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

2004

Table of contents

1. Team

Scientific leader
Monique Thonnat [DR Inria]

Administrative Assistant
Catherine Martin

Inria Researchers
François Brémond [CR Inria]
Sabine Moisan [CR Inria]
Annie Ressouche [CR Inria]

Engineers
Alberto Avanzi [Bull Engineer]
Julien Canet [Fellow Engineer]
Nicolas Chleq [IR2 Inria from july 2004]

Research Engineers
Binh Bui [RATP Paris, up to june 2004]
Frédéric Cupillard [Videa/Vigitec Brussels]
Gabriele Davini [SAMSIT project]
Florent Fusier [European project AVITRACK]
Magali Mazière [Cassiopée/Crédit Agricole]
Christophe Tornieri [SNCF Paris]
Valery Valentin [European project AVITRACK]

PhD Students
Bernard Boulay [INRIA Grant, Nice Sophia-Antipolis University]
Benoit Georis [Louvain Catholic University (UCL) Belgium]
Céline Hudelot [Regional Grant, Nice Sophia-Antipolis University]
Naoufel Khayati [ENSI Tunis, since January 2004]
Nicolas Maillot [INRIA Grant, Nice Sophia-Antipolis University]
Vincent Martin [Regional Grant, Nice Sophia-Antipolis University,since october 2004]
Jean-Philippe Vidal [CEMAGREF Grant, Toulouse University]
Thinh Van Vu [INRIA Grant, Nice Sophia-Antipolis University, up to October 2004]

Associates
Jean-Paul Rigault [Professor, Nice Sophia-Antipolis University]

Undergraduates
Jihene Bannour [DEA Saint-Etienne, mars-august 2004]
Vincent Martin [DEA, Astro UNSA, april-september 2004]
Le Son Pham [DEPA,IFI Hanoi, april-september 2004]

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 **intelligent reusable systems** and **cognitive vision**.

2.1.1. Research Themes

More precisely, our research themes deal with 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 to facilitate the construction of knowledge-based 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 about objects, of events and of scenarios to be recognized, as well as the reasoning processes useful 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 which with Silogic S.A. Toulouse, CCI Aeroport 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;
- Contract with RATP (in video interpretation) for passenger detection and classification in real time;
- Participation to projects with ALSTOM, SNCF, CEA and INRETS, both 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 distribution of medical imagery program supervision;
- Cooperation with National Cheng Kung University (Taiwan), National Taiwan University (Taiwan), MICA (Vietnam), IPAL (Singapor), I2R (Singapor), NUS (Singapor), CLIPS-IMAG (CNRS-INPG-UJF Grenoble) and IRIT (Toulouse). This cooperation 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;
- Cooperation with ST-Microelectronics (Le Rousset Aix en provence) on Intelligent Cameras.

3. Scientific Foundations

3.1. Introduction

In order to facilitate the design of KBS, we design generic engines, but yet *dedicated* to a particular task. The tasks that we study are program supervision and image understanding.

Developing dedicated tools allows us to provide systems that are adapted to express the necessary knowledge and that can be used in a wide range of application domains.

To design such engines, it is necessary to rely on models of both knowledge and reasoning mechanisms (problem solving methods) that are involved in the tasks we study.

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: both 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 [10], which is devoted both to knowledge base and inference engine design.*

Program supervision aims at automating the reuse 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 KBS 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, but 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 use 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 where complex processing is necessary 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 powerful control mechanisms, like sophisticated repair mechanisms.

Program Supervision Model

To better understand the general characteristics of the program supervision activity and to improve the (re)use of existing programs, the knowledge involved in this activity has to be modeled independently of any application. The knowledge model defines the structure of program descriptions and what issues play a role in the composition of a solution using the programs. It is thus a guideline for representing reusable programs. We have thus used knowledge modeling techniques to design an explicit description of program supervision knowledge to allow the necessary expertise to be captured and stored for supporting of a novice user or an autonomous system performing program supervision. We have modeled concepts and mechanisms of program supervision first for the OCAPI [4] engine, and then for our more recent engines. A preliminary work with KADS expertise model has been improved by using recent component reuse techniques (from Software Engeneering area), planning techniques (from Artificial Intelligence area), 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 LAMAplatform 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. [43]

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 KBS (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 belonging to 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* (inference engines, interfaces, knowledge base description languages, verification tools, etc.). That is why we decided to design a generic and adaptable software development platform, namely the LAMA platform [10]. 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 KBSs 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.

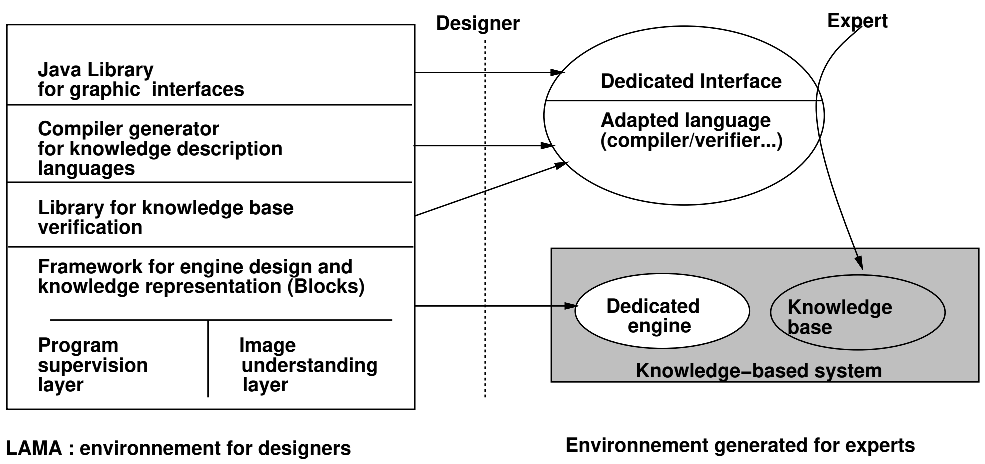


Figure 1. LAMA *architecture and tools for engine design, knowledge base description, verification, and visualization*

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 the support of 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 languages) –, 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 subtyping. 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 KBS 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 (KBS) 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 KBS (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 KBSs (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 Image Interpretation

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

Participants: Francois Brémond, Monique Thonnat.

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

Automatic image interpretation is a difficult problem which is the basis of many research activities in both computer vision and artificial intelligence. The difficulty of the interpretation task first comes from the type of the entities to be recognized. It is easier to recognize manufactured rigid objects than the behaviour of several natural objects in a dynamic context. The difficulty also depends on the type of interpretation to be performed. The problem can be a simple labeling of an entity detected in an image associated with a model or the detection and the consistency checking (e.g. spatial, temporal, structural) of an entity set.

The Orion team aims at the automatic interpretation of spatial and/or temporal images. Interpretation results can be object categorization but also event, situation or scenario recognition. The approach is composed of two steps. (1) An image processing step which aims at extracting entities of interest for interpretation. (2) The analysis of the extracted entities which can be selected for object categorization or for behavior recognition.

The difficulty of the problem is that two types of knowledge are required. First, primitive and descriptor extraction relies on the execution of image processing programs and requires knowledge on vision algorithms. Second, the interpretation task requires expert knowledge of the application domain related to the entities to be recognized or analyzed. Automatic image processing execution is related to program supervision and is a research activity by itself (cf. module 3.2).

The next points describe the proposed approaches for complex object recognition and for image sequence analysis.

Complex Object Recognition :

Complex object recognition refers to recognize non-geometric objects from abstract semantic models. In a first stage, image processing techniques are used to detect regions of interest so as to compute numerical descriptors. In a second stage, extracted descriptors are used by the interpretation system to classify the object in a predefined taxonomy of classes which define the semantic models. Three recursive tasks are involved in the classification process : data abstraction, class matching, recognition refinement. During the classification process, more information may have to be computed from the image. The operationalization of such systems requires an important work for the design of knowledge bases and for the implementation of image processing programs.

Interpretation of Image Sequences :

The interpretation of image sequences refers to giving a semantic explaining the human activities depicted from image streams provided by color, monocular and fixed cameras. The general base of the scene interpretation algorithm is based on the a priori knowledge (containing the scene model and a library of scenario models) and on the cooperation of the 4 following modules (cf figure 2): 1) mobile object detection and frame to frame tracking, 2) multi-cameras combination, 3) long term tracking, and 4) behavior recognition. The first module is implemented as three instances, one for each camera. It 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 module is optional in the case of one camera. The third module tracks the mobile objects on a long term basis using model of the expected objects to be tracked. The last module 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.

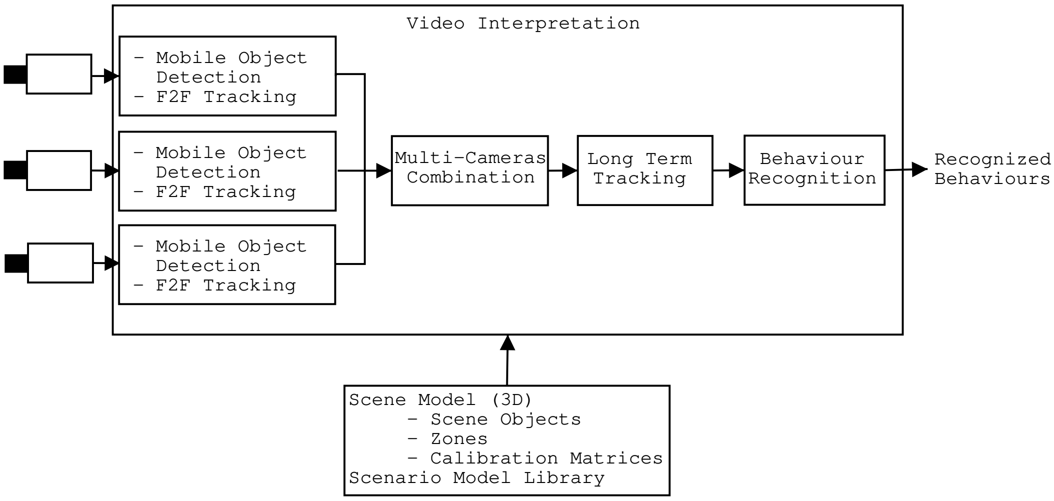


Figure 2. Algorithm of the interpretation of image sequences.

4. Application Domains

4.1. Overview

Keywords: *astronomy*, *bio computer science*, *environment*, *health*, *hydraulics*, *multimedia*, *pollen*, *safety properties*, *transportation*, *visual-surveillance*.

Applications achieved in the Orion team are useful to validate both our research directions and ours models. We are mainly involved in the following applicative domains : astronomy, health, environment, video-surveillance and transportation. Our approach supports two goals. A first scientific goal is to bring new techniques in other domains. For instance, in astronomy for the automatic classification of galaxies. A second goal concerns industrial issues in order to develop operational systems as for instance intelligent visual-surveillance of underground stations. Besides imagery is the main focus of the team, we have also developed 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 discretisation 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.

Moreover, our theoretical approach in the software engeenering field may be applied in a more general context: as a consequence, the theory we developed to enforce safety properties of software tools we developed, can be applied to critical system verification.

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 [48][45]. We are expert in this domain both concerning the image processing of galaxies field and theoretical models of morphological classification. This application is a reference to validate our models and software related to interpretation for complex objects understanding and to program supervision [46][50].

4.3. Video-surveillance

The growing feeling of insecurity among the population led the private companies and also the public authorities to deploy security systems in order to protect their equipment or their commercial interests. For the safety of the public places, the video camera based surveillance techniques are more and more 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 has to 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 project ADVISOR and industrial projects RATP, CAssiopee, ALSTOM and SNCF. A first set of very simple applications for the indoor night surveillance of supermarket (AUCHAN) showed the feasability of this approach. A second range of applications has been investigated and it corresponds to 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 shaked 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 and train car surveillance. 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 healthcare and environment are dedicated to automatic pollen grain recognition. We aim at providing tools with the palynologist so that they can process 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 technician 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 ([39] and [49]). Nevertheless, individual recognition remains of great interest. That is why the Orion team has been studying this problem since 1996 [42].

