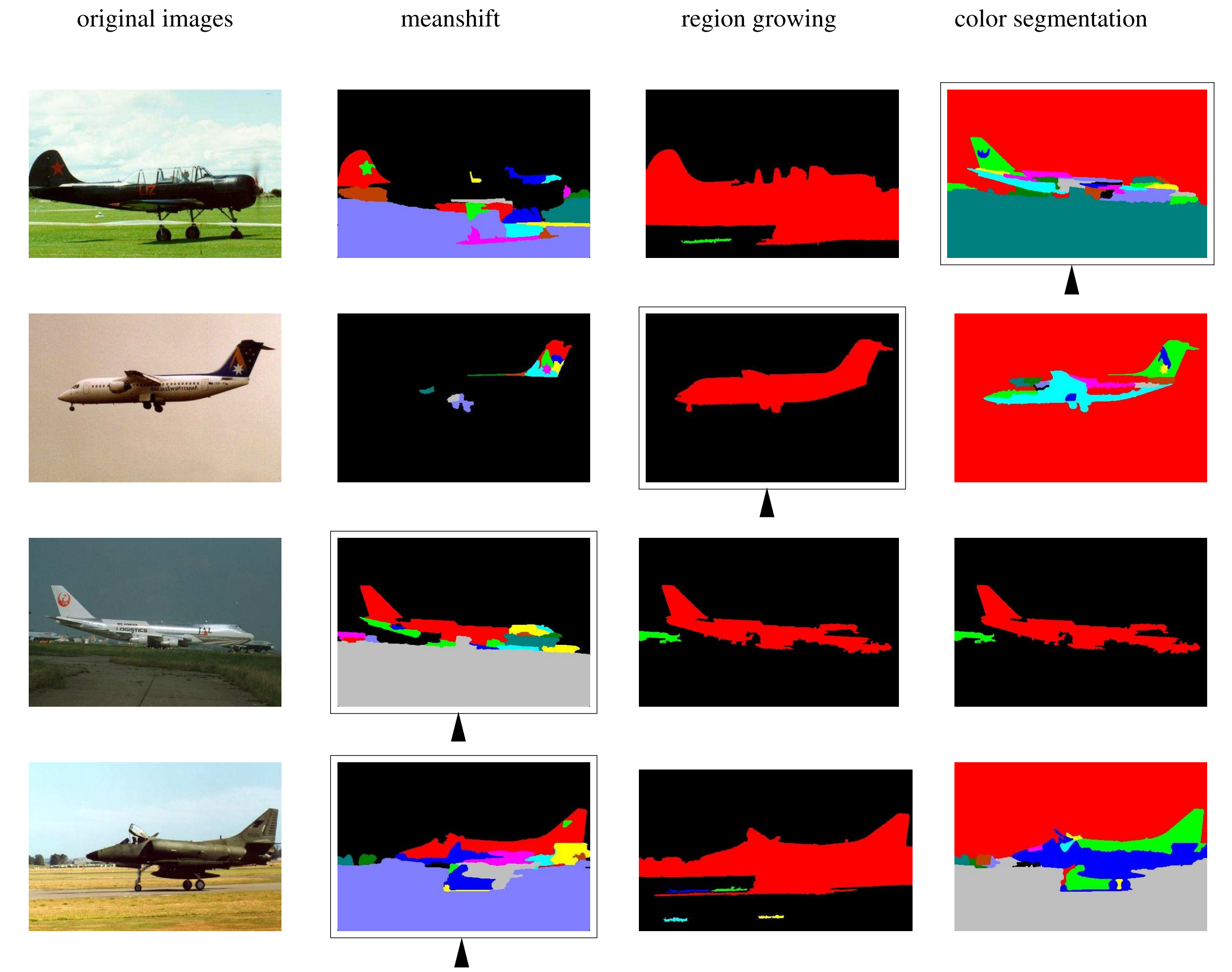


Figure 16. Segmentation results in automatic phase. For each test image, results of the three algorithms are presented. Surrounded images correspond to the selected algorithm by the system.

Moreover, we have

- considered goal-directed behavior for each sub-task through the generation of specific requests;
- made explicit the context useful for each sub-task.

