INRIA

Project-Team PRIMA

Perception, recognition and integration for interactive environments

Grenoble - Rhône-Alpes

THEME COG

Activity Report

2008

Table of contents

1. Team

Research Scientist

Frédéric DEVERNAY [CR1 INRIA]

Faculty Member

James CROWLEY [Professeur CE INPG, HdR]

Augustin LUX [Professeur INPG, HdR]

Patrick REIGNIER [MdC Université Joseph Fourier]

Dominique VAUFREYDAZ [MdC Université Pierre Mendes France]

Alexandre DEMEURE [MdC Université Joseph Fourier]

External Collaborator

Julien LETESSIER [HiLabs]

Martin GUILLON [Binocle SARL]

Sergi PUJADES [Binocle SARL]

Technical Staff

Nicolas GOURIER [Engineer CDD INRIA]

Matthieu LANGET [Engineer CDD INRIA]

Stan BORKOWSKI [Engineer CDD INRIA]

Jérôme MAISONNASSE [Engineer CDD INRIA]

PhD Student

Olivier BERTRAND [Doctoral Student (ENS Cachan)]

Sofia ZAIDENBERG [Doctoral Student (BDI CNRS), ATER UPMF]

Suphot CHUNWIPHOT [Doctoral Student (Thailand)]

Rémi EMONET [Doctoral Student (Bourse MENSR), ATER ENSIMAG]

Han YUE [Doctoral Student (Bourse MENSR)]

Jean-Pascal MERCIER [Doctoral Student (Bourse MENSR)]

Rémi BARRAQUAND [Doctoral Student (Bourse MENSR)]

Antoine MELER [Doctoral Student (Bourse MENSR)]

Frédéric HUGUET [Doctoral Student (Bourse MENSR)]

John Alexander RUIZ HERNANDEZ [Doctoral Student (Bourse INRIA)]

Administrative Assistant

Annie SIMON [Secretary (SAR) Inria]

2. Overall Objectives

2.1. Perception, Recognition and Integration for Interactive Environments.

Keywords: *Computer Vision*, *Interactive Environments*, *Machine Perception*, *Man-Machine Interaction*, *Perceptual User Interfaces*.

The objective of Project PRIMA is to develop the scientific and technological foundations for human environments that are capable of perceiving, acting, communicating, and interacting with people in order to provide services. The construction of such environments offers a rich set of problems related to interpretation of sensor information, learning, machine understanding, dynamic composition of components and man-machine interaction. Our goal is make progress on the theoretical foundations for perception and cognition, as well as to develop new forms of man machine interaction, by using interactive environments as a source of example problems.

An environment is a connected volume of space. An environment is said to be "perceptive" when it is capable of recognizing and describing things, people and activities within its volume. Simple forms of applications-specific perception may be constructed using a single sensor. However, to be general purpose and robust, perception must integrate information from multiple sensors and multiple modalities. Project PRIMA creates and develops machine perception techniques fusing computer vision, acoustic perception, range sensing and mechanical sensors to enable environments to perceive and understand humans and human activities.

An environment is said to be "active" when it is capable of changing its internal state. Common forms of state change include regulating ambient temperature, acoustic level and illumination. More innovative forms include context-aware presentation of information and communications, as well as services for cleaning, materials organisation and logistics. The use of multiple display surfaces coupled with location awareness offers the possibility of automatically adapting information display to fit the current activity of groups. The use of activity recognition and acoustic topic spotting offers the possibility to record a log of human to human interaction, as well as to provide relevant information without disruption. The use of steerable video projectors (with integrated visual sensing) offers the possibilities of using any surface for presentation, interaction and communicartion.

An environment may be considered as "interactive" when it is capable interacting to humans using tightly coupled perception and action. Simple forms of interaction may be based on observing the manipulation of physical objects, or on visual sensing of fingers, hands or arms. Richer forms of interaction require perception and understanding of human activity and context. PRIMA has developed a novel theory for situation modeling for machine understanding of human activity, based on techniques used in Cognitive Psychology [37]. PRIMA explores multiple forms of interaction, including projected interaction widgets, observation of manipulation of objects, fusion of acoustic and visual information, and systems that model interaction context in order to predict appropriate action and services by the environment.

For the design and integration of systems for perception of humans and their actions, PRIMA has developed:

- A theoretical foundation for machine understanding of human activity using situation models
- Robust, view invariant methods for computer vision systems using local appearance,
- A software architecture model for reactive control of multi-modal perceptual systems.

The experiments in project PRIMA are oriented towards developing interactive services for smart environments. Application domains include health and activity monitoring services for assisted living, context aware video recording for lectures, meetings and collaborative work, context aware services for commercial environments new forms of man-machine interaction based on perception and new forms of interactive services for education, research and entertainment. Creating interactive services requires scientific progress on a number of fundamental problems, including

- Component-based software architectures for multi-modal perception and action,
- Service-oriented software architectures for smart environments,
- Situation models for observing and understanding human to human interaction,
- Robust, view-invariant image description for embedded services based on computer vision,
- New forms of multi-modal human-computer interaction.

2.2. Highlights

- In March 2008, PRIMA officially began participation in OSEO project MinImage. Project MinIMage, directed by ST Micro-eloctronics, is a 5 year program to create ultra-low cost high quality miniature image acquisition and processing systems for cell phones, video-phones, computers and mobile devices. PRIMA is providing fast view invariant detection, tracking and recognition methods for novel mobile applications such as video communications, and interactive services.