Due to the complexity of the different types of pollen grains, palynologist knowledge is taken into account [8]. 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 [47][44][40][41].

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 the 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 Chamber of Agriculture as a major issue in ornemental 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 appearences. 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 unsufficient 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 takes 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. Synchronous Model Development

In the domain of Formal Methods, Orion (A. Ressouche) is interested in applying a mathematical based approach to get a safe re-usability of the LAMA platform libraries. To apply synchronous modeling results, in collaboration with V. Roy (CMA Ecole des Mines de Paris) and D. Gaffé (CNRS and UNSA), we study the problem of synchronous language modular compilation. Last year, we defined a synchronous language with modularity facilities and sound behavioral semantics in terms of process algebra model. This year, we defined a circuit semantics that allows us to compile a program into a boolean equation system and we prove that this semantics is equivalent to the behavioral semantics. The circuit semantics gives us a practical means to compile programs. Moreover, we define 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.

The motivation for this work is the specification and verification of critical systems. Many realistic applications accept different levels of description involving entities of various kind which will be modeled differently : some via a controller (synchronous), some via a model specific to their sequential activity (asynchronous). Very often, the critical part of a system is its controller part and specifying it with synchronous formalisms offers a suitable framework for applying model-checking techniques. Moreover, compositional synchronous formalisms guarantee scaling up feasibility since it allows for separate compilation and verification. In order to be efficient in application specification and verification, we must be able to deal with large systems.

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 à 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.

5.3. VSIP

VSIP (figure 3) is a Video Surveillance Interpretation Platform written in C and C++. The goal is to build a generic environment applicable as a first step to video surveillance of banks and subways. Besides the image acquisition hardware, the platform is built from three software components: image processing for people detection, human tracking, and interpretation of behaviors relative to the people evolving in the scene. The platform takes as input video streams from several cameras, a geometric description of the unoccupied scene and a set of behaviors of interest specified by experts of the application domain. For each detected event, the algorithms emit automatically an alert and store an annotation in accordance with the set of predefined behavior models. The platform has been adapted to work in embedded environment such as train. It has also been tested in outdoor scenes, more particularly in airport apron areas, to monitor the surrounding of the parked aircrafts. It will be validated in a bank agency, December, 2004, in live conditions.

6. New Results

6.1. Program Supervision

Participants: Julien Canet, Naoufel Khayati, Sabine Moisan, Jean-Philippe Vidal.

This year we have first focused on extensions of the program supervision task for distributed applications on the Web and we have also introduced a new module for symbolic curve management in the platform. Our major applications are medical imaging and river hydraulics.

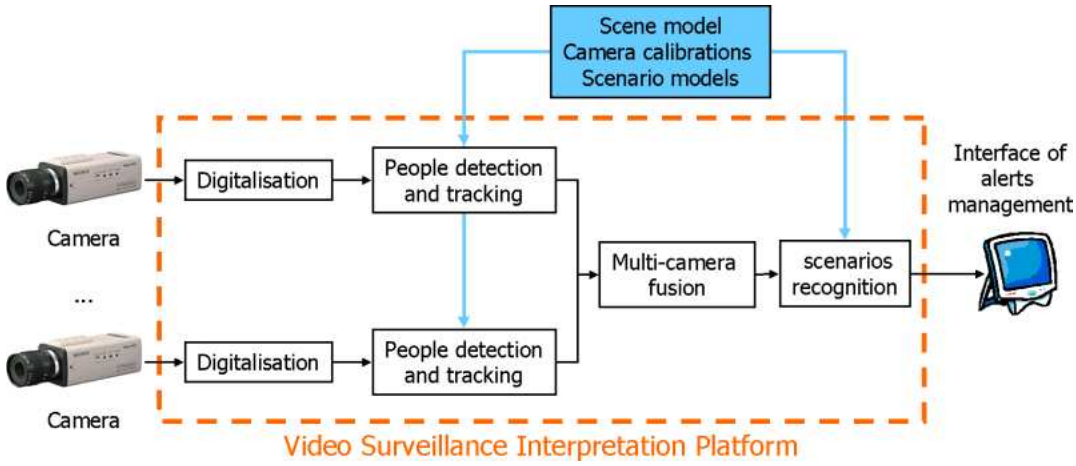


Figure 3. Components of the Video Surveillance Interpretation Platform (VSIP).

6.1.1. Program Supervision Site

Participants: Julien Canet, Naoufel Khayati, Sabine Moisan, Najet Soudi.

The purpose of the program supervision site is to offer a distributed access to program supervision knowledge bases, programs, or knowledge-based systems. A remote user (for instance, a physician or an image processing researcher) can connect to that site and use a delocated graphic interface to either simply visualize knowledge, or submit a request (data to process) to an existing KBS (and follow the execution steps), or even to share some programs and their usage knowledge by building a common knowledge base.

We have first developed a Java Applet version of the graphic interface of our main program supervision engine. This interface (named GIPSE) can now run on a Web Browser (with the Java Plug-In). It offers different graphic facilities to follow the evolution of a KBS reasoning or to build a common knowledge base, involving programs and knowledge from different sites.

In addition, we have set up an application server on an INRIA's experimental network computer in order to prepare a program supervision site. The first version of this site contains some explanations on program supervision and a few knowledge bases. It allows users to edit knowledge bases with GIPSE Applet. The program supervision site also features an administrator section that allows authorized users to manage permissions on knowledge base files. This section has been developed using JSP (Java Server Pages) and Java Servlet technologies.

We have also created a Java remote file system that allows to mount knowledge bases files located on distant computers on a common file system tree and to access them transparently as if they were located on the same computer. GIPSE has been upgraded to access knowledge bases either on the local file system or using the remote file system. Thus it becomes possible to build distributed program supervision knowledge-based systems. Users can keep control on their (local) programs and on the knowledge about the use of these programs, yet sharing them with other users. Our test application is an image processing supervision KBS for osteoporosis detection, in collaboration with physicians and image processing researchers from France and from Tunisia.

6.1.2. Symbolic Curves Module Integration

Participants: Julien Canet, Sabine Moisan, Jean-Philippe Vidal.

Evaluating output data of programs is often complex, even for a human expert, and in real-world applications few automatic methods exist. A usual way to manage complexity in many domains is to display results in the form of curves that offer a simplified summary of results and that can be interpreted and compared to expected results.

This year, we have integrated a symbolic curve module into the LAMA platform. This module allows to reason about symbolic properties of curves in knowledge-base systems rules and thus to obtain a fully automatic evaluation in the cases where results can be obtain in the form of curves, which is fortunately the case in many scientific applications. Using the curve module for result evaluation combined with our general program supervision repair mechanism, it is possible to design systems that can run automatically and that can accommodate changing situations.

The module computes a symbolic description of a curve first by filtering and segmenting the curve and then by describing the different elements (segments, peaks...) of this segmented curve using a "dictionary" of expert-defined symbolic qualifiers for their features (such as the width of a segment or the shape of a peak). Such symbolic qualifiers highly depend on the target domain and on the type of data represented by the curves. The mapping between feature symbols and numerical (ranges of) values must therefore be provided by experts.

In addition to curve descriptions, the module comes with comparison facilities between two curves and between points and a curve, in terms of distance and position. Symbols describing relative position (e.g. above/under) and distance (e.g. close/far) of points with regards to a curve are also defined by the expert in the dictionary. Thus experts can symbolically compare curves or sets of points to reference ones, which is useful to compare a curve resulting from experiments to a theoretical model..

We have extended the expert knowledge description language YAKL with symbolic curves management operations. Symbolic feature qualifiers and results of comparison can be accessed and tested in rules. The expert is now able to define specific symbolic dictionaries for application domains and then to reason on symbolic features of curves and sets of points to decide whether results are correct or not.

The integration of the symbolic curve module has been tested and validated on both Linux and Windows systems. It is currently used by Jean-Philippe Vidal for the hydraulic calibration prototype.

6.1.3. Distributed Program Supervision

Participants: Naoufel Khayati, Sabine Moisan.

In the framework of a cooperation with Tunisia (ENSI Tunis) on distributed program supervision for medical imagery, a new co-directed PhD has started at the beginning of this year. The objective is to study various distribution methods of program supervision knowledge-based systems (KBS) for medical imagery. 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 team work and sharing of knowledge for new knowledge-based systems.

As a first step during this year, we started a bibliographical study in the domain of distributed (knowledge-based) systems. The result of this study is a synthesis which enabled us to trace possible directions for the distribution of program supervision systems. We studied the applicability of P2P architecture, of GRID architecture and of Multi-Agents Systems. Our application will utilize a few tens of peers and of physicians' sites and few centralized servers, and will have to be powerful enough to offer complex functionalities such as the execution of distant programs and the access in various modes to data and knowledge.

As a result of this study, we propose a first version of a distributed program supervision system, using a simple GRID architecture equipped with a metadata warehouse – to localize the various resources (programs, data, knowledge files, etc.) – and managed by a multi-agent system based on mobile agents. Hence, as a second step, we established a scenario of use for this first version, based on an architecture containing a supervision server, which plays the role of an intermediary between the physician and the rest of the system, composed of program servers, knowledge servers and the PEGASE+ supervision engine.

- The supervision server is accessible from our program supervision site at Inri;.
- In this first version, the PEGASE+. engine will run on this server.

- The knowledge description language for program supervision (YAKL) must be adapted to distribution needs (e.g., to indicate remote calls for programs);
- The agent platform that we chose to implement mobile agents is the "IBM Aglets" which is 100% Java. It will implement agents that communicate necessary information and data between the engine and the programs or among programs.

Currently, we are implementing and testing these first architecture and scenario on local machines. Our test application is a KBS for medical image processing, to detect osteoporosis on bone images.

6.2. Software Platform for Cognitive Systems

Participants: Julien Canet, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Jean-Philippe Vidal.

This year, concerning the verification of engine behavior, we have extended the formal approach for component verification. Concerning knowledge-based system engines, we mainly concentrated on a new engine for the model calibration task, with an application in river hydraulics. We have also changed the storage format of the LAMA *knowledge bases to use standard XML.*

6.2.1. Model Calibration Task

Participants: Sabine Moisan, Jean-Philippe Vidal.

In the framework of a co-directed PhD thesis (with Cemagref Lyon and INPT Toulouse), we are studying the model calibration task. We propose an approach to support model calibration that combines knowledge-based techniques and standard numerical modeling ones to help end-users of simulation codes.

Our current work deals with two main topics: first, enhance an hydraulic knowledge base and the corresponding prototype of knowledge-based system in order to face various real-life cases of river model calibration and second, design a new engine dedicated to numerical model calibration.

This year, we have first completed our knowledge base for numerical model calibration task, applied on river hydraulics. This development was done using the existing program supervision engine PEGASE+. Indeed program supervision and calibration tasks have a lot in common even though PEGASE+ is not completely adapted to calibration needs. This experiment allowed us to formally identify three levels of knowledge, which have been described in [33].

From this knowledge base, we built a prototype knowledge-based system. This prototype is dedicated to assist an end-user of the MAGE simulation code to achieve a river model calibration activity. The prototype has been used to solve a real-life case: the calibration of a model of the Hogneau River. Encouraging results have been presented in [32].

Moreover, on the basis of the identified domain-independent calibration knowledge and on the river hydraulics knowledge, we have also established a generic methodology for "good calibration practice" in river hydraulics. This methodology includes an ontology for river hydraulics together with a paradigm for the calibration process. This approach is detailed in an [25].

Finally, from this theoretical work and from our experiment with the prototype knowledge-based system, we have fully specified a new engine dedicated to model calibration. We have started to design this engine, based on our LAMA software platform. The platform allowed us to reuse code and concepts from the general layer, as well as from the program supervision one, such as the notions of rules, data or operators. We have also developed a new expert language – an extension of the existing YAKL language – to accommodate model calibration expression needs.

6.2.2. XML Format for Knowledge Base Contents

Participants: Julien Canet, Sabine Moisan.