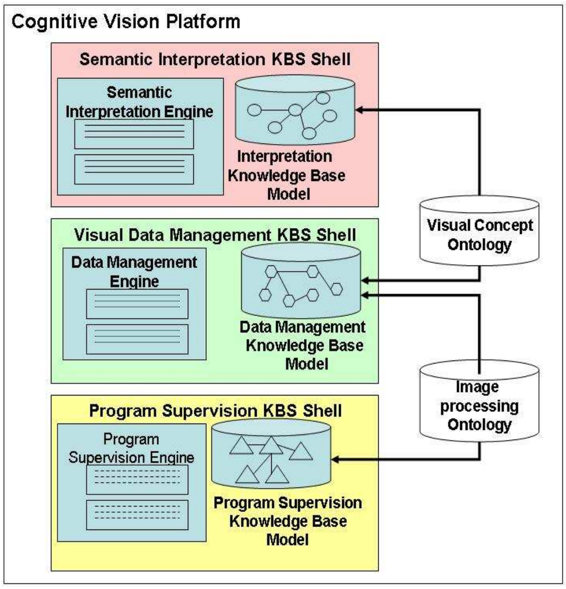


Figure 17. Global minimal architecture of our cognitive vision platform

High-level semantic interpretation

The high-level semantic interpretation process uses the domain specific knowledge acquired during the ontology based knowledge acquisition phase. This domain knowledge is organized as a hierarchy of domain classes described by visual concepts (e.g. a pollen grain is described as a circular surface which hue is pink). This hierarchy contains specialization and composition links. This kind of representation reflects the domain taxonomy.

The system is activated by end-user requests. An end-user request contains:

- an image or a sequence of images to be interpreted;
- the image acquisition context;
- a domain context;
- a high level goal. The system can process through three different modes : object classification (recognition of an instance of a specified object in a scene), multiple object classification (recognition of all the instances of the specified object in the scene) and scene analysis (recognition of all the objects of interest in the current scene).

The reasoning is based on the depth first traversal of the domain class tree. In a top down way, the engine generates physical object hypotheses according to the request of the end user and to a priori knowledge. The order in which objects are scheduled is dependent on the high level goal and on the current state of the interpretation. In a bottom up way the engine matches visual object instances found in images with the predefined classes of the hierarchy and refines the classification. This interpretation process is recursively applied to sub-parts of domain classes.

Visual data management

The main goal of this module is to establish the correspondence between symbols (abstract representation of physical objects) and sensor data : in our case image data. This task was often considered as part of the recognition algorithm. We think it is a task on its own which involves complex reasoning as uncertainty reasoning, spatial reasoning and data management reasoning (grouping or splitting of data, ...). We have designed a dedicated knowledge based system for this task : the visual data management knowledge based system. The study of works about **symbol grounding** and **anchoring** was useful for the design of this module. The knowledge base is composed of:

- **image data** organized in **image data primitives** (*edge, region, ...*) and **image data features** (*compactness, circularity, red value, ...*);
- **visual concepts** which are the descriptions from a visual data management point of view of the concepts of the visual concept ontology. In this framework, each visual concept is associated with a set of image data features. Each image data feature of this link is modeled as a fuzzy linguistic variable with a domain, a set of possible linguistic values and their associated fuzzy sets. This link is called the symbol grounding link. It explicitly encodes the link between visual concepts and realistic numerical values of image data features;
- **spatial relations** divided into topological, distance and orientation relations.

It also contains inferential knowledge : **object extraction criteria** to build image processing request according to the object hypothesis in terms of visual concepts, **spatial deduction criteria** to manage multi-object hypotheses by the use and propagation of spatial constraints, **data management criteria** to diagnose the image processing results according to visual object hypotheses.
From the reasoning point of view, the main goals of this module are:

- building of image processing requests according to visual object hypotheses and according to the current state of the interpretation (*complete*, *partial*, *hypothesized*, *missing*);
- matching between image processing results (in terms of image data and low level descriptors) and symbolic data (in terms of visual concepts). To manage uncertainty and the fuzziness of symbols we use fuzzy mapping techniques;
- instantiation of complete, partial or missing visual object instances found in images for the mapping with domain concepts by the interpretation modules;
- top down and bottom up spatial reasoning.

Image processing

This module has to perform the extraction and the numerical description of objects from images. We have to make good use of program supervision approach to manage different contexts and to process images in an intelligent and adaptable way. As explained before, program supervision techniques allow the automation of the configuration and execution of a library of programs for a given objective. Most of image processing objectives and programs are common and reusable for many applications.
Thus, the minimal image processing knowledge base contains knowledge about generic image processing functionalities like object segmentation, image filtering, feature extraction, low level grouping... The generic image processing requests are generated by the data management knowledge based system. They are linked

with an **Image High Level Context** which contains top down constraints. This **Image High Level Context** is used by the program supervision engine as additional information to manage programs.