- In October 2008 members of PRIMA officially founded HILabs (Human Interaction Laboratories). HILabs will commercialize large-scale interactive public information displays based on a patented technology for multi-touch interactive surfaces developed by members of the PRIMA project-group.

- Doctoral Thesis presented by members of PRIMA in 2008

    – Christophe Braillon, (Avec C. Laugier), Détection d'obstacles par fusion de flux optique et stéréovision dans des grilles d'occupation, Thèse doctorale de l'I.N.P.G, 15 january 2008.

    – Suphot Chunwiphat, A formal framework for representation and recognition of human activities, Thèse doctorale de l'INPG, 20 march 2008

3. Scientific Foundations

3.1. Context aware interactive environments

Keywords: *Context Aware Environments*, *Situation Modeling*, *Smart Environments*.

3.1.1. Summary

Interactive environments have the potential to provide many new services for communications and access to information. A major barrier to providing such services is the problem of unwanted disruption of human activity. Information and communication technologies are autistic. They have no sense of the social roles played by interacting humans, no abilities to predict appropriate or inappropriate service actions, and no sensitivity to the disruption to activity caused by inappropriate service behavior. Disruption renders information and communications services impractical for many applications.

Over the last few years, the PRIMA group has pioneered the use of context aware observation of human activity in order to provide non-disruptive services. In particular, we have developed a conceptual framework for observing and modeling human activity, including human-to-human interaction, in terms of situations. A situation model acts as a non-linear script for interpreting the current actions of humans, and predicting the corresponding appropriate and inappropriate actions for services. This framework organizes the observation of interaction using a hierarchy of concepts: scenario, situation, role, action and entity.

Many human activities follow a loosely defined script in which individuals assume roles. Depending on the activity, actions and interaction may be more or less constrained and limited by implicit compliance with a shared script. Deviating from the script is considered impolite and can often provoke reprobation or even terminate the interaction. Some activities, such as class-room teaching, formal meetings, purchasing items in a shop, or dining at a restaurant, follow highly structured scripts that constrain individual actions to a highly predictable sequence. Other human activities occur in the absence of well-defined scripts, and are thus less predictable. We propose that when a stereotypical social script does exist, it can be used to structure observation and to guide the behavior of services that avoid disruption.

Encoding activity in situation models provides a formal representation for building systems that observe and understand human activity. Such models provide scripts of activities that tell a system what actions to expect from each individual and the appropriate behavior for the system.

No generic approach currently exists for domain independent recognition of situations from machine perception. Approaches such as logic programming, Bayesian reasoning, and fuzzy logic systems each have a number of domain dependent strengths and weakness. In the Prima project, we explore a Model Driven Engineering (MDE) approach that allows us to explicitly separate a context model and a program implementation. Such an approach allows to integrate different programming tools as required for an application, and to automate the process of transforming a context model into an interactive service.

Current technology allows us to handcraft real-time systems for a specific services. The current hard challenge is to create a technology to automatically learn and adapt situation models with minimal or no disruption of human activity. We are curently exploring Reinforcement Learning approachs to solve this problem

3.1.2. Detailed Description

An environment is a connected volume of space. An environment is said to be "interactive" when it is capable of perceiving, acting, and communicating with its occupants. The construction of such environments offers a rich set of problems related to interpretation of sensor information, learning, machine understanding and man-machine interaction. Our goal is make progress on a theoretical foundation for cognitive or "aware" systems by using interactive environments as a source of example problems, as well as to develop new forms of man machine interaction.

The experiments in project PRIMA are oriented towards context aware observation of human activity. Over the last few years, the group has developed a technology for describing activity in terms of a network of situations. Such networks provide scripts of activities that tell a system what actions to expect from each individual and the appropriate behavior for the system. Current technology allows us to handcraft real-time systems for a specific service. The current hard challenge is to create a technology for automatically learning and adapting situation models with minimal or no disruption of users.

We have developed situation models based on the notion of a script. A theatrical script provides more than dialog for actors. A script establishes abstract characters that provide actors with a space of activity for expression of emotion. It establishes a scene within which directors can layout a stage and place characters. Situation models are based on the same principle.

A script describes an activity in terms of a scene occupied by a set of actors and props. Each actor plays a role, thus defining a set of actions, including dialog, movement and emotional expressions. An audience understands the theatrical play by recognizing the roles played by characters. In a similar manner, a user service uses the situation model to understand the actions of users. However, a theatrical script is organised as a linear sequence of scenes, while human activity involves alternatives. In our approach, the situation model is not a linear sequence, but a network of possible situations, modeled as a directed graph.

Situation models are defined using roles and relations. A role is an abstract agent or object that enables an action or activity. Entities are bound to roles based on an acceptance test. This acceptance test can be seen as a form of discriminative recognition.

There is no generic algorithm capable of robustly recognizing situations from perceptual events coming from sensors. Various approaches have been explored and evaluated. Their performance is very problem and environment dependent. In order to be able to use several approaches inside the same application, it is necessary to clearly separate the specification of context (scenario) and the implementation of the program that recognizes it, using a Model Driven Engineering approach. The transformation between a specification and its implementation must be as automatic as possible. We have explored three implementation models :

- *Synchronized petri net.* The Petri Net structure implements the temporal constraints of the initial context model (Allen operators). The synchronisation controls the Petri Net evolution based on roles and relations perception. This approach has been used for the Context Aware Video Acquisition application (more details at the end of this section).