Knowledge base contents are expressed by experts using one of the proposed knowledge description language (for example in YAKL language, for program supervision). These languages are expert-oriented and they contain a lot of “syntactic sugar”. A more tractable format for knowledge manipulation and exchange (with graphic interfaces or other tools) is therefore useful. We have introduced in LAMA an XML storage and exchange format, in replacement of the previous ascii one, in order to represent the knowledge described by experts in their knowledge bases. It is not only a mere storage format, but also a communication means with other tools, and especially tools that are outside the scope of the LAMA platform. All the knowledge bases written on top of LAMA toolkits have now a standard format, that can be understood and handle by many other tools on the market place. We kept an expert-friendly form (such as the YAKL language), but its translation into XML (or even OWL) allows systems designed using the LAMA platform to interface, for instance, with existing Web-tools for ontology descriptions, or with any other tool relying on XML for its data.

This format is described in XML Schema and the corresponding schema is hosted on Orion web site (for verification with XML parsers) at http://www-sop.inria.fr/orion/yaklschema.xsd. In addition to this schema, we have developed libraries, available in C++ and Java, that allows to save and restore knowledge pieces into and from this format. The knowledge pieces correspond to the meta-model of the target tasks of LAMA, namely, program supervision, calibration and data interpretation tasks. These libraries have been integrated into the expert language parsers in order to generate XML files.

The Java toolkit of LAMA for graphic interface design has been updated to support this new format. In particular GIPSE, the GUI of PEGASE+ program supervision engine, now understands that format. We are also working on the connection of our systems with front-end OWL-based tools, to facilitate “upstream” knowledge capture.

6.2.3. Component Framework Verification

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

Last year, we defined both a dedicated language (BDL) to express BLOCKS component behavior and a *synchronous* mathematical model to give a semantics to BDL programs. 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.

This notion of safe substitutability can lead to at least two different approaches: define design rules that, by construction, ensure the correct usage of the components or verify afterwards that correct substitutability holds.

We have already defined several design rules, thus this 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.

Another point is that, in order to incrementally support design of engines from components, our verification process should be both compositional, that is going bottom-up, and modular, that is going top-down. To get compositional model-checking, we proved that temporal logic properties are stable with respect to BDL language operators. Concerning composition, we proved that the "assume-guarantee method" is possible for BDL programs. In this method a global property is decomposed into local properties following the natural decomposition of the system. Of course, the conjunction of the local properties should imply the global one. Our formalism and its properties are detailed in and [31].

However, our present formalism to deal with behavioral substitutability is somewhat too strong and we are studying ways to relax it. We are also investigating how to apply the assume-guarantee method automatically; this involves finding an algorithm for decomposing global properties into local ones. Finally, the implementation of a substitutability analyzer for KBS engines is underway.

6.3. Automatic Interpretation of Image Sequences

Participants: Alberto Avanzi, Jihène Bannour, François Brémond, Bernard Boulay, Binh Bui, Frédéric Cupillard, Gabriele Davani, Benoit Georis, Florent Fusier, Magali Mazière, Le Son Pham, Monique Thonnat, Christophe Tornieri, Thinh Van Vu, Valery Valentin.

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 : two transfer actions Telescope3 and VIDEA, the European project AVITRACK, the Predit project (departments of research and defence) and the five following industrial projects: Intelligent Cameras(STmicroelectonics), PFC, CASSIOPEE, SAMSIT and RASV.

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. So, we try to recognize predefined scenarios based on visual invariants. The system in charge of the second problem is the module of scenario recognition. Our approach to image sequence interpretation is based on the a priori modeling of the observed environment.

This year, we have refined our works on mobile object detection to correct mobile object position in case of static occlusions. We have extended the platform to recognize scenarios also involving people interacting with vehicles. We have improved our approach for the recognition of human postures and complete the algorithm for the real time recognition of temporal scenarios. We have also focus on the evaluation of the video understanding platform at two levels, using both recorded videos and pilot sites for different applications. We have started using the program supervision framework to control the interpretation platform and started conceiving a platform architecture to easily build a dedicated interpretation system from the platform for new applications. We have also started building tools (user center interface) to help users to defined scenarios models.

6.3.1. Learning and evaluation of mobile object lateral forms

Participants: Binh Bui, François Brémond, Monique Thonnat.

One of the main stage of automatic video interpretation is to classify the mobile objects in the scene because they can correspond to several types of different objects (a person, a push-chair, etc). To solve this problem, we have proposed a new method using a Bayesian network for the classification of mobile object shapes observed from the side (lateral form). This year, we have evaluated the system on both recorded sequences and real-time prototype.

Learning of lateral forms

Using the software developed last year that automates the learning stage, we have built five models of lateral forms: "adult alone", "two adults side by side", "adult with luggage", "luggage", "child" and "noise". The class child is defined as a person under 1.10 meter. For each model (class), we have used ten typical sequences (representative of the class) in the learning stage, corresponding to hundreds of frames.

Classification of the lateral forms

For each blob b detected, we calculate its degree of membership $D(b|F)$ to all models F of lateral forms. The degree of membership $D(b|F)$ is the ratio of the probability that the blob b belongs to the class F divided by the probability that the blob b belongs to all other classes. The class of the blob b is the class having the highest degree of membership.

Results

This year, we have focus on the evaluation of the interpretation system. We have tested the module with several hundreds of images taken from sequences which are different from the sequences used to build the

models: 5 videos (about 400 images) characteristic for "adult alone", 5 videos characteristic for "two adults side by side", 3 videos (about 250 images) characteristic for "adult with luggage" and 5 videos characteristic for "child". The result is convincing. More than 90% of adults and luggage are correctly recognized. More than 80% of children and adults with luggage are also correctly recognized. The success rate for the class "two adults side by side" is about 60% due to the lack of examples used in the learning process.
The results are satisfying, the system enables to classify mobile object lateral forms even in case of corrupted data, thanks to the uncertainty management. The next step is to handle missing information due to sensor failures.

6.3.2. Human Posture Recognition

Participants: Bernard Boulay, François Brémond, Monique Thonnat.

This year, we have continued previous work on human posture recognition by combining a classical 2D approach and use of a 3D human model. The 3D model (cf. figure 4-c) is defined by 9 degrees of freedom (articulations): the abdomen, shoulders, elbows, hips and knees. We first define for each posture of interest a specific set of parameters. These parameters are the three Euler angles for each articulation.

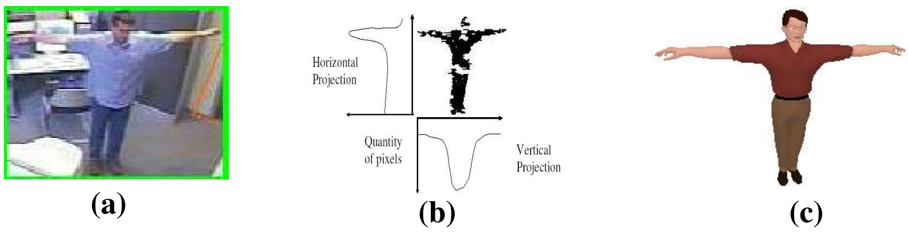


Figure 4. The image (a) depicts a man in T-shape posture, in image (b) we can see a blob and its horizontal and vertical projections, image (c) represents the 3D human model in T-shape posture

For each detected person, the videounderstanding platform (VSIP) provides us with the position of this person in the 3D space. Then we compare the silhouette of the detected person with reference silhouettes generated from the 3D model. We project the 3D model on the image plan for reference postures which have been generated using the 3D position of the person. To determine the orientation of 3D model we test all possible orientations using a rotation step. Then we compare horizontal and vertical (H. & V.) projections of these images with the (H. & V.) projections of the detected person silhouette. The horizontal (resp. vertical) projection on the reference axis is obtained by counting the quantity of motion pixels which correspond to the detected person for each row (resp. column). An example is shown in figure 4 for T-shape posture.

This year, our first goal was to tune the parameters of our posture recognition algorithm to obtain better results. The most important parameter is the rotation step. To determine the optimal rotation step, the recognition rate of postures on synthetic data are computed for different rotation step values. For this experiment we have used a 3D woman model (cf. figure 6) which is different of the 3D model used for recognition. We choose a 36 degrees step because it gives a better ratio between recognition rate (76 %) and computation time (1.5 im/s). We have performed tests on synthetic data by giving the correct orientation which show that computation time is divided by ten, and recognition rate increase from 76% to 95%.

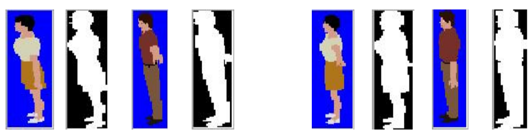
Our second goal was to evaluate our approach. We obtain good results on real videos of about 2000 frames depicting targeted postures (figure 5): 89% of correct recognition for standing postures and 88% for seated-bending postures. A more accurate analysis gives 100% of correct recognition for standing posture with arms up (left and right arms are not discriminated), 62% for standing posture with arms near the body, 87% for T-shape posture, 75% for seated postures, and only 40% for bending posture.

Standing postures: 89%						Seated–bending postures: 88%					
Standing posture with arms up: 100%		Standing posture with arms near the body: 62%		T–shape posture: 87%		Seated on a chair posture: 54%		Seated on the floor posture: 57%		Bending posture: 40%	
na:100	a:100	na:83	a:41	na:93	a:0	na:61	a:40	na:71	a:13	na:70	a:24

Figure 5. Recognition rate of postures organized in hierarchical way for real videos (a: ambiguous case, na: no-ambiguous case)

We define an ambiguous case as a situation where the silhouettes of a person in two different postures are very similar (cf. figure 7). We determine ambiguous case by using synthetic data. A synthetic video is made for each posture and for all possible orientations. Then our algorithm is applied, an ambiguous case is found when the posture is not correctlty detected. For example, no-ambiguous case of the bending posture corresponds to the situation where the camera observes the person from the side. By discriminating ambiguous cases we obtain an average of recognition rate of 70% for non ambiguous cases and only 24% for ambiguous cases on real videos.

Figure 6. 3D woman model used for experiments with synthetic data



(a) **(b)**

Figure 7. (a) Standing posture with arms near the body recognized as T-shape posture,(b) T-shape posture recognized as standing posture with arms near the body

About 35% of wrong recognitions are due to the ambiguity problem, and the other to problem of segmentation. To accelerate our algorithm we plan to compute features from the detected person (orientation, 3D height, ...). We also want to use information provided by the tracking phase to improve results by taking advantages of temporal analysis.

A part of this work was described in [35].

6.3.3. A context-aware classification module for occluded objects

Participants: Gabriele Davini, François Brémond, Monique Thonnat.

In many video interpretation applications, mobile objects may be partially occluded and thus not visible in their wholeness. As a consequence, when computing their 3D position and dimensions, if only the visible part is taken into account, errors may be committed. Moreover, standard appearance-based context-unaware classification algorithms may fail, assigning the objects with wrong types or even considering the objects as noise.

We can identify three types of occlusions:

1. the occlusions due to the vision field of camera, which occur when the mobile object is on the border of the vision field of camera;
2. the occlusions due to contextual objects, that occur when one or more contextual objects are placed between the mobile object and the camera;
3. the occlusions due to (other) mobile objects.

Occlusions may be static or dynamic, depending on whether the occluded regions change in time. Usually, the first two types of occlusion are static, since the camera and contextual objects do not move; anyhow, there are some exceptions as, for example, the doors, the windows or the chairs on wheels. On the contrary, third type occlusions are typically dynamic.

In order to cope with static problems of occlusion, we propose an algorithm which exploits the *a priori* knowledge of the vision field of camera and of contextual objects so as to estimate the occluded parts of analyzed objects. Second type dynamic problems of occlusion are left to specific occluding object-based algorithms, while third type and mixed, i.e. multiple type, occlusion's issues are left to the tracking module.