Figure 18 depicts how the cognitive vision platform is used in a cooperative way, by three different experts (i.e. an application domain expert, a visual data management expert and an image processing expert) to build a complete semantic interpretation system for a specific application. We argue that experts of a specific domains are the best persons to deal with their domain and to explicit their knowledge. The modularity of the cognitive vision platform enables experts to contribute only at their level of expertise, using specific knowledge description languages provided by the cognitive vision platform.

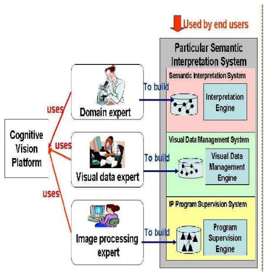


Figure 18. Cooperation of different experts for the design of a particular semantic image interpretation system using the cognitive vision platform

6.3.4. Applications

Detection of plant diseases

In cooperation with INRA (URIH of Sophia Antipolis), the early detection of plant diseases, in particular rose diseases is used to evaluate and validate our platform [1]. The data we are working on are two-dimensional macroscopic and microscopic (magnified per 60) images of rose leaves. The objects of interest can be fungi which appear as thin networks more or less developed or insects of various shapes and appearances. Examples can be seen in fig. 19. Indeed in these examples, the two images of powdery mildew emphasize the complexity and variability of object appearance and of contexts : we need domain knowledge and intelligent management of image processing programs. The image of insects is a case of multiple object scene : we need spatial knowledge and spatial reasoning. Using the cognitive vision platform, the building of the rose disease detection system consists in the building of three dedicated knowledge bases:

- a domain knowledge base describes the early signs and symptoms on leaves of rose diseases;
- a visual data management knowledge base contains the symbolic description of the visual concepts, the spatial relations and the image data useful to manage the problem of rose disease detection;
- an image processing program supervision knowledge base describes the knowledge on how to use generic image processing programs. We have used the image processing library PANDORE; some missing functionalities (e.g. ridge extraction on images) were added to complete this library.

The knowledge bases contain 218 frames and 73 rules.

Semantic Image Indexing and Retrieval

The cognitive vision platform is currently used to perform image indexing. As described in [46], we propose to index a set of images (e.g. from a video stream) by using domain knowledge. We have evaluated the cognitive vision platform on 7000 images. Our goal was to index aircraft and ship images. Once indexing is performed, querying of the indexed image database is done by keywords (e.g. Aircraft, Blue Sky) The categorization process is illustrated in fig. 20. The output of the system is an interpretation of the input image.

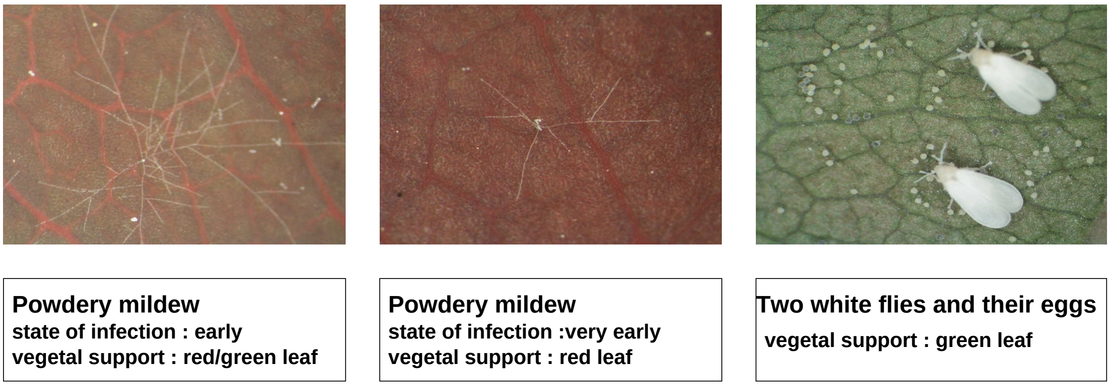


Figure 19. Examples of analyzed images

Image processing tasks (i.e. segmentation, feature extraction) are needed to achieve the image interpretation. Trained visual concept detectors are used in the object categorization process. They are used to map image data to high-level symbolic description of domain classes. We have obtained a recall of 0.19 for a precision of 0.73.