- *Fuzzy Petri Nets.* The Fuzzy Petri Net naturally expresses the smooth changes of activity states (situations) from one state to another with gradual and continuous membership function. Each fuzzy situation recognition is interpreted as a new proof of the recognition of the corresponding context. Proofs are then combined using fuzzy integrals. This approach has been used to label videos with a set of predefined scenarios (context).

- *Hidden Markov Model.* This probabilistic implementation of the situation model integrates uncertainty values that can both refer to confidence values for events and to a less rigid representation of situations and situations transitions. This approach has been used to detect interaction groups (in a group of meeting participants, who is interacting with whom and thus which interaction groups are formed)

Currently situation models are constructed by hand. Our current challenge is to provide a technology by which situation models may be adapted and extended by explicit and implicit interaction with the user. An important

aspect of taking services to the real world is an ability to adapt and extend service behaviour to accommodate individual preferences and interaction styles. Our approach is to adapt and extend an explicit model of user activity. While such adaptation requires feedback from users, it must avoid or at least minimize disruption. We are curently exploring reinforcement learning approaches to solve this problem.

With a reinforcement learning approach, the system is rewarded and punished by user reactions to system behaviors. A simplified stereotypic interaction model assures a initial behavior. This prototypical model is adapted to each particular user in a way that maximizes its satisfaction. To minimize distraction, we are using an indirect reinforcement learning approach, in whichser actions and consequences are logged, and this log is periodically used for off-line reinforcement learning to adapt and refine the context model.

Adaptations to the context model can result in changes in system behaviour. If unexpected, such changes may be disturbing for the end users. To keep user's confidence, the learned system must be able to explain its actions. We are currently exploring methods that would allow a system to explain its model of interaction. Such explanation is made possible by explicit describing context using situation models.

The PRIMA group has refined its approach to context aware observation in the development of a process for real time production of a synchronized audio-visual stream based using multiple cameras, microphones and other information sources to observe meetings and lectures. This "context aware video acquisition system" is an automatic recording system that encompasses the roles of both the camera-man and the director. The system determines the target for each camera, and selects the most appropriate camera and microphone to record the current activity at each instant of time. Determining the most appropriate camera and microphone requires a model of activities of the actors, and an understanding of the video composition rules. The model of the activities of the actors is provided by a "situation model" as described above.

Version 1.0 of the video acquisition system was used to record 8 three-hour lectures in Barcelona in July 2004. Since that time, successive versions of the system have been used for recording testimonial's at the FAME demo at the IST conference, at the Festival of Science in Grenoble in October 2004, and as part of the final integrated system for the national RNTL ContAct project. In addition to these public demonstrations, the system has been in frequent demand for recording local lectures and seminars. In most cases, these installations made use of a limited number of video sources, primarily switching between a lecturer, his slides and the audience based on speech activity and slide changes. Such actual use has allowed us to gradually improve system reliability. Version 2.0, released in 2005, incorporated a number of innovations, including 3D tracking of the lecturer and detection of face orientation and pointing gestures. This version has been used to record the InTech lecture series a the INRIA amphitheater. This system has been installed in a meeting room and in an amphitheater of the LIG laboratory and is currently undergoing real-world trials. This installation is a part of the LIG plateform demonstrating ambient informatics.

In collaboration with France Telecom, we have adapted this technology to observing social activity in domestic environments. Our goal is to demonstrate new forms of services for assisted living to provide non-intrusive access to care as well to enhance informal contact with friends and family.

3.2. Robust architectures for multi-modal perception

Keywords: *Autonomic Computing*, *MultiModal Perception*, *Process Architectures*, *Robust Perceptual Components*.

3.2.1. Summary

Machine perception is notoriously unreliable. Even in controlled laboratory conditions, programs for speech recognition or computer vision generally require supervision by highly trained engineers. Practical real-world use of machine perception requires fundamental progress in the way perceptual components are designed and implemented. A theoretical foundation for robust design can dramatically reduce the cost of implementing new services, both by reducing the cost of building components, and more importantly, by reducing the obscure, unpredictable behaviour that unreliable components can create in highly complex systems. To meet this challenge, we propose to adapt recent progress in autonomic computing to the problem of producing reliable, robust perceptual components.

Autonomic computing has emerged as an effort inspired by biological systems to render computing systems robust [35]. Such systems monitor their environment and internal state in order to adapt to changes in resource availability and service requirements. Monitoring can have a variety of forms and raises a spectrum of problems. An important form of monitoring relies on a description of the system architecture in terms of software components and their interconnection. Such a model provides the basis for collecting and integrating information from components about current reliability, in order to detect and respond to failure or degradation in a component or changes in resource availability (auto-configuration). However, automatic configuration, itself, imposes constraints on the way components are designed, as well as requirements on the design of the overall system [31].

Robust software design begins with the design of components. The PRIMA project has developed an autonomic software architecture as a foundation for robust perceptual components. This architecture allows experimental design with components exhibiting, auto-regulation, self-description, self-Monitoring and self-configuration and self-repair. Maintenance of such autonomic properties can result in additional computing overhead within components, but can pay back important dividends in system reliability.

3.2.2. *Detailed Description*

Components based programming makes it possible to design systems that can be dynamically reconfigured during run-time. Reconfiguration can be achieved by having each component provide a description of its parameters, input data and output data using a standardized XML schema. Such XML descriptions can be recorded in a component registry and used to adapt interfaces either manually or automatically. Such XML descriptions are an example of the principle of self-description that characterizes Autonomic Systems [38]. Other such principles are defined at the component level and the systems integration level. At the component level, in addition to self-description one finds techniques for "auto-initialization", "self-regulation", self-monitoring and "performance reporting". At the systems level, one finds methods for "self-configuration", self-repair, and system supervision.