The proposed algorithm is composed of two parts: the first one deals with first type occlusions, the second one deals with second type static occlusions. Both sections start with a series of tests that are run to establish whether the mobile object is occluded and where the mobile object may have a hidden part. For example, if a person is seated behind an armchair and the visible part is just his/her head, we may conclude that there is an occlusion due to the armchair and then a hidden part may be found below the head.

At this point, in order to proceed with the analysis, we check whether the visible part of the object is compatible with a mobile object model. This means that we test whether the mobile object's 3D dimensions are akin to those of a mobile object, as specified by its model. For example, if a mobile object is found to be occluded on its left side, we first verify whether its height is greater than the minimum height of the considered model object, since the occlusion on object's left side does not affect its height. Then, we check whether its width is reasonably close to the minimum mobile object's width. Exceptions to this rule are made on a per model basis, for some specific situations, as when the visible part is a head, that is not compatible with person's standard width.

Since the visible part of an object may comply with more than one mobile object model, the preferred model is that with the the highest compatibility.

In the case of first type occlusions, an occluded mobile object, which complies for at least one 3D dimension (width or height or length) with a mobile object model, is modified so as the two other dimensions are the standard 3D dimensions of the model; eventually, its 3D position may also be recomputed. At this point, a new classification is run so as to assign the object with a type, which is that of the mobile object model considered or that of a compatible model with the latter.

In the case of second type static occlusions, static contextual objects are first scanned in order to find those that partially occlude the mobile object; also, this allows for evaluating the maximum extension of the mobile object behind them. In the following, a number of hypotheses about the analyzed object 3D position and dimensions are made, till the first of them succeeds in complying with the considered mobile object model. At this point, as before, the analyzed object is modified and reclassified.

The proposed algorithm has been successfully tested on video sequences with different scenarios; an example of its application is provided in the pictures in figure 8.

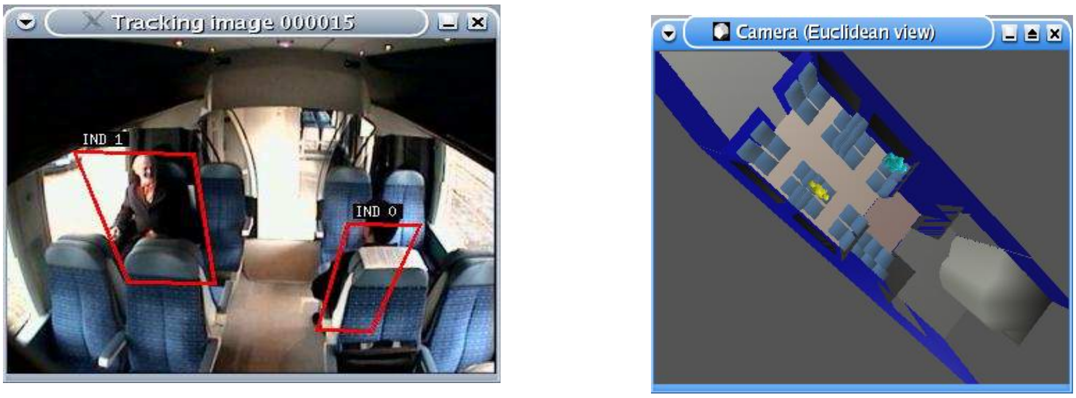


Figure 8. Two occluded objects correctly recognized and positioned.

6.3.4. Vehicle Activity Recognition

Participants: Florent Fusier, François Brémond, Monique Thonnat.

An important task for this year was to extend the VSIP functionalities to allow vehicle activity recognition for new application domains like apron monitoring. We want to detect people and vehicle (cars, trucks, aircraft...) and understand their activities and interactions :

- an adapted segmentation to outdoor conditions;
- a classification module which classifies people and vehicles;
- a tracking module which tracks people and vehicles;
- a background management algorithm which integrates parked vehicles into the reference image;
- a scenario recognition module which recognizes scenarios involving vehicles.

Segmentation (motion detection) :
The new challenge of the segmentation consists in being robust to outdoor conditions : illumination changes, strong shadows, weather conditions, vehicles having a low contrast with the background. The segmentation algorithm has been tested with several representative sequences. Without modifications, the current algorithm seems robust to illumination changes and the detection of mobile objects gives good results even when vehicles have a low contrast with the background. The algorithm is real time and robust to noise. The point to improve is mainly shadow removing, because the sun causes strong shadows for vehicles and people and leads to classification errors.
Classification :
We need to classify several types of mobile objects : people and a set of vehicles with different shapes. Thus we have added vehicle models with different sizes (small vehicles, deformed vehicles, large vehicles, aircraft) into the knowledge base characterized mainly by their bounding box aspects (height, width, height/width ratio) and their moving pixels density.

Tracking :
The VSIP Tracker was basically developed to track people and we have made some improvements on it to allow vehicle tracking. We have adapted the tracking parameters and added tracking vehicle models specifying how can evolute their characteristics (size, trajectory). The platform can track people and vehicles such as aircraft as shown on Fig 9.
Background management :

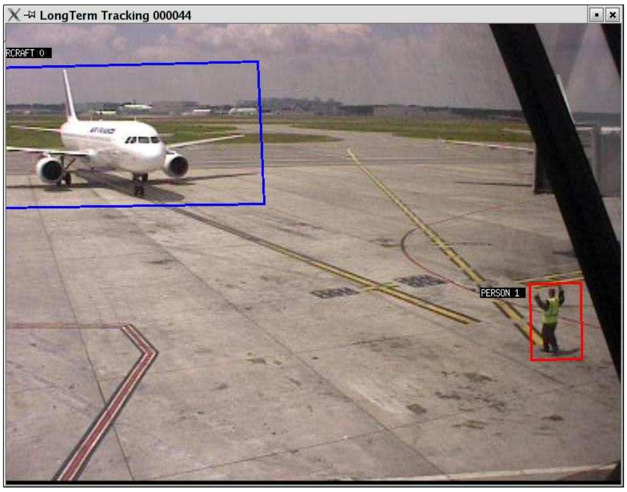


Figure 9. An aircraft and a person tracked on an airport apron

An issue in recognizing vehicle activities is to handle people interactions with vehicles. Thus we need to detect and track people around vehicles when they have just parked. To allow this, we have developed a new background management algorithm which integrates the parked vehicles into the reference image by determining precisely when vehicles stop and start. This method allows to detect and track people who interact with vehicles, such as in maintenance operations.
Scenario Recognition :
We have extended the scenario recognition module too. This module was mainly used to recognize human activities, but in new application domains, activities can involve vehicles too. So we have defined new types of physical objects and events in the video events ontology. Without modifying the recognition engine, we have recognized several scenarios such as a person preparing the aircraft arrival (see Fig 10).

Figure 10. A maintenance person deposited the chocks before the aircraft arrival

The VSIP platform is now opened on more application domains in which not only human beings are involved. It is now possible to classify and track vehicles, and to recognize activities with people and vehicles. This adaptation was mainly achieved by adding new models for vehicles and demonstrates the capacity of the platform to adapt easily to new functionalities.

6.3.5. Global tracker

Participants: Frédéric Cupillard, François Brémond.

ORION video interpretation system is based on the cooperation of a vision and an interpretation module. The final task of the vision module is to track individuals evolving in the scene thanks to the **Long Term Tracker** (**LTT**). Then, trajectories of tracked individuals are provided to the interpretation module for further human behavior analysis.

Currently, the main limitations of the LTT are:

1. loss of tracks caused by dynamic or static occlusion (individual occluded by another individual or a contextual object) or bad segmentation;
2. over detection: detection of one individual who is not in the scene (for example, when there is significant noise);
3. under detection: detection of less individuals than there are in the scene (for example, when there is a lack of contrast);
4. mix of individual identities when individuals are merging and crossing each others.

To cope with these limitations, we have developed a higher level module, called **Global Tracker**. The goal of the Global Tracker is to detect potential LTT errors and correct them. To achieve this, it maintains a global understanding of the scene using the LTT outputs (trajectories) and the 3D scene model (zones of interest, objects). The Global Tracker can be seen as a set of generic rules enabling to adapt the output of the vision module to requirements of a specific application. For example, to adapt the Global Tracker to a new scene, we can define the zones fitting with geometry and semantic of that scene.

The Global Tracker currently reduce LTT limitations 1 and 2 (defined above). To solve the LTT limitation 1, a finite state automaton (as described on figure 11) maintains individual tracks whenever the LTT loses tracks while it should not (corresponding to "Incorrect cases" in the automaton). Indeed the meaning of figure 11is the following: a transition in the automaton (characterized by a white rectangle with dotted boundary lines) corresponds to a particular configuration when a person disappears (that person is not tacked anymore by the LTT). A state (characterized by a white rectangle with continuous boundary lines) represents the way a disappeared person is managed. Either this disappearance is considered as a mistake; the track of the person is maintained ("Incorrect cases") or this disappearance is considered as normal; nothing is done ("Correct case"). The initial automaton state (characterized by the white ellipse) corresponds to a person who is normally tracked or tracked again by the LTT.

To solve the LTT limitation 2, the Global Tracker relies on the life duration of individuals tracked by the LTT. When the life duration of an individual is under a predefined threshold (for example, 4 frames) the Global Tracker considers the detection of that individual as a wrong detection.

First, we have defined two semantic attributes characterizing the zones of interest: An **"In/Out"** zone, where individuals enter or leave the field of view of the camera (for example, a zone in front of a door) and a **"Problem"** zone, where significant dynamic or static occlusions occur or where segmentation is not reliable. A zone can be both an "In/Out" and a "problem" zone. These new semantic attributes aim at determining whether the loss of tracks is justified or not. When it is not justified, the automaton allows individual tracks to be maintained according to three situations: **dynamic occlusion**, **static occlusion** or **no occlusion**.

The Global Tracker has been evaluated on four functionalities for a building access control application with a large set of recorded video sequences (about 150 000 frames): 21 videos of an inside building corridor and 17 videos of an entrance building interlock. The four functionalities are: (1) people presence detection, (2) people unicity control (to detect if there are one or several persons), (3) people counting and (4) people origin/destination estimation. The Global tracker has allowed a significant improvement for functionalities (3) and (4): 20% of improvement for the functionality (3) and 62.4% for the functionality (4) and a small improvement for functionalities (1) and (2). Indeed, the Global Tracker performs better when tracking individual trajectories. The next step of the Global Tracker consists in extending its set of rules to bring other solutions to carry on improving LTT limitations 1 and 2 and cope with the two other LTT limitations: under detection and mix of individual identities.

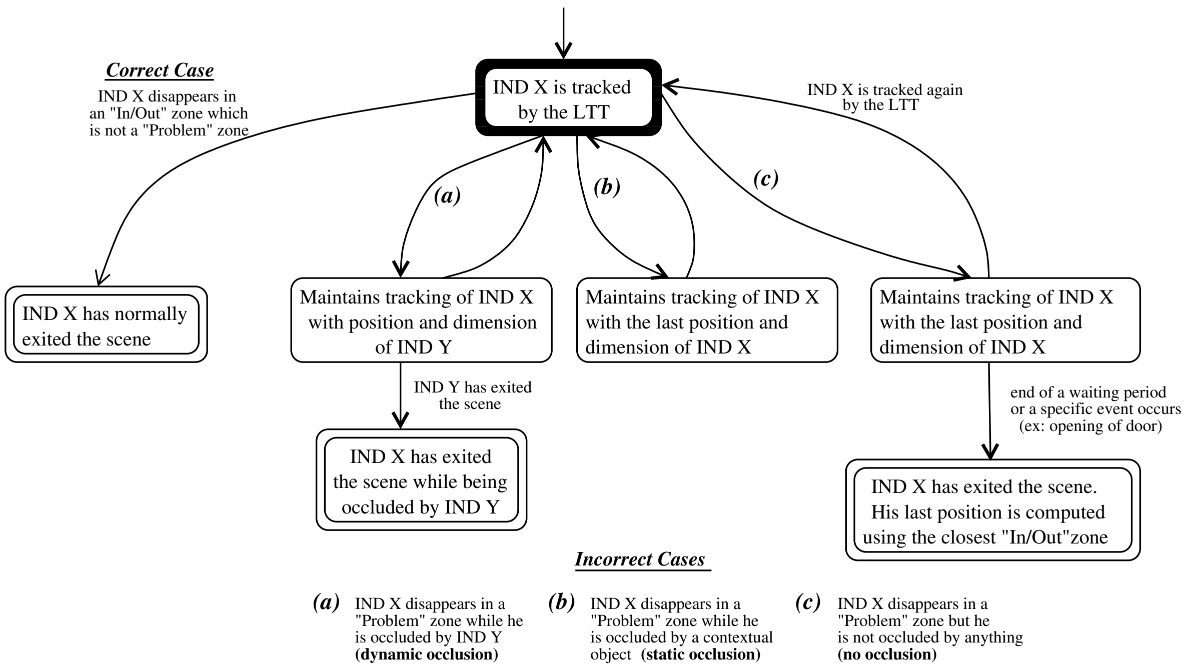


Figure 11. This figure illustrates the Global Tracker set of rules represented by a finite state automaton.