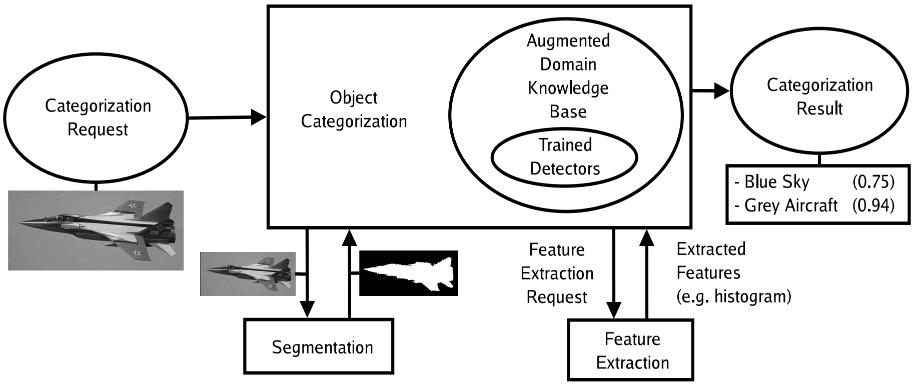


Figure 20. Object categorization for image indexing

Precision is defined as the ratio between the number of relevant retrieved images and the number of retrieved images. Recall is defined as the ratio between the number of relevant retrieved images and the number of relevant images in the image database. These results are promising. We are currently working to improve the performance of the system.

In [32] is presented how a semantic image indexing and retrieval system can be built in a weakly-supervised approach. An image indexing and retrieval system dedicated to the domain of vehicles (i.e. aircraft, cars, motorbikes) has been obtained. The results obtained on a database of 1850 images are shown in fig. 21. This figure shows that results have been significantly improved: for a recall of 0.5, precision is between 0.75 and 0.78 for the domain classes $Aircraft$, $MotorBike$ and Car and of 0.90 for class Sky. These results show that even with very little effort of knowledge acquisition, this approach offers both good results and semantic richness. This is mainly due to an improvement at the feature extraction level by the use of SIFT features.

Note that depending on the user needs, high precision can be obtained at the cost of low recall. High recall can be obtained at the cost of low precision.

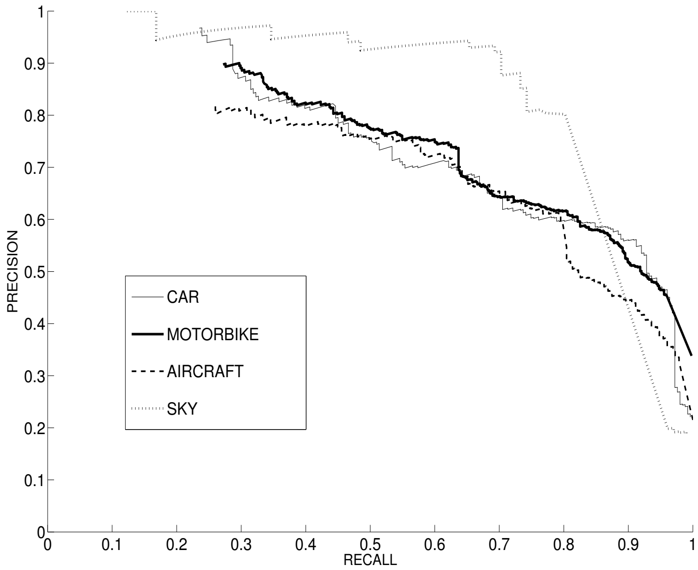


Figure 21. Results obtained for image indexing and retrieval system dedicated to vehicles.

7. Contracts and Grants with Industry

7.1. Industrial Contracts

In 2005, 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 VIDEA project to transfer a part of ORION activities to Vigitec company.

- CASSIOPEE : This project aims 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 is planned for 51 months, from first of January 2002 to thrity-first of March 2006.
- SAMSIT : SAMSIT is a project in collaboration with ALSTOM, CEA, SNCF, INRETS. It has begun in January, 2004 and will last 2 years. 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.
- TELESCOPE 3 : The aim of this project with Bull is 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.

- VIDEA : This project with Vigitec (Brussels) began in November 2003 for 2 years. The aim of this project was to transfer a part of the video surveillance technology of the ORION team into industrial products. During this project, the ORION team has developed and transfered to a video surveillance company two applications enabling the recognition of specific human behaviors.