An important form of monitoring relies on a description of the system architecture in terms of software components and their interconnections. Such a model provides the basis for collecting and integrating information from components about current reliability, in order to detect and respond to failure or degradation in a component or changes in resource availability (auto-configuration). However, automatic configuration, itself, imposes constraints on the way components are designed, as well as requirements on the design of the overall system [31], [38]. The PRIMA group has taken a leading role in introducing autonomic system approachs to programming perceptual systems.

Robust software design begins with the design of components. The PRIMA project has developed an autonomic software architecture as a foundation for robust perceptual components. This architecture allows experimental design with components exhibiting:

Auto-criticism: Every computational result produced by a component is accompanied by an estimate of its reliability.

Auto-regulation: The component regulates its internal parameters so as to satisfy a quality requirement such as reliability, precision, rapidity, or throughput.

Self-description: The component can provide a symbolic description of its own functionality, state, and parameters.

Self-Monitoring: the component can provide a report on its internal state in the form of a set of quality metrics such as throughput and load.

Self-configuration: The component reconfigures its own modules so as to respond to changes in the operating environment or quality requirements [47].

Self-repair: The component can react to errors detected by self-monitoring by changing operating parameters, switching among execution modules, or editing internal data structures.

Maintenance of such autonomic properties can result in additional computing overhead within components, but can pay back important dividends in system reliability.

The PRIMA software architecture for supervised autonomic perceptual components [27], [28], is shown in figure 1. In this design, perceptual components use a supervisory controller to dynamically configure, schedule and execute a set of modules in a cyclic detection and tracking process.
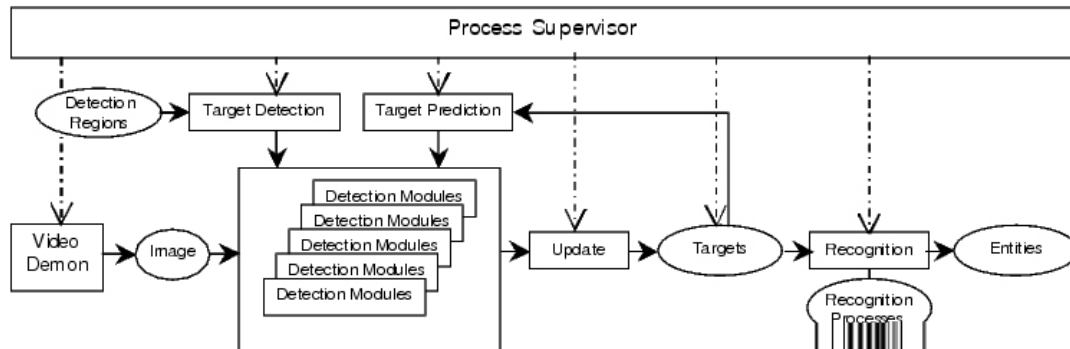

Figure 1. Architecture for an autonomic perceptual component

The supervisory controller provides five fundamental functions: command interpretation, execution scheduling, event handling, parameter regulation, and reflexive description. The supervisor acts as a programmable interpreter, receiving snippets of code script that determine the composition and nature of the process execution cycle and the manner in which the process reacts to events. The supervisor acts as a scheduler, invoking execution of modules in a synchronous manner. The supervisor handles event dispatching to other processes, and reacts to events from other processes. The supervisor regulates module parameters based on the execution results. Auto-critical reports from modules permit the supervisor to dynamically adapt processing. Finally, the supervisor responds to external queries with a description of the current state and capabilities.

Real-time visual processing for the perceptual component is provided by tracking. Tracking conserves information about over time, thus provides object constancy. Object constancy assures that a label applied to a blob at time T1 can be used at time T2. Tracking enables the system focus attention, applying the appropriate detection processes only to the region of an image where a target is likely to be detected. Also the information about position and speed provided by tracking can be very important for describing situations.

Tracking is classically composed of four phases: Predict, observe, detect, and update. The prediction phase updates the previously estimated attributes for a set of entities to a value predicted for a specified time. The observation phase applies the prediction to the current data to update the state of each target. The detect phase detects new targets. The update phase updates the list of targets to account for new and lost targets. The ability to execute different image processing procedures to process target information with an individual ROI is useful to simultaneously observe a variety of entities.

The PRIMA perceptual component architecture adds additional phases for interpretation, auto-regulation, and communication. In the interpretation phase, the tracker executes procedures that have been downloaded to the process by a configuration tool. These are interpreted by a RAVI interpreter [43] and may result in the generation of events or the output to a stream. The auto-regulation phase determines the quality of service metric, such as total cycle time and adapts the list of targets as well as the target parameters to maintain a desired quality. During the communication phase, the supervisor responds to requests from other processes. These requests may ask for descriptions of process state, or capabilities, or may provide specification of new recognition methods.

Homeostasis, or "autonomic regulation of internal state" is a fundamental property for robust operation in an uncontrolled environment. A process is auto-regulated when processing is monitored and controlled so as to maintain a certain quality of service. For example, processing time and precision are two important state variables for a tracking process. These two may be traded off against each other. The component supervisor maintains homeostasis by adapting module parameters using the auto-critical reports from modules

An auto-descriptive controller can provide a symbolic description of its capabilities and state. The description of the capabilities includes both the basic command set of the controller and a set of services that the controller may provide to a more abstract supervisor. Such descriptions are useful for both manual and automatic assembly of components.