6.3.6. Temporal scenarios for automatic video interpretation

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

Our goal is to study the problem of *Temporal Scenario Representation and Recognition for Automatic Video Interpretation*. In particular, the most interesting point is to design an algorithm recognizing in real-time temporal scenarios predefined by experts and taking as input mobile objects tracked by a vision module and a priori knowledge of the observed environment. We have proposed a novel approach taking advantages of two state of the art approaches: one proposed by N. Rota (a former Orion Phd thesis) and the other proposed by M. Ghallab and C. Dousson at LAAS in 1994. To do this, we have to solve two issues: scenario representation and scenario recognition.

Concerning the issue of temporal scenario representation: we have proposed a video event ontology (in collaboration with an ARDA workshop series on video events) that can facilitate the representation of temporal scenarios. This ontology is composed of concepts to describe physical objects (e.g. mobile-objects, contextual-objects), video events (e.g. primitive-state, composite-event) and relations between concepts (e.g. temporal-relations, spatial-relations) [[34]]. Then, based on this ontology, we have proposed a description language helping experts of different domains to describe easily their scenarios of interest. The ontology and the language are/have been used by experts of four European/French projects for video surveillance (CASSIOPEE for bank agency surveillance, AVITRACK for apron monitoring, SAMSIT for inside train surveillance and ADVISOR for metro station surveillance). For example, Fig. 12 shows a representation of a scenario model "bank_attack" in the context of a bank agency.

```
Scenario(bank_attack,
  physical-objects(emp : Person, rob : Person)
  components(
     (emp_at_pos : primitive-state inside_zone(emp, "back_counter"))
     (rob_enters : primitive-event changes_zone(rob,"entrance","counter"))
     (emp_at_safe: primitive-state inside_zone(emp, "safe"))
     (rob_at_safe: primitive-state inside_zone(rob, "safe")) )
  forbidden-scenarios(
     (any_in_bank: primitive-state inside_zone(anyP, "bank")) )
  constraints(
     (emp_at_pos  before emp_at_safe)      // (1): temporal
     (rob_enters  before rob_at_safe)      // (2): temporal
     (rob_enters  during emp_at_pos)       // (3): temporal
     (rob_at_safe finish emp_at_safe)      // (4): temporal
     (anyP ≠ emp)                          // (5): forbidden
     (anyP ≠ rob)                          // (6): forbidden
     (any_in_bank[ ≤ rob_enters[)          // (7): forbidden
     (rob_enters[ ≤ any_in_bank]) )        // (8): forbidden
  decisions(Alert("Bank Attack!"))  )
```

Figure 12. Representation of the scenario model "bank_attack": this scenario model is composed of four components. This scenario description contains six constraints. The constraints (1)-(4) express a sequence of components composing the scenario. Constraints (5) - (8) in connection with the forbidden scenario express that there is no other person (anyP is a person different from com and rob) in the bank during the attack.

Concerning the recognition issue, a challenge is to reduce the complexity of the temporal scenario recognition algorithm. The recognition can be viewed as a Temporal Constraint Satisfaction Problem (TCSP). So, the algorithm to solve this problem belongs generally to the NP-complete class. Thus, the other tasks of our work concern the reduction of the processing time for the recognition algorithm. To solve this problem, we have proposed a temporal scenario recognition method that is able to recognize in real-time predefined scenarios in real-time (in video cadence).

Concerning the comparison between the Proposed Approach (PA) and the Chronicle Recognition Technique (CRT) [Dousson & Ghallab, 1994], we would like to suggest recommendations on how to use these two techniques for video interpretation: - CRT does not store recognized scenarios but all partially recognized scenarios. Thus, CRT is efficient for the recognition of frequent scenarios. This allows us to suggest that CRT can be (efficiently) used for monitoring applications. - CRT gets into a combinatorial explosion in function of the number of physical object variables defined within each scenario model. Thus, this technique should not be used for multi-physical-object applications. - Proposed Approach does not store partially recognized scenarios but stores recognized scenarios. Thus, PA is really efficient for the recognition of rare scenarios. This allows us to suggest that PA can efficiently be used for surveillance applications. The experiments that we have realized show that PA is also efficient for monitoring applications. Thus, PA can be used for monitoring applications. - Moreover, PA can process efficiently multi-physical-object scenario models. Thus, PA can be used for multi-physical-object scenario applications.

To validate the recognition algorithm, we have cooperated with the experts of four projects for video surveillance/monitoring to realize three types of tests with: recorded videos, live videos and simulated data. The obtained results answer requirements. The recognition algorithm together with the vision routines can recognize correctly in real-time (10 frames/second) scenario occurrences in a longtime interval (a week).

Moreover, the recognition algorithm can cope with complex videos containing 240 persons and with complex scenarios defined with 10 physical objects and 10 components.

6.3.7. A Program Supervision Architecture for Video Understanding

Participants: Benoît Georis, François Brémond, Monique Thonnat.

During this year, we have focused our research on supervising (control of the execution) the automatic video interpretation platform called VSIP (Video Surveillance Interpretation Platform) using the PEGASE+ program supervision engine. To this end, we have modified the VSIP platform in order to obtain a library of programs which can be supervised. We have also built a knowledge base containing the knowledge about programs using the YAKL description language. This knowledge base is used by the engine to supervise the VSIP platform. Up to now, we have defined 29 operators. An operator corresponds to a video processing or interpretation task with a well defined functionality but that can be implemented by several techniques. There are 2 types of operators: primitive or composite. A primitive operator corresponds to a concrete and elementary video processing or interpretation program (e.g., a background sub-tractor, a moving region classifier) while a composite one is an abstract task describing a combination (sequence, alternative) of several operators. These combinations are defined by a video processing or interpretation expert. Currently, we have defined 23 primitive operators and 6 composite operators in the knowledge base. The system is currently able to track mobile objects on supervised mode in real time and has been tested on several recorded video sequences.

The control of the whole process of video interpretation is achieved by the engine which uses several types of criteria. These criteria are defined at the operator level in the knowledge base:

- Selection criteria: they correspond to a choice between several techniques having the same functionality but having distinct characteristics. For instance, the merge of moving regions can be specialized for merging people or vehicles depending on the expected objects in the target application. In a bank monitoring application, we only need the merging people algorithm;
- Initialization criteria: they assign an initial value for operator parameters. For example, depending on the camera orientation, the 3D position of mobile objects is more or less reliable. If the camera is in a top view configuration, the 3D position is accurate. This initialization criterion enables to set up the default value of the confidence in the 3D position computation;
- Evaluation criteria: performance assessment on operator results is done by using evaluation criteria. Currently, they are based on a comparison between operator outputs and ground truth reference data. For instance, we can assess the quality of the motion detection by comparing the true positives (moving region in good overlap with a ground truth data) with the false positives (moving region not overlapping a ground truth data);
- Adjustment criteria: they enable to modify the value of operator parameters. For instance, when the performance of the motion detection is low (e.g.,the ratio of true positives over false positives is one third, so there is an over-detection problem) we increase the motion detection threshold used for background subtraction.

Operators and their associated criteria have been defined thanks to the program supervision framework. The proposed architecture can thus exhibit several good properties:

- Isolated knowledge property: all the knowledge is extracted and separated away from the code and stored in a knowledge base using this formalism, so that knowledge is represented in a human understandable way. This separation between code and knowledge also allows to better control the tasks at hand;
- Modular knowledge property: knowledge is structured in small parts corresponding to clearly identified video processing or interpretation functionalities;

- Upgradable knowledge property: knowledge is easily maintained and/or extended (capitalization of human expertise).

The challenge is now to determine the knowledge to control operators (e.g., under which hypotheses an operator can be used successfully). To acquire this knowledge we have conceived an evaluation framework which consists in quantifying results, classifying errors and eventually diagnosing problems. This precise evaluation is needed for the video processing expert to be able to define adjustment criteria attached to operators. We are currently collecting and analyzing results we have obtained in automatic evaluation by comparing operator outputs and ground truth data. The objective is to enlarge the knowledge base and to provide repair mechanisms to improve system results and robustness. Nevertheless, obtaining a meaningful evaluation enabling error diagnosis and repair is a complex problem. The main problem is due to dependencies between an operator and its predecessors. For instance, once a repair mechanism is defined for an operator A which enables an improvement of performances at the present time or on given video sequences, we cannot guarantee that we will not encounter a decrease in the performances of A in the future due to a change in an operator B which is preceding A in the video interpretation process. The second problem is that we still lack a clear functionality definition for each operator, mainly because there are different levels of complexity which can be handled by an operator. This complexity is due to the large diversity of real world data which do not correspond to ideal input data described in the operator specification. We thus need to refine our methodology and we are currently investigating an evaluation taking into account dependencies by refining specifications which models the real world data complexity.

A part of this work has been presented in [29].

6.3.8. Design of a Video Understanding Platform

Participants: Alberto Avanzi, Frédéric Cupillard, Christophe Tornieri, Benoît Georis, Magali Maziére, Valery Valentin, François Brémond, Monique Thonnat.

We define three requirements that an activity monitoring platform should meet to enable its re-usability for different applications and to insure performance quality: modularity and flexibility of the architecture, separation between algorithms and a priori knowledge used by them, and automatic evaluation of algorithm results. We also applied a development methodology which enables to fulfill the last two properties and which consists of the interaction between end-users and developers during the whole development of a specific monitoring system for a new application.

We believe that to obtain a reusable and efficient activity monitoring platform, a unique global and sophisticated algorithm is not adapted because it cannot handle the large diversity of real world applications. However such a platform can be achieved if many algorithms can be easy combined and integrated to handle such diversity. Therefore we decided to use software engineering and knowledge engineering techniques to meet these major requirements.

1. **Modularity and flexibility of the architecture**, is a classical software engineering property. To use a platform for deriving new systems for specific applications, it is often necessary to add new algorithms, to remove existing ones or to replace some of them with others which have the same functionality but are able to cope with more challenging situations. **Modularity** enables the platform to be adapted to various applications (metro surveillance, bank agency surveillance, people counting, ...) in various environment (different metro stations for example) because the platform can efficiently combine the modules corresponding to the particular requirements of the application. Nevertheless, a problem is still open: the management of the data exchanges between modules. Our solution to this issue is based on the notion of *shared data manager*. A shared data manager contains a data structure where modules read and write input/output data. A *module* corresponds to a platform functionality: in our case, for example, video acquisition functionality, segmentation functionality or frame to frame tracking. *Input/Output* are all the data exchanged between modules. The shared data manager can be thought as a module which performs “data management and distribution” task, following

the modularity philosophy. The shared data manager manages the way data are exchanged between modules. A module is directly connected to the shared data manager only. To put some data in the shared data manager, the module calls the appropriate method of the shared data manager. The module is not aware of how and when data will be used. Separating data management from module functionality allows, for instance, an application to be distributed on different machines changing only shared data manager implementation.