8. Other Grants and Activities

8.1. European projects

ORION team has been involved this year in three european projects on image interpretation : ECVision European Research Network (IST-type), a project on airport surveillance (AVITRACK) and a new project on crowd behavior analysis (SERKET).

8.1.1. ECVision European Research Network

The ECVision European Research Network has begun in march 2002 for 3 years. This research network was formed to promote and merge activities of 50 European laboratories working in cognitive vision (see http://www.ECVision.info/home/Home.htm).

8.1.2. 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. 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.

8.1.3. 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 helping for terrorist prevention.

8.2. International Grants

Orion is involved in two international programs STIC-Asie (ISERE) and STIC-Tunisie.

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. STIC-Tunisie

Orion team cooperates with ENSI Tunis (Tunisia) in the framework of STIC Franco-Tunisian cooperations. A joint PhD thesis (N. Khayati) dedicated to research on distributed medical imagery program supervision. 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.

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. A joint PhD thesis [17] on cognitive vision has been defended in may 2005. Real data and biological expertise are provided by INRA to test and develop a new cognitive vision platform enabling natural object recognition (see also 4.5).

8.3.2. Program Supervision for River Hydraulics Simulation

Orion cooperates with Cemagref (Lyon) and INPT (Toulouse) for solving an environmental problem: river flood prediction. In practice, program supervision techniques developed by Orion are used together with hydraulics simulation codes to build a prototype knowledge-based system. This protype is dedicated to end-users of the MAGE simulation code to achieve a river model calibration activity (see 6.1.4). A joint PhD thesis [19] has been defended.

8.3.3. 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 (see also 6.2.4).

8.3.4. Long-term Monitoring Person at Home

This year, Orion has started a collaboration with CSTB (Centre Scientifique du Batiment) 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 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 (see 6.2.8), a platform to provide services and to perform experiments should be devised.

8.3.5. Classification of Lateral Forms for Control Access Systems

In the framework of a collaboration with RATP, 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, (see also 6.2.3).

8.3.6. 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 defense 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 Creation

This year, Orion has created a spin off: Keeneo.

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 Chair of the Industrial Liaison Committee of the IAPR (International Association for Pattern Recognition) for the period 2004-2006.
- M. Thonnat is an area leader of the ECVision European Excellence Network in cognitive vision domain since march 2002 for 3 years (50 teams and 12 countries).
- M. Thonnat is a reviewer for the journals IEEE Transactions - Pattern Analysis and Machine Intelligence PAMI, CVIU (Journal of Computer Vision and Image Understanding), Image and Vision Computing Journal and I3 (Information - Interaction - Intelligence).
- M. Thonnat is a Program Committee member for the following conferences: IJCAI05 (Poster Track), TAIMA05, Acivs 2005, the 2005 conference on Advanced Concepts for Intelligent Vision Systems, HAREM2005 - Human Activity Recognition and Modelling 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 Program Committee Chair with Bruce Draper of the IEEE International Conference on Computer Vision Systems January 5-7, 2006 St. Johns University, Manhattan, New York City NY, USA, http://www.cs.colostate.edu/icvs06/OrganizingCommittee.htm.
- M. Thonnat is an expert for ANVAR the French agency for research valorization.
- M. Thonnat is member of the steering committee of the Technovision program from the French Ministry of Research and the French Ministry of Defence.
- M. Thonnat is an expert for the French Ministry of Defence.
- M. Thonnat is reviewer for the following theses: Yifan SHI PhD at Georgia Tech (Atlanta USA) under the supervision of Aaron Bobick , Pierre Regnier HDR (Habilitation à diriger des recherches, IRIT Toulouse University), Francois Portet PhD (Rennes University) under the supervision of Marie-Odile Cordier.
- M. Thonnat is member of the INRIA Evaluation board since 2003.
- M. Thonnat is member of the scientific board of INRIA Sophia Antipolis (bureau du comité des projets) since september 2005.
- M. Thonnat 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 a reviewer for the following conferences: WAMOP Workshop on Performance Evaluation of Tracking and Surveillance and the CVPR05, ACIVS05, ICCV05, IJCAI05, AVSS05, ECCV06, VIE06, VS-Pets2005, VS2006, ICDP05, ICDP06,ICNSC06.