In the context of recent National projects (RNTL ContAct) and European Projects (FAME, CAVIAR, CHIL), the PRIMA perceptual component has been demonstrated with the construction of perceptual components for

1. Tracking individuals and groups in large areas to provide services,
2. Monitoring a parking lot to assist in navigation for an autonomous vehicle.
3. Observing participants in an meeting environment to automatically orient cameras.
4. Observing faces of meeting participants to estimate gaze direction and interest.
5. Observing hands of meeting participants to detect 2-D and 3D gestures.
6. Observing and monitoring the activities of groups to understand social interaction

3.3. Service Oriented Architectures for Intelligent Environments

Keywords: *Semantic Description of Functionalities*, *Service Composition*, *Service Factories*, *Service Oriented Computing*, *Software Architecture*.

Intelligent environments are at the confluence of multiple domains of expertise. Experimenting within intelligent environments requires combining techniques for robust, autonomous perception with methods for modeling and recognition of human activity within an inherently dynamic environment. Major software engineering and architecture challenges include accomodation of a heterogeneous of devices and software, and dynamically adapting to changes human activity as well as operating conditions.

The PRIMA project explores software architectures that allow systems to be adapt to individual user preferences. Interoperability and reuse of system components is fundamental for such systems. Adopting a shared, common Service Oriented Architecture (SOA) architecture has allowed specialists from a variety of subfields to work together to build novel forms of systems and services.

In a service oriented architecture, each hardware or software component is exposed to the others as a "service". A service exposes its functionality through a well defined interface that abstracts all the implementation details and that is usually available through the network.

The most commonly known example of a service oriented architecture are the Web Services technologies that are based on web standards such as HTTP and XML. Semantic Web Services proposes to use knowledge representation methods such as ontologies to give some semantic to services functionalities. Semantic description of services makes it possible to improve the interoperability between services designed by different persons or vendors.

Taken out of the box, most SOA implementations have some "defects" preventing their adoption. Web services, due to their name, are perceived as being only for the "web" and also as having a notable performance overhead. Other implementations such as various propositions around the Java virtual machine, often requires to use a particular programming language or are not distributed. Intelligent environments involves many specialist and a hard constraint on the programming language can be a real barrier to SOA adoption.

The PRIMA project has developed OMiSCID, a middleware for service oriented architectures that addresses the particular problematics of intelligent environments. OMiSCID has emerged as an effective tool for unifying access to functionalities provided from the lowest abstraction level components (camera image acquisition, image processing) to abstract services such (activity modeling, personal assistant). OMiSCID has creately facilitated cooperation by experts from within the PRIMA project as well as in projects with external partners.

Experiments with semantic service description and spontaneous service composition are conducted around the OMiSCID middleware. In these experiments, attention is paid to usability. A dedicated language has been designed to allow developers to describe the functionalities that their services provide. This language aims at simplifying existing semantic web services technologies to make them usable by a normal developer (i.e. that is not specialized in the semantic web). This language is named the User-oriented Functionality Composition Language (UFCL).

UFCL allows developers to specify three types of knowledge about services:

- The knowledge that a service exposes a functionality like a "Timer" functionality for a service emitting message at a regular frequency.
- The knowledge that a kind of functionality can be converted to another one. For example, a "Metronome" functionality issued from a music centered application can be seen as a "Timer" functionality.
- The knowledge that a particular service is a factory and can instantiate other services on demand. A TimerFactory can for example start a new service with a "Timer" functionality with any desired frequency. Factories greatly helps in the deployment of service based applications. UFCL factories can also express the fact that they can compose existing functionalities to provide another one.

To bring the UFCL descriptions provided by the developers to life, a runtime has been designed to enable reasoning about what functionalities are available, what functionalities can be transformed to another one and what functionalities could be obtained by asking factories. The service looking for a particular functionality has just to express its need in term of functionalities and properties (e.g. a "Timer" with a frequency of 2Hz) and the runtime automates everything else: gathering of UFCL descriptions exposed by all running services, compilation of these descriptions to some rules in a rule-based system, reasoning and creation of a plan to obtained the desired functionality, and potentially invoking service factories to start the missing services.

3.4. Robust view-invariant Computer Vision

Keywords: *Affine Invariance*, *Local Appearance*, *Receptive Fields*.

3.4.1. Summary

A long-term grand challenge in computer vision has been to develop a descriptor for image information that can be reliably used for a wide variety of computer vision tasks. Such a descriptor must capture the information in an image in a manner that is robust to changes the relative position of the camera as well as the position, pattern and spectrum of illumination.

Members of PRIMA have a long history of innovation in this area, with important results in the area of multi-resolution pyramids, scale invariant image description, appearance based object recognition and receptive field histograms published over the last 20 years. The group has most recently developed a new approach that extends scale invariant feature points for the description of elongated objects using scale invariant ridges. PRIMA is currently working with ST Microelectronics to embed its multi-resolution receptive field algorithms into low-cost mobile imaging devices for video communications and mobile computing applications.

3.4.2. Detailed Description

The visual appearance of a neighbourhood can be described by a local Taylor series [40]. The coefficients of this series constitute a feature vector that compactly represents the neighbourhood appearance for indexing and matching. The set of possible local image neighbourhoods that project to the same feature vector are referred to as the "Local Jet". A key problem in computing the local jet is determining the scale at which to evaluate the image derivatives.