We call **flexibility** the property of having a set of tunable parameters. This property implies the possibility to configure algorithms and to define different scenarios without changing source code. To fulfill this property, we have decided to make all the internal parameters of every module tunable. To do that, parameters are defined in separate files outside the code (i.e. outsourcing of parameters) using a description formalism. These files are handled by the shared data manager as regular input/output data. These parameters can be changed during processing enabling dynamic parameters optimization.

2. **Separation between algorithms and a priori knowledge used by them**. Using a priori knowledge is not new but keeping it separate from algorithms which use it enables re-usability. Complex systems performing activity monitoring use a huge amount of contextual knowledge of different types. Knowledge is often application dependent and, for the same application, camera dependent, so it should never be embedded into the algorithms. We have decided to use 3D descriptions of the observed empty scenes, as well as predefined scenarios, as a priori knowledge available to the system. For example, the predefined scenarios change when the application changes, but thanks to this separation we can reuse the same algorithm for a different system without modifying it. The a priori knowledge is composed of two different types of information: **image acquisition related** (e.g. camera calibration parameters) and **models** (e.g. models of expected physical objects evolving in the scene). If the use of a priori knowledge enables to better and more efficiently solve the interpretation problems, only the separation between knowledge and algorithms allows algorithms to be independent of the application and to be reusable in other situations. All a priori knowledge in our platform is stored in specific configuration files independently of the code. For example, all cameras observing the same scene are processed by computers having the same scene description file. Also, applying an activity monitoring system to a new scene requires only to change the context file. We have also proposed an adapted formalism to describe each type of knowledge. For example, the scenario models are described using a special declarative user-oriented language. Because of this knowledge organization we have managed to separate knowledge from the algorithms.

3. **Automatic evaluation** allows to evaluate the results of the different activity monitoring systems built from the platform after the integration of a new algorithm or the modification of an existing one. When addressing a new application, it is normal to face with new problems which require to handle new situations and to use more a priori knowledge (or to use more precisely the already existing knowledge). The development and the insertion into the platform of such new algorithms is made possible by the modularity and the flexibility of the architecture. Thus, the difficulty is to insure that the new algorithm keeps the quality of the results previously obtained by the activity monitoring system dedicated to other applications. A solution can be the development of an automatic evaluation framework based on ground truth data, which is able to evaluate the performances of a set of activity monitoring systems on a wide set of predefined sequences. Thanks to that, it is possible to evaluate the impact of a new algorithm on the platform, insuring that it globally increases the quality of the results. Moreover, a framework of this type allows to apply statistical learning methods for parameters tuning, necessary to find the best parameter set for a given application. The framework allows to apply a statistical learning technique for parameter tuning. For example, the algorithm can be run with a modified set of parameters for all ground-truth sequences. If the results improve, the parameters are validated, if not they are modified using an algorithm that explores heuristically the N-dimensional space of parameters (N being the number of parameters).

When a new algorithm is added to the platform, the set of ground-truth-ed sequences are run automatically by each activity monitoring system. The evaluation results are compared with those obtained before the integration of the new algorithm. In case of lower quality results, a first possibility is to recompute the set of parameters separately for each activity monitoring system. If the improved set of parameters still gives worse results than the one obtained before the introduction of the new algorithm, then this algorithm is said to be not generic enough to be used for all applications. The next step is to understand precisely why the new algorithm fails and under which hypotheses it can be used.

Finally, a development methodology to fulfill the last two properties consists in the **interaction between end-users and developers** (end users are for example metro security or bank agency operators). This interaction is useful because end-users provide the a priori knowledge (e.g. predefined scenario models) used by the system and the scenario-level ground-truth used to perform the automatic evaluation. The end-users a priori knowledge is particularly useful in three tasks:

1. End-users, helped by system developers, can find out which are the interesting activities to monitor and how to describe them precisely. The importance of this approach is to avoid to have a system which does not meet user needs.
2. it is often necessary to ask professional actors to act a set of scenes showing either normal activities and the activities to monitor. These video sequences are essential during the development of the system to tune and to test algorithms. Actors are needed because there are often too few recorded sequences showing these activities. Only end users can explain to actors how to act in a realistic way, which are the activities to monitor and how to describe them precisely.
3. end-users can perform, at the end of the development, an assessment of the system, measuring its efficiency and evaluating its utility.

In our case, we have been working closely with end-users of different application domains. We have build from our platform three activity monitoring systems for different applications which have been validated by end-users: an activity monitoring system in metro stations, a bank agency monitoring system and a lock chamber access control system for building security. These applications present some characteristics which make them interesting for research purposes: the observed scenes vary from large open spaces (like metro halls) to small and closed spaces (corridors and lock chambers); cameras can have both non overlapping (like in metro stations and lock chambers systems) and overlapping fields of view (metro stations and bank agencies); humans can interact with the equipment (like ticket vending machines or access control barriers, bank safes and lock chambers doors) either in simple ways (*open/close*) or in more complex ones (as the interaction occurring during *vandalism against equipment* or *jumping over the barrier* scenarios). All these AMS have been validated and an end-user assessment has been done or it is scheduled before the end of the year 2004.

We are currently building from our platform three other applications. All these applications are illustrated on figure 13,figure 14 and figure 15. A first application is apron monitoring on an airport where vehicles from various types are evolving in a cluttered scene. The dedicated system has been able to successfully detect at the same time vehicles and people getting in and out on several videos. A second application consists of detecting abnormal behaviors inside moving trains. The dedicated system is able to handle situations in which people are partially occluded by the train equipment like dividing walls and seats. A third application is traffic monitoring on highway; the dedicated system has been built in few weeks to show the adaptability of the platform. These systems are currently under development and validation and end-user assessment will be done in the near future.

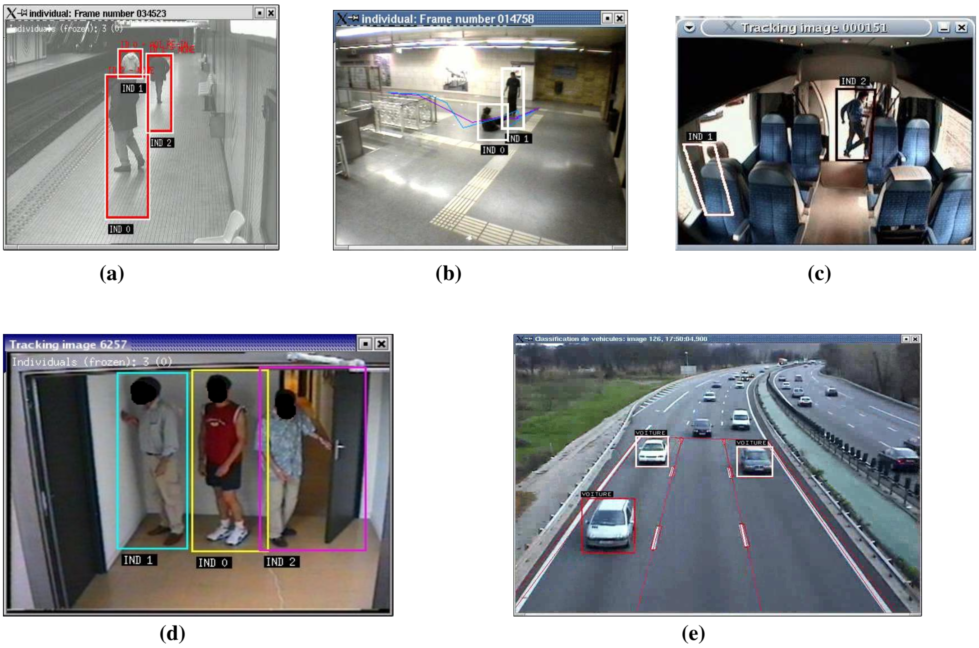


Figure 13. Image (a) illustrates a metro monitoring application system running on black and white cameras of the YZER station in Brussels. Image (b) illustrates the same system but analyzing images from a color surveillance camera of the SAGRADA FAMILIA station in Barcelona. Image (c) illustrates a system for unruly behaviors detection on trains. Image (d) illustrates a single-camera system for a lock chamber access control application for building entrances. Finally, image (e) illustrates an highway traffic monitoring application.

Figure 14. Images (a) and (b) illustrate an application for aprons monitoring on airports; this system has to combine a total of 8 cameras with overlapped fields of view.

(a) **(b)**

Figure 15. Images (a) and (b), taken with 2 synchronized cameras with overlapping field of view working in a cooperative way, illustrate a bank agency monitoring system detecting an abnormal "bank attack" scenario.

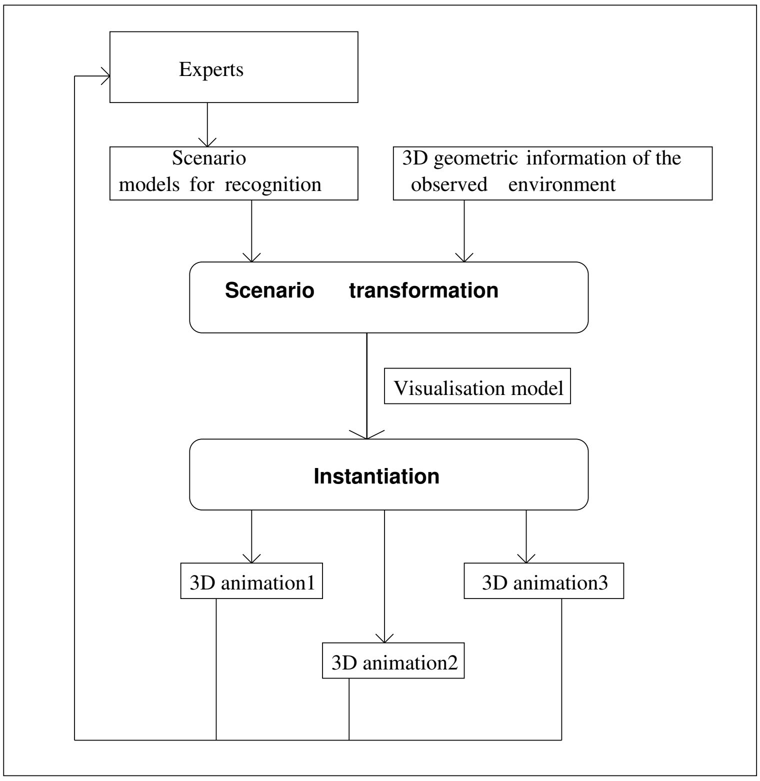


Figure 16. Generation of 3D animations: process overview.

6.3.9. User-centered tools for scenario modeling

Participants: Jihène Bannour, Benoît Georis, Le Son Pham, Nicolas Maillot, François Brémond, Monique Thonnat.

Generation of 3D Animations from Scenario Models

One research axis of the ORION team is automatic interpretation of video sequences using a knowledge-based approach. This knowledge can be classified in two main categories: knowledge on programs (e.g., how to use them, how to assess their performances) and knowledge on the application domain. This a priori knowledge depends on the application domain and is usually given by experts of the application domain (e.g., for a bank monitoring application, the expert can be a bank agent or a video-surveillance human operator). We have developed a tool for the generation of 3D animations from scenario models. This work is intended to ease the knowledge acquisition process and takes place in the context of user-centered interface.

The goal is to provide the expert with generated 3D animations of a scenario model he/she has described in a declarative language developed in ORION. This is of prime importance since the knowledge acquisition process is usually a long and iterative process: the expert is able to refine his/her models once he/she sees the system running on a video sequence illustrating the scenario. By giving him/her several generated 3D animations corresponding to several instances of the same scenario model, the expert will be able to verify directly his/her definition when writing the scenario model, thus making the knowledge acquisition process more efficient.

The overall schema of the generation process is presented in figure 16.

It consists of two main stages: transformation and instantiation. The transformation stage takes as input the scenario model of the expert and a 3D description of the environment (e.g., location of walls and doors) and outputs what we have called a visualization model. The visualization model contains:

- a linear system related to time instants,
- a linear system related to positions,
- trajectories,
- 3D geometry information of the observed environment,
- rendering parameters such as contrast, illumination level, position and number of lights,
- person's speed.

The instantiation stage is responsible for assigning values to each element contained in the visualization model. The variability we have in choosing values for those variables enables us to generate several instances of the same scenario model.