- F. Brémond is a reviewer for the following journals: EURASIP JASP, Pattern Recognition Letters, CVIU.
- F. Bremond is a member of the program commitee of the WAMOP Workshop on Performance Evaluation of Tracking and Surveillance and VIE06, VS-Pets2005, VS2006, ICDP05, ICDP06,ICNSC06.
- F. Bremond is the organizer of the ETISEO workshop on Performance Evaluation on Video Understanding Techniques.
- N. Maillot is a reviewer for Image and Vision Computing Journal (Elsevier).
- J-P Rigault is member of the AITO Committee and participates to its General Meeting (the steering committee of ECOOP conferences).
- A. Ressouche is a member of the Inria Cooperation Locales de Recherches (Colors) committee.

9.2. Teaching

- Orion is a hosting team for the DEA 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 Essi (UNSA), Usage of Synchronous languages dedicated tools TP (12h A. Ressouche).
- Teaching at UNSA, Image processing Introduction (TD, 20h, C. Hudelot), Web Computing (TD, 20h, C. Hudelot), Unix Introduction (TD, 9h, C. Hudelot).
- Teaching at ESSI (UNSA), Object-Oriented Design and Programming in C++ (50h N .Maillot).
- Teaching at IUT (UNSA), Java programming (25h N. Maillot).
- Teaching at CIV (Centre International Valbonne) TD Informatique - Classes Preparatoires MPSI 1ere annee (56h B. Boulay).

9.3. Theses

9.3.1. Theses in progress

- Bernard Boulay : Reconnaissance de postures pour l'interprétation d'activités humaines, 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.
- 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.
- Marcos Zùniga : Unsupervised primitive event learning and recognition in video , Nice-Sophia Antipolis University.

9.3.2. Theses defended

- Céline Hudelot : Vers une plate forme de vision cognitive pour l'interprétation sémantique d'images : application à la reconnaissance d'organismes biologiques, Nice Sophia Antipolis University (thesis defended in May 2005).
- Benoit Georis : Knowledge-based reconfigurable tracker, Louvain Catholic University (thesis defended in December 2005).
- Nicolas Maillot : Système cognitif d'interprétation d'images pour la reconnaissance d'images d'objets 3D, Nice Sophia Antipolis University (thesis defended in December 2005).
- Jean-Philippe Vidal : Validation opérationnelle en hydraulique fluviale – Approche par un système à base de connaissances, Institut National Polytechnique de Toulouse (thesis defended in March 2005).

9.4. Participation to Conferences, Seminars, and Invitations

Members of the Orion team have presented papers in the following conferences:

- ICCV05: International Conference on Computer Vision 17 -20 October 2005 Beijing,China (M Thonnat and Celine Hudelot).
- CIVR05: International Conference on Image and Video Retrieval 22-22 July 2005 Singapore (N. Maillot).
- SKCV: International Workshop on Semantic Knowledge for Computer Vision 16 October 2005 Beijing,China (M Thonnat and Celine Hudelot).
- AVSS05:IEEE International Conference on Advanced Video and Signal-Based Surveillance (F Bremond).
- MB05: Measuring Behavior , Wageningen The Netherlands, September 2005, (F Bremond).
- CAP05:Invited talk at "Atelier Connaissance et Documents Temporels" (AFIA 2005), Nice, France (F Bremond).
- ICDP05: IEE International Symposium on Imaging for Crime Detection and Prevention, London, UK, (F. Bremond, B. Boulay) .
- ECOOP'2005: The 19th European Conference on Object-Oriented Programming, Glasgow, Scotland, July 25-29 2005, (J-P Rigault).
- SLAP'05: Synchronous Languages Applications and Programming, Edinburgh, Scotland, (A Ressouche).
- AMBS'05: Invited talk at the Annual Meeting of Bank Security, 31st of May 2005 Toulouse(M. Thonnat).

10. Bibliography

Major publications by the team in recent years

[1] P. BOISSARD, C. HUDELOT, G. PÉREZ, P. BÉAREZ, P. PYRRHA, M. THONNAT, J. LÉVY. *La détection précoce in situ des bioagresseurs : application à l'oidium du rosier de serre*, in "Revue Horticole", septembre 2004.

[2] P. BOISSARD, C. HUDELOT, M. THONNAT, G. PEREZ, P. PYRRHA, F. BERTAUX. *Automated Approach to Monitor the Sanitary Status and to Detect Early Biological Attacks on Plants in Greenhouse*, in "International Symposium on Greenhouse Tomato, Avignon", CTIFL, 2003.