Lindeberg [41] has described scale invariant features based on profiles of Gaussian derivatives across scales. In particular, the profile of the Laplacian, evaluated over a range of scales at an image point, provides a local description that is "equi-variant" to changes in scale. Equi-variance means that the feature vector translates exactly with scale and can thus be used to track, index, match and recognize structures in the presence of changes in scale.

A receptive field is a local function defined over a region of an image [51]. We employ a set of receptive fields based on derivatives of the Gaussian functions as a basis for describing the local appearance. These functions resemble the receptive fields observed in the visual cortex of mammals. These receptive fields are applied to color images in which we have separated the chrominance and luminance components. Such functions are easily normalized to an intrinsic scale using the maximum of the Laplacian [41], and normalized in orientation using direction of the first derivatives [51].

The local maxima in x and y and scale of the product of a Laplacian operator with the image at a fixed position provides a "Natural interest point" [42]. Such natural interest points are salient points that may be robustly detected and used for matching. A problem with this approach is that the computational cost of determining intrinsic scale at each image position can potentially make real-time implementation unfeasible.

A vector of scale and orientation normalized Gaussian derivatives provides a characteristic vector for matching and indexing. The oriented Gaussian derivatives can easily be synthesized using the "steerability property" [30] of Gaussian derivatives. The problem is to determine the appropriate orientation. In earlier work by PRIMA members Colin de Verdiere [26], Schiele [51] and Hall [34], proposed normalising the local jet independently at each pixel to the direction of the first derivatives calculated at the intrinsic scale. This has provided promising results for many view invariant image recognition tasks as described in the next section.

Color is a powerful discriminator for object recognition. Color images are commonly acquired in the Cartesian color space, RGB. The RGB color space has certain advantages for image acquisition, but is not the most appropriate space for recognizing objects or describing their shape. An alternative is to compute a Cartesian representation for chrominance, using differences of R, G and B. Such differences yield color opponent receptive fields resembling those found in biological visual systems.

Our work in this area uses a family of steerable color opponent filters developed by Daniela Hall [34]. These filters transform an (R,G,B), into a cartesian representation for luminance and chrominance (L,C1,C2). Chromatic Gaussian receptive fields are computed by applying the Gaussian derivatives independently to each of the three components, (L, C1, C2). The components C1 and C2 encodes the chromatic information in a Cartesian representation, while L is the luminance direction. Chromatic Gaussian receptive fields are computed by applying the Gaussian derivatives independently to each of the three components, (L, C1, C2). Permutations of RGB lead to different opponent color spaces. The choice of the most appropriate space depends on the chromatic composition of the scene. An example of a second order steerable chromatic basis is the set of color opponent filters shown in figure 2.
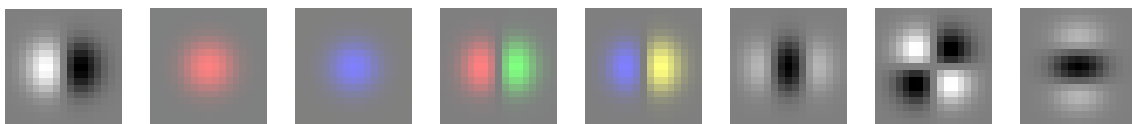

Figure 2. Chromatic Gaussian Receptive Fields ($G_x^L, G^{C_1}, G^{C_2}, G_x^{C_1}, G_x^{C_2}, G_{xx}^L, G_{xy}^L, G_{yy}^L$).

Key results in this area include

1. Fast, video rate, calculation of scale and orientation for image description with normalized chromatic receptive fields [29].
2. Real time indexing and recognition using a novel indexing tree to represent multi-dimensional receptive field histograms [48].
3. Robust visual features for face tracking [33], [32].
4. Affine invariant detection and tracking using natural interest lines [54].
5. Direct computation of time to collision over the entire visual field using rate of change of intrinsic scale [44].

We have achieved video rate calculation of intrinsic (characteristic) scale from interpolation within a Binomial Pyramid computed using an O(N) algorithm [29]. This software provides a practical method for obtaining invariant image features for detection, tracking and recognition at video rates. This method has been used in the real time BrandDetect system, for detecting publicity panels in broadcast video of sports events, as described below.

Daniela Hall and Nicolas Gourier have developed machine learning techniques to statistically learn robust visual features for face tracking [33], [32].

Augustin Lux and Hai Tranh have developed a method with provides a direct measurement of affine invariant local features based on extending natural interest points to "natural interest ridges" [56], [55]. The orientation of natural interest ridges provides a local orientation in the region of an image structure. Early results indicate an important gain in discrimination rates compared to SIFT and and other histogram based detection approaches. An example of the dominant interest ridges used for tracking of people in the entrance hall of INRIA Rhone Alpes is shown in 3.
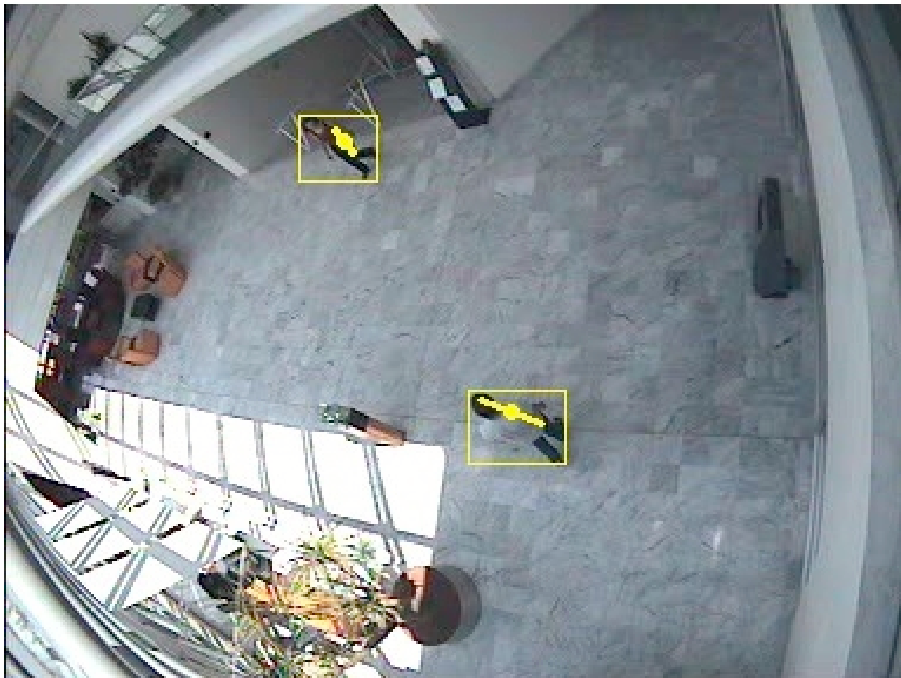