We have experienced this tool on 2 applications: in a bank monitoring application with a bank robbery scenario and a train monitoring with a pickpocket scenario. An expert feedback on the pickpocket scenario model has led to the addition of a speed parameter in the scenario model. The expert has noticed when seeing 3D animations generated from the model which he/she has defined that the speed of the robber was too slow and thus too far from reality. First results are promising and experts who have already used this tool have been able to refine their scenario model definition.

This work has been presented in [26].

Knowledge acquisition interface based on video event ontology

The scenario description language (SDL) used in the ORION team is not end-user oriented. To cope with this issue, knowledge engineering techniques are of interest. From April to September 2004, we have worked on a knowledge acquisition interface [38]. This interface aims at guiding the experts in the definition of new video events. The proposed knowledge acquisition process is based on a video event ontology which has been implemented in the standard language OWL. The knowledge acquisition tool allows the expert to perform the following tasks :

- defining physical objects involved in a video event,

- definining components of a video event,
- definining constraints of video event.

Each task is driven by the the video event ontology. Future works include an integration of the knowledge acquisition interface with the 3D visualization tool proposed by J.Bannour.

6.4. Cognitive Vision Platform

Keywords: *classification*, *cognitive vision*, *environment*, *image formation*, *learning*, *plant diseases*, *pollen*.

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

This year, research activites on object recognition have been continued. A cognitive vision platform is now partially operational. The platform is based on reasoning, learning and image processing mechanisms. The platform is used for the detection of plant diseases and for image indexing purposes. Ontology based knowledge acquisition for object description are still studied. A PhD thesis started in September 2004 on "Cognitive Vision : Supervised Learning for Image Segmentation".

6.4.1. Introduction

We study the problem of semantic image interpretation. We want to design a generic and reusable cognitive vision platform dedicated to object recognition and 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 users and 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. 17):

1. High-level semantic interpretation;
2. Mapping between high level representations of physical objects and image numerical data;
3. 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 :

[1] The European research Network for Cognitive Computer Vision Systems,

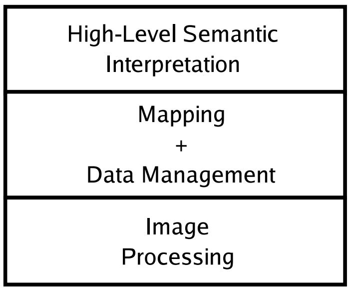


Figure 17. Image interpretation as a three-stage process.

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

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 new functionalities can be added easily. Current implementation is based on the development platform LAMA.

6.4.2. 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 (e.g. 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), context concepts (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 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. This new approach for designing knowledge bases for vision systems is described in [24].

A part of the knowledge acquisition process consists of manual segmentation of the objects of interest in image samples. We are currently experimenting unsupervised learning techniques (i.e. k-means and Kohonen maps) to ease this task. Instead of manually segmenting one region of interest per image, our goal is to guide the expert by automatically gathering sets of image data (e.g. regions) which have similar properties (i.e similar color, similar shape, similar texture). Once a set is obtained, the expert only has to semantically label the set by adequate visual concepts provided by the ontology.

Learning for visual concept detection

One difficult task when using knowledge based approaches is the definition of inferential knowledge. During 2004, our efforts have been focused on the use of machine learning techniques to facilitate the mapping between high-level knowledge and image data. In [30], it is explained how machine learning techniques (e.g. support vector machines, multi layer perceptions) are used to map image data to visual concepts used in the description of domain classes.

In order to obtain an operational learning module, a learning phase is needed. This supervised learning phase consists of training a set visual concept detectors. To achieve this task, manually segmented image samples provided during knowledge acquisition phase are used.

Once learning phase is over, trained visual detectors are used to perform automatic image interpretation. They are involved for mapping purposes. Visual concept detectors detect visual concepts in the image to interpret. A probabilistic combination of detected visual concepts gives a confidence value about the mapping between high-level hypotheses and image data.

Learning for automatic parameterization of segmentation algorithms

During his master thesis, V.Martin proposed a method which uses supervised learning techniques and a priori knowledge to perform better parameterization of segmentation algorithms. This approach is composed of two phases : (1) a supervised learning phase and (2) an automatic segmentation phase. Phase 1 uses a training set composed of representative images of the given domain of interest. These images have to be manually segmented and symbolically described. The symbolic description is based on a priori knowledge of the domain of interest. Such a symbolic description can be used to guide segmentation. The learning phase first consists in inferring adapted segmentation algorithms from symbolic description of training images. Then optimization of selected algorithms (e.g. finding best thresholds) is performed thanks to manually segmented images. Next, statistical features (e.g. entropy, histogram analysis) are extracted from images. These features show context variations. Finally, matching between algorithms optimal parameters and different image context features distribution is learned by neural networks. Previously trained neural networks are used during automatic segmentation phase to perform automatic parameterization of algorithms selected thanks to a priori knowledge. First experimental results of automatic phase proves that learning techniques can improve image segmentation. By tuning algorithms parameters within image features related to context information, we obtain more accurate segmentation results than non-adaptive methods. More details of our approach can be found in [37].

6.4.3. *Semantic Object Recognition*

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

As explained in [7], the proposed architecture for our cognitive vision platform, is a distributed architecture based on three highly specialized Knowledge Based System (KBS) see figure 18. Each KBS is specialized for the corresponding sub-task of image interpretation. We choose KBS techniques to emulate the strategy of the expert of the corresponding task. For each involved KBS, we have:

- formalized the different types of knowledge adapted to the different types of expertise,
- specified dedicated engine with adapted reasoning,
- involved goal-directed behavior for each sub-task through the generation of specific requests,
- made explicit the context useful for each sub-task.

High-level semantic interpretation

High-level semantic interpretation process uses knowledge acquired during ontology based knowledge acquisition phase: the domain specific knowledge. 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,

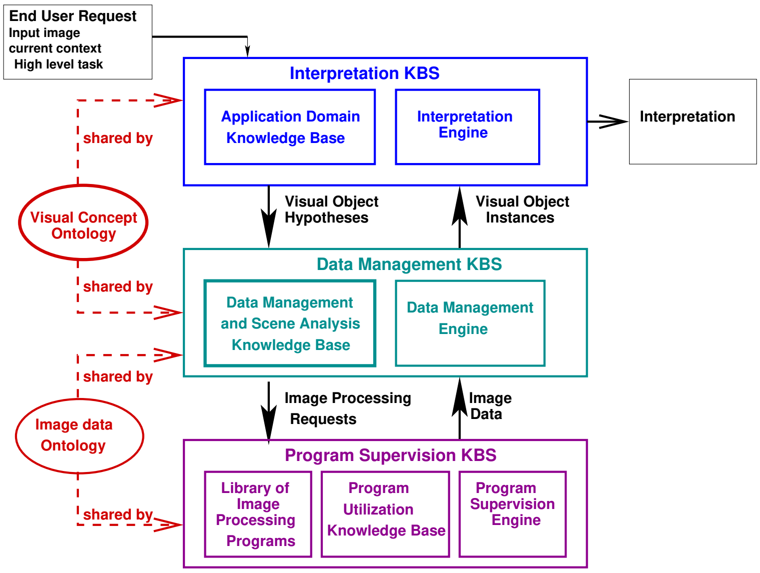


Figure 18. Global minimal architecture of our cognitive vision platform

- 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.

Mapping between high level representations and image data

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 reasonings as uncertainty reasoning, spatial reasoning and data management reasoning (grouping or splitting of data, ...). We have designed a dedicated KBS for this task : the 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 organized in generic image data concepts (**image data primitives**), generic image features (**image data descriptors**) and **spatial relations** divided into topological, distance and orientations relations. It also contains inferential knowledge : **object extraction criteria** to build image processing request according to the object hypothesis in terms of visual concepts, **scene analysis 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*, *hypotheses*, *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 a good use of program supervision approach to manage different contexts and to process images in an intelligent and adaptable way. As explaines 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 and low level grouping. The generic image processing requests are generated by the data management KBS. They are linked to 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.

6.4.4. Applications

Detection of plant diseases

In cooperation with INRA (URIH de Sophia Antipolis), the early detection of plant diseases, in particular rose diseases is used to evaluate and validate our platform ([23], [27]). 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. To test the cognitive vision platform, a domain knowledge base describing disease symptoms on rose leaves has been written. It contains about 20 domain classes.

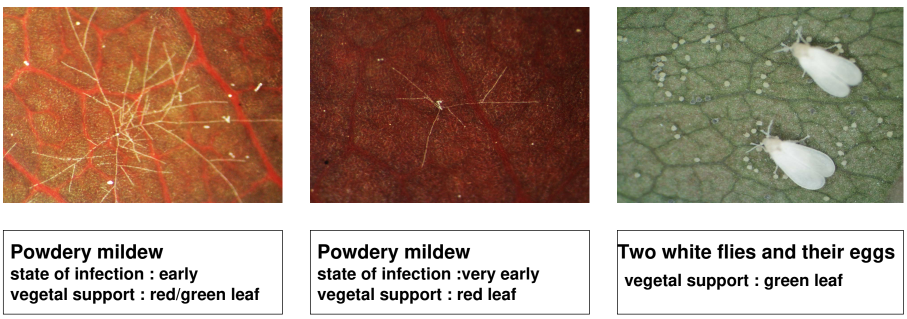


Figure 19. Examples of analyzed images. The two images of powdery mildew emphasize the complexity and variability of object appearance and of contexts : we need domain knowledge and intelligent management of image processing programs. The image of insects is a case of a multiple object scene : we need spatial knowledge and spatial reasoning.

Image Indexing and Retrieval

The cognitive vision platform is currently used to perform image indexing. As described in [30], 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. Image processing tasks (i.e. segmentation, feature extraction) are needed to achieve 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.

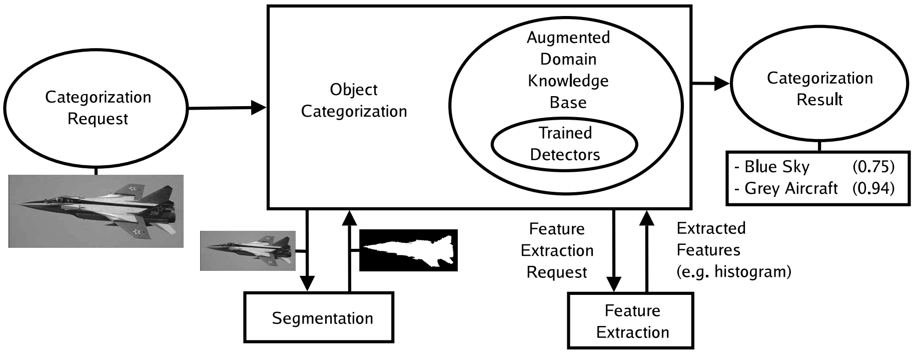


Figure 20. Object categorization for image indexing

We have obtained a recall rate of 19% for a precision of 73%. 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.

7. Contracts and Grants with Industry

7.1. Industrial Contracts

Participants: Alberto Avanzi, François Brémond, Binh Bui, Frédéric Cupillard, Magali Mazière, Monique Thonnat, Christophe Tornieri, Florent Fusier, Gabriele Davini.

In 2004, ORION team has been involved in 5 industrial projects : RATP project on subway user classification, CASSIOPEE project on bank agency visual surveillance and two projects, SAMSIT and RASV on train visual surveillance.