[3] 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.

[4] N. CHLEQ, F. BREMOND, 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.

[5] 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.

[6] N. DEY. *Etude de la Formation de l'Image d'un Objet Microscopique 3D Translucide - Application à la microscopie optique*, Ph. D. Thesis, Université de Nice - Sophia Antipolis, November 2002.

[7] N. MAILLOT, M. THONNAT. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", 2004.

[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, 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.

[11] M. THONNAT, S. MOISAN. *Knowledge-based systems for program supervision.*, in "First international workshop on Knowledge-Based systems for the (re)Use of Programs libraries KBUP'95, Sophia Antipolis, France", INRIA, March 1995, p. 4–8.

[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] J.-P. VIDAL, J.-B. MOISAN. *Knowledge-Based Hydraulic Model Calibration*, in "Knowledge-Based Intelligent Information and Engineering Systems (KES), Oxford, UK", V. PALLADE, R. HOWLETT, L. JAIN (editors). , LNAI, vol. I, n° 2773, Springer, September 2003, p. 676–683.

[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 for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03), Acapulco, Mexico", 2003.

Doctoral dissertations and Habilitation theses

[16] B. GEORIS. *Knowledge-based reconfigurable tracker*, Ph. D. Thesis, Louvain Catholic University, 2005.

[17] C. HUDELOT. *Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of of Biological Organisms*, Ph. D. Thesis, Nice-Sophia Antipolis University, 2005.

[18] N. MAILLOT. *Système cognitif d'interprétation d'images pour la reconnaissance d'images d'objets 3D*, Ph. D. Thesis, Nice-Sophia Antipolis University, 2005.

[19] J.-P. VIDAL. *Assistance au calage de modèles numériques en hydraulique fluviale : apports de l'intelligence artificielle*, Ph. D. Thesis, INP Toulouse, March 2005.

Articles in refereed journals and book chapters

[20] A. AVANZI, F. BRÉMOND, C. TORNIERI, M. THONNAT. *Design and Assessment of an Intelligent Activity Monitoring Platform*, in "EURASIP Journal on Applied Signal Processing, special issue in "Advances in Intelligent Vision Systems: Methods and Applications"", 2005.

[21] N. MAILLOT, M. THONNAT. *Ontology Based Complex Object Learning and Recognition-Under Minor Revision*, in "Image and Vision Computing", 2005.

[22] J.-P. VIDAL, S. MOISAN, J.-B. FAURE, D. DARTUS. *Towards a reasoned 1D river model calibration*, in "J. Hydroinformatics", vol. 7, n° 2, April 2005, p. 79–90.

Publications in Conferences and Workshops

[23] M. BORG, D. THIRDE, J. FERRYMAN, F. FUSIER, F. BRÉMOND, M. THONNAT. *An Integrated Vision System for Aircraft Activity Monitoring*, in "Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), Breckenridge, Colorado, USA", IEEE Computer Society, January 2005.

[24] M. BORG, D. THIRDE, J. FERRYMAN, F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT, J. AGUILERA, M. KAMPEL. *Visual Surveillance for Aircraft Activity Monitoring*, in "Proceedings of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China", IEEE Computer Society, October 2005.

[25] M. BORG, D. THIRDE, J. FERRYMAN, F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT. *Video Surveillance for Aircraft Activity Monitoring*, in "Proceedings of IEEE International Conference on Advanced Video and Signal-Based Surveillance, Como, Italy", IEEE Computer Society, September 2005.

[26] B. BOULAY, F. BRÉMOND, M. THONNAT. *Posture Recognition with a 3D Human Model*, in "Proceedings of IEE International Symposium on Imaging for Crime Detection and Prevention, London, UK", Institution of Electrical Engineers, 2005.

[27] B. BUI, F. BRÉMOND, M. THONNAT, J. FAURE. *Shape Recognition Based on a Video and Multi-Sensor System*, in "IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2005), Como, Italie", September 2005.

[28] F. CUPILLARD, F. BRÉMOND, M. THONNAT. *Automatic Visual Recognition for Metro Surveillance*, in "Measuring Behavior, Wageningen The Netherlands", September 2005.

[29] J. FERRYMAN, M. BORG, D. THIRDE, F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT, J. AGUILERA, M. KAMPEL. *Automated Scene Understanding for Airport Aprons*, in "Proceedings of 18th Australian Joint Conference on Artificial Intelligence, Sidney, Australia", Springer-Verlag, December 2005.