Figure 3. Dominant natural interest ridges for tracking people

Doctoral student Amaury Negre has adapted the scale invariant ridge detection for use in detection and tracking of obstacles for autonomous vehicle navigation. In this work, the characteristic size of objects, provided by the scale of dominant ridges is directly used to calculate time to contact. Using this method, he has demonstrated direct computation of time to contact using rate of change of intrinsic scale [44]. This approach is currently being adapted for use in visual navigation in joint work with project EMOTION. Amaury Negre will defend his doctoral dissertation on these methods in early 2009.

Doctoral student Jean-Pascal Mercier has recently begun work extending the ridge description methods to 3D and 4D spatio-temporal volumes in order to detect structures for recognizing human actions.

3.5. New forms of man-machine interaction based on perception

Keywords: *Augmented Reality*, *Projected interaction widgets*, *Steerable Camera Projector*.

Surfaces are pervasive and play a predominant role in human perception of the environment. Augmenting surfaces with projected information provides an easy-to-use interaction modality that can easily be adopted for a variety of tasks. Projection is an ecological (non-intrusive) way of augmenting the environment. Ordinary objects such as walls, shelves, and cups may become physical supports for virtual functionalities [46]. The original functionality of the objects does not change, only its appearance. An example of object enhancement is presented in [24], where users can interact with both physical and virtual ink on a projection-augmented whiteboard.

Combinations of a camera and a video projector on a steerable assembly [23] are increasingly used in augmented environment systems [45] [49] as an inexpensive means of making projected images interactive. Steerable projectors [23] [46] provide an attractive solution overcoming the limited flexibility in creating interaction spaces of standard rigid video-projectors (e.g. by moving sub windows within the cone of projection in a small projection area [57]).

The PRIMA group has constructed a new form of interaction device based on a Steerable Camera-Projector (SCP) assembly. This device allows experiments with multiple interactive surfaces in both meeting and office environments. The SCP pair, shown in figure 4, is a device with two mechanical degrees of freedom, pan and tilt, mounted in such a way that the projected beam overlaps with the camera view. This creates a powerful actuator-sensor pair enabling observation of user actions within the camera field of view. This approach has been validated by a number of research projects as the DigitalDesk [59], the Magic Table [24] or the Tele-Graffiti application [53].
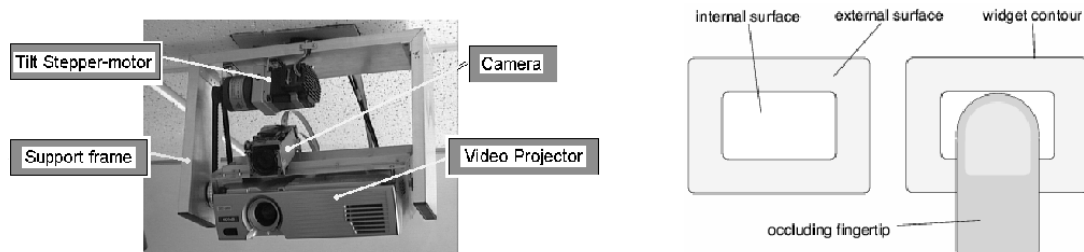

Figure 4. Steerable camera-projector pair (left) and surfaces defined to detect touch-like gestures over a widget (right)

For the user interaction, we are experimenting with interaction widgets that detect fingers dwelling over button-style UI elements, as shown to the right in figure 4.

Given the limited personnel available to pursue this area, we have concentrated our efforts on

1. Analysis of the the mathematical foundations for projected interaction devices, and

2. Developing software toolkits that provide easy programming for a wide variety of interaction models.

An important challenge is real time rectification for both the projected interaction patterns, and the perceptual field in which actions are observed. When the projected workspace is fixed, it is possible to pre-calibrate the homographies that relate the projected pattern and sensitive field. However, when the interaction surface is free to travel around the environment, these homographies must be re-computed in real time.

To provide real time re-calibration, we have implemented a procedure that detects and tracks the boundaries of a rectangular screen, referred to as the "portable display screen" or PDS. The intersection of the four boundary lines provides the image location of the observed corners of the PDS, which are then used to directly recalculate the transformation from camera to screen. Because the camera is rigidly mounted to the projector, the relation between the camera and the projector is also a homography. This homography is precalibrated using projected patterns as a calibration grid, The product of the homography from projector to camera, and the homography from camera to screen, gives the homography from projector to screen.