- PFC : This project with RATP (Paris) and Timeat (Rennes)ended in June, 2004 and provided 1.6MF financial resources to ORION. The aim of this project was to classify in real-time different types of subway users;
- CASSIOPEE : This project with Crédit Agricole, Ciel and Eurotélis has a duration of 3 years and will provide 450KEuros financial resources to ORION. The aim of this project is to develop and test a visual-surveillance platform allowing automatic detection of predefined scenarios in a bank agency environment;
- SAMSIT : SAMSIT is a European project in collaboration with ALSTOM, CEA, SNCF, INRETS. It has begun in January, 2004 and will last 2 years. The aim of this project is 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;

- RASV : This project with SNCF has begun in september 2003 for a duration of 18 months (funding for Orion). The aim of this project is to automatically detect human behaviors in trains. During this project ORION team will develop techniques to recognize human behaviors and scenarios of interest;
- 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;
- VIDEA : This project with Vigitec began in November 2003 and has a 2 years duration. The aim of this project is to transfer a part of the video-surveillance technology of the ORION team into industrial products. During this project, the ORION team will develop and transfer to a video-surveillance company two applications enabling the recognition of specific human behaviors;
- KINOMAÏ : This project has begun in December 2003 for a duration of four months in collaboration with Kinomaï team. The aim of this project is to develop an object recognition platform for image indexing and retrievial purposes.

8. Other Grants and Activities

8.1. European projects

ORION team has been involved this year in two european projects on image interpretation : ECVision European Research Network (IST-type) and a project on airport surveillance AVITRACK.

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) . It has begun in February 2004 and will last 2 years. 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.

8.2. International Collaborations

Orion is involved in two local program 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 Collaborations

Orion Team has three national cooperations each involving a PhD thesis.

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 (Celine Hudelot) is on-going on cognitive vision. 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 and 6.4.4).

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-user of the MAGE simulation code to achieve a river model calibration activity. A joint PhD thesis (J-P. Vidal) is on-going.

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 6.3.2).

9. Dissemination

9.1. Scientific Community

- 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 Artificial Intelligence Journal, PATREC, Image and Vision Computing Journal, IEEE Systems, Man and Cybernetics, IEEE Transactions - Pattern Analysis and Machine Intelligence PAMI, CVIU (Journal of Computer Vision and Image Understanding),EURASIP Journal on Applied Signal Processing, IJCV (International Journal on Computer Vision),IEE Proceedings Vision, Image and Signal Processing;
- M. Thonnat is a reviewer for the conferences RFIA04, KR04, ECAI04, MDDE 04, Event Mining2004;
- 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 area chair for the conference CVPR04;

- M. Thonnat is an expert for ACI Masse de donnees program from the French Ministry of Research;
- M. Thonnat is member of the steering committee of the Technovision program from theFrench Ministry of Research and the French Ministry of Defence;
- M. Thonnat is an expert for IFREMER PhD thesis program;
- F. Brémond is a reviewer for the BMVC'04, CVPR'04 and ECAI'04 conferences;
- F. Brémond is a reviewer for IEEE Transactions Neural Networks, PAAJ and EURASIP-JASP;
- S. Moisan is a member of the program commitee of the IC conference about knowledge software;
- S. Moisan is a member of the 27^{e} department of specialist committee at UNSA (Nice Sophia Antipolis University);
- N. Maillot is a reviewer for Image and Vision Computing Journal (Elsevier).

9.2. Teaching

- Orion is a hosting team for the DEA of computer Science of UNSA;
- Teaching at Summer Schools in Cognitive Vision organised by the network of excellence ECVision at Bonn in august (9h M. Thonnat);
- Teaching at DESS of Computer Science at Essi (UNSA), Object-oriented Analysis and Conception (15h S. Moisan);
- Teaching at DEA of Astronomy, image and gravitation (UNSA) classification lectures (9h M. Thonnat and 3h F. Brémond);
- Conference at EURECOM on Video Understanding (2h 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 ar DESS 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);

9.3. Thesis

9.3.1. Thesis in progress

- Bernard Boulay : Reconnaissance de postures pour l'interprétation d'activités humaines, Nice Sophia Antipolis University;
- Benoit Georis : Knowledge-based reconfigurable tracker, Louvain Catholic University;
- Céline Hudelot : Vers une plate forme de vision cognitive pour l'interpretation sémantique d'images : application à la reconnaissance d'organismes végétaux, 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;
- Nicolas Maillot : Système cognitif d'interprétation d'images pour la reconnaissance d'images d'objets 3D, Nice Sophia Antipolis University;
- Vincent Martin : Vision cognitive: apprentissage supervisé pour la segmentation d'image, Nice-Sophia Antipolis University;
- Jean-Philippe Vidal : Validation opérationnelle en hydraulique fluviale – Approche par un système à base de connaissances, Institut National Polytechnique. de Toulouse.

9.3.2. Thesis defended

- Thinh van Vu : Visualisation de comportements humains pour l'interprétation de séquences video, Nice Sophia Antipolis University (thesis defended in october 2004).

9.4. Participation to Conferences, Seminars, and Invitations

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

- RFIA 2004 (14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle) (Toulouse, France);
- SEFM 2004 (Software Engineering and Formal Methods) (Beijing, China);
- IDSS (Intelligent Distributed Surveillance System) (London, UK);
- 15ième Journées Francophones d'Ingienerie des Connaissances (Lyon, France);
- ICTAI (16th IEEE International Conference on Tools with Artificial Intelligence) (Boca Raton,USA);
- VIIP (Visualization, Imaging and Image Processing) (Marbella, Spain);
- ICNSC (IEEE International Conference on Networking, Sensing and Control) (Tapei, Taiwan);
- Summer School on Cognitive Vision (Bonn, Germany).

10. Bibliography

Major publications by the team in recent years

[1] B. BOULAY, F. BREMOND, M. THONNAT. *Human Posture Recognition in Video Sequence*, in "Proceedings Joint IEEE International Workshop on VS-PETS, Visual Surveillance and Performance Evaluation of Tracking and Surveillance", OCT 11-12 2003, p. 23-29.

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

[3] N. CHLEQ, F. 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.

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

[5] N. DEY, A. BOUCHER, M. THONNAT. *Un modèle pour la formation de l'image d'un objet microscopique 3D translucide*, in "Actes du Troisième colloque francophone Traitements & Analyse d'Informations, Méthodes & Applications TAIMA'2003", vol. 1, September 2003, p. 219-224.

[6] B. GEORIS, F. BREMOND, M. THONNAT, B. MACQ. *Use of an Evaluation and Diagnosis Method to Improve Tracking Performances*, in "Proceedings of VIIP'03- 3rd IASTED International Conference on Visualization, Imaging and Image Processing, Benalmadera, SPAIN", SEP 8-10 2003.

[7] C. HUDELOT, M. THONNAT. *A cognitive Vision Platform for Automatic Recognition of Natural Complex Objects*, in "International Conference on Tools with Artificial Intelligence, ICTAI, Sacramento", IEEE (editor)., 2003.

[8] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Computer Vision Systems, Third International Conference, ICVS", J. L. CROWLEY, J. H. PIATER, M. VINCZE, L. PALETTA (editors)., Lecture Notes in Computer Science, vol. 2626, Springer, 2003.

[9] N. MOENNE-LOCCOZ, F. BRÉMOND, M. THONNAT. *Recurrent bayesian network for the recognation of human behaviors from video*, in "Third International Conference On Computer Vision Systems (ICVS 2003), Graz, Austria", vol. LNCS 2626, Springer, 2003, p. 44-53.

[10] S. MOISAN. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*, Habilitation à diriger les recherches, université de Nice, avril 1998.

[11] 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° 4, November 2001, p. 501-507.

[12] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Behavioral Substitutability in Component Frameworks: a For-*

mal Approach, in "Specification and Verification of Component-Based Systems (SAVCBS'2003) Workshop at ESEC/FSE 2003, Helsinki, Finland", September 2003, p. 22–28.

[13] A. RESSOUCHE, V. ROY, J. TIGLI, D. CHEUNG. *SAS architecture: Verification Oriented Formal Modeling of Concrete Critical Systems*, in "2003 IEEE International Conference on Systems, Man and Cybernetics, SMC'03, Washington, USA", October 2003.

[14] 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° 5, 2000.

[15] 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.

[16] M. THONNAT, N. ROTA. *Image Understanding for Visual Surveillance Applications*, in "Third International Workshop on Cooperative Distributed Vision, Kyoto, Japan", November 1999.

[17] 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, octobre 2003.

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

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

[20] V.-T. VU, F. BRÉMOND, M. THONNAT. *Video surveillance: human behaviour representation and on-line recognition*, in "The Sixth International Conference onKnowledge-Based Intelligent Information & Engineering Systems (KES'02), Crema, Italy", 2002.

[21] 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.

[22] V.-T. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Recognition Algorithm for Temporal Scenarios Based on Pre-compiled Scenario Models*, in "The 3rd International Conference on Vision System (ICVS'03), Graz, Austria", 2003.

Articles in referred journals and book chapters

[23] 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.

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

[25] J.-P. VIDAL, S. MOISAN, J.-B. FAURE, D. DARTUS. *Towards a Reasoned 1-D River Model Calibration.*, in "Journal of Hydroinformatics", accepted for 2005, 2004.

Publications in Conferences and Workshops

[26] J. BANNOUR, B. GEORIS, F. BREMOND, M. THONNAT. *Generation of 3D Animations from Scenario Models*, in "Proceedings of VIIP'04- 4th IASTED International Conference on Visualization, Imaging and Image Processing", September 6-8th, 2004.

[27] P. BOISSARD, G. PÉREZ, P. BÉAREZ, P. PYRRHA, C. HUDELOT, J. LÉVY. *Maladies et ravageurs des cultures ornementales; Raisonner la protection des plantes*, in "Journées techniques ASTREDHOR, Rennes, France", janvier 2004.

[28] F. CUPILLARD, A. A. F. BRÉMOND, M. THONNAT. *Video Understanding For Metro Surveillance*, in "IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems, Tapei(Taiwan)", March 2004.

[29] B. GEORIS, M. MAZIERE, F. BREMOND, M. THONNAT. *A Video Interpretation Platform Applied To Bank Agency Monitoring*, in "Proceedings of IDSS'04 - 2nd Workshop on Intelligent Distributed Surveillance Systems", February 23th, 2004.

[30] 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.

[31] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Towards Formalizing Behavorial Substitutability in Component Frameworks*, in "2nd International Confernece on Software Engeniering and Formal Methods, Beijing, China", IEEE Computer Society Press, 2004, 122,131.

[32] J.-P. VIDAL, J.-B. FAURE, S. MOISAN, D. DARTUS. *Decision Support System for Calibration of 1-D River Models.*, in "6th International Conference on Hydroinformatics, Singapore", S. Y. LIONG, K.-K. PHOON, V. BABOVIC (editors)., vol. 2, World Scientific Publishing, June 2004, p. 1067–1074.

[33] J.-P. VIDAL, S. MOISAN. *Opérationalisation de connaissances pour le calage de modèles numériques – Application en hydraulique fluviale.*, in "IC'2004, 15èmes Journées Francophones d'Ingénierie des Connaissances, Lyon, France", Presses Universitaires de Grenoble, 2004, p. 261–272.

Internal Reports

[34] F. BRÉMOND, N. MAILLOT, M. THONNAT, T. VU. *Ontologies For Video Event*, Technical report, n° 5189, INRIA Sophia Antipolis, May 2004, http://www.inria.fr/rrrt/rr-5189.html.

Miscellaneous

[35] B. BOULAY, M. THONNAT, F. BREMOND. *PS26-27 : Environnement Intelligent*, Rapport d'avancement, March 2004.

[36] M. DORTEL, J. GALIZZI. *Création d'une interface graphique de commande pour un ordonnanceur de prises de vues sur télescope léger*, Rapport de projet dernière année d'école d'ingénieur, ESSI, 2004.

[37] V. MARTIN. *Aprentissage Supervisé pour la Segmentation d'Images*, rapport de DEA d'Astronomie, Université de Nice Sophia Antipolis, Juin 2004.

[38] L. S. PHAM. *Acquisition de Connsaissances pour la Reconnaissances de Comportements Humains*, rapport de DEPA, IFI, Vietnam, September 2004.

Bibliography in notes

[39] 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° 9, 1995, p. 1150–1158.

[40] 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° 2, April 1996, p. 113–123.

[41] 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.

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

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

[44] 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.

[45] 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.

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

[47] 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° 5, November 1992, p. 373–387.

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

[49] 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.

[50] 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.