[30] N. KHAYATI, W. LEJOUAD-CHAARI, S. MOISAN. *Systèmes distribués de pilotage de programmes : application à l'imagerie médicale*, in "Proc. Méthodes et Heuristiques pour l'Optimisation des Systèmes industriels (MHOSI'2005), Hammamet, Tunisie", Avril 2005.

[31] N. KHAYATI, W. LEJOUAD-CHAARI, S. MOISAN, J.-P. RIGAULT. *A Mobile Agent-based Architecture for Distributed Program Supervision Systems*, in "Proc. 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain", December 2005.

[32] N. MAILLOT, M. THONNAT. *A Weakly Supervised Approach for Semantic Image Indexing and Retrieval.*, in "CIVR", 2005, p. 629-638.

[33] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Synchronous Formalism and Behavioral Substitutability in Component Frameworks*, in "Synchronous Languages Applications and Programming", F. MARANINCHI (editor). , ENCS, Elsevier Science, 2005.

[34] D. THIRDE, M. BORG, J. FERRYMAN, F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT. *Video Event Recognition for Aircraft Activity Monitoring*, in "Proceedings of 8th IEEE International Conference on Intelligent Transportation Systems, Vienna, Austria", IEEE Computer Society, September 2005.

Miscellaneous

[35] ETISEO. *Video Understanding Evaluation Project*, 2005, http://www.etiseo.net.

Bibliography in notes

[36] M. BORG, D. THIRDE, J. FERRYMAN, F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT. *A Real-Time Scene Understanding System for Airport Apron Monitoring*, in "Proceedings of 2006 IEEE International Conference on Computer Vision Systems, New York, USA", IEEE Computer Society, January 5-7, 2006.

[37] J. CHAMBERS, ET AL.. *The R Project for Statistical Computing*, http://www.r-project.org.

[38] 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.

[39] 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)(To appear), New-York, NY, USA", January 2006.

[40] S. KAWASHIMA, Y. TAKAHASHI, S. AIKAWA, T. NAGOYA. *An attempt of applying the image processing for the automatic estimation of sampled airborne pollen*, in "Japanese Journal of Allergology", vol. 44, n^o 9, 1995, p. 1150–1158.

[41] E. KEOGH, M. PAZZANI. *Derivative Dynamic Time Warping*, in "Proceedings of the 1st SIAM International Conference on Data Mining (SDM'2001), Chicago, USA", 2001.

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

[43] 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.

[44] S. LIU, M. THONNAT, M. BERTHOD. *Automatic Classification of Planktonic Foraminifera by a Knowledge-based System*, in "The Tenth Conference on Artificial Intelligence for Applications, San Antonio, Texas", IEEE Computer Society Press, March 1994, p. 358–364.

[45] 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.

[46] N. MAILLOT, M. THONNAT, C. HUDELOT. *Ontology Based Object Learning and Recognition : Application to Image Retrieval*, in "Proceedings of 16th IEEE International Conference on Tools For Artificial Intelligence, Boca Raton, USA", IEEE Computer Society Press, 2004.

[47] V. MARTIN, N. MAILLOT, M. THONNAT. *A Learning Approach for Adaptive Image Segmentation*, in "ICVS", IEEE (editor). , Jan 2006.

[48] M. MAZIÈRE. *Étude de faisabilité pour la reconnaissance automatique de grains de pollen*, Rapport de DEA, Université de Nice-Sophia Antipolis, 1997.

[49] S. MOISAN. *Program Supervision:* YAKL *and* PEGASE+ *Reference and User Manual*, Tecnichal Report, n^o 5066, INRIA, December 2003, http://www.inria.fr/rrrt/rr-5066.html.

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

[51] J. OSSOLA. *Coopération de systèmes à base de connaissances pour l'analyse et la reconnaissance d'objets naturels complexes : application au classement de galaxies ou de zooplanctons*, Ph. D. Thesis, Université de Nice-Sophia Antipolis, May 1996.

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

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

[54] 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.

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

[56] R. TOMCZAK, P. BONTON, C. AUROYER, D. CAILLAUD, C. ROUQUET. *Traitement d'images et reconnaissance des formes appliquées à la mesure des taux de pollens dans l'air*, in "Journées Automatique, Agriculture et Agro-alimentaire, Clermont-Ferrand, France", 1997.

[57] 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.

[58] 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.