Evaluating the entire Hough space from scratch can be costly, and can lead to errors. In order to provide fast, robust, estimation, we track each peak in the Hough space using a robust tracking procedure based on a Kalman filter. The result is a fast, robust method for real time estimation of the projections from camera and projector to display screen. This method has been published at the first ProCams workshop, [23] and is now often cited in the camera-projector community.

In order to develop experiments with projected interaction widgets, we have recently developed a component-oriented programmers tool-kit for vision-based interactive systems, taking inspiration from [39]. In this toolkit, we separate vision components for interaction from the functional core of the application. The implementation of the vision-components draws on the VICs framework presented by Ye et. al in [60].

This tool-kit approach to interactive system design seeks to minimize the difficulties related to the deployment of perceptual user interface by:
a) encapsulating vision components in isolated services,
b) imposing these services to meet specific usability requirements, and
c) limiting communications between the services and the interactive applications to a minimum.

Our early demonstrations of the device were met with enthusiasm by our partners at Xerox Research XRCE, Univeristy of Karlsruhe, IRST Trento, and France Telecom. In October 2003, we participated in a workshop on Camera-Projector systems organised by IBM Watson and Carnegie Mellon University. Since then an enthusiastic community of researchers has formed around this subject, and devices and innovations are moving quickly from laboratory to commercial applications.

Several important commercial opportunities have recently presented themselves. The start up company HiLabs has recently been created such for use in interactive publicity and information kiosk in public places.

4. Application Domains

4.1. The Smart Environments Plateforme

Keywords: *Augmented Reality*, *Collaborative Work*, *Multi-modal Interaction*.

Participants: Patrick Reignier, Dominique Vaufreydaz, Augustin Lux, Rémi Emonet, Matthieu Langet, Jérôme MAISONNASSE, Oliver Brdiczka, Sofia Zaidenberg, Nicolas Gourier, James Crowley.

Project PRIMA has constructed Smart Environments Plateforme, show in figure 5, in order to test and develop context aware interactive services. The PRIMA Smart Environments Plateforme is equipped with a microphone array, wireless lapel microphones, wide angle surveillance cameras, panoramic cameras, steerable cameras, scanning range sensors and two camera-projector video-interaction devices. The microphone array is used as an acoustic sensor to detect, locate and classify acoustic signals for recognizing human activities. The wide-angle and panoramic cameras provide fields of view that cover the entire room, and allows detection and tracking of individuals. Steerable cameras are used to acquire video of activities from any viewing direction.
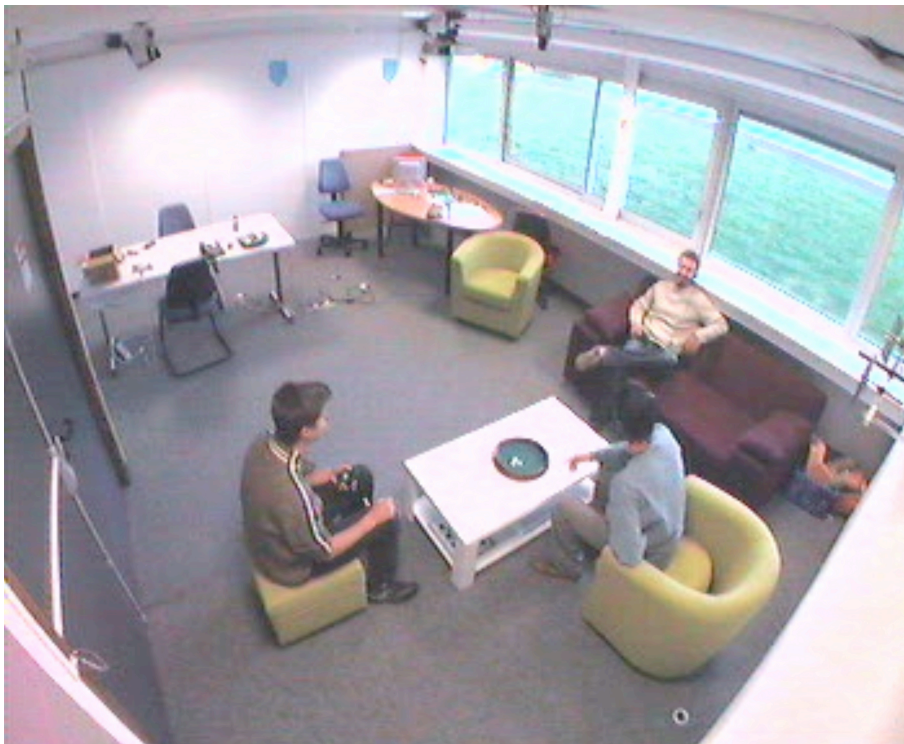
Figure 5. The Smart Environments Plateforme is an modular laboratory environment equipped with a microphone array, 32 wireless lapel microphones, wide angle surveillance cameras, a panoramic camera, steerable cameras, scanning range sensors, contact sensors and camera-projector video-interaction devices. This plateforme is used for experiments in context aware interactive services.

Video interaction devices associate a camera with a video projector to provide new modes if man-machine interaction. Such devices may be used for interaction, presentation or capture of information based on natural activity. Examples include selecting menus and buttons with a finger and capturing drawings from paper or a whiteboard. Fixed video interaction devices in the AME have been constructed for a vertical surface (a wall mounted white board) and a horizontal desk-top work-space. Recently a steerable interaction device has been constructed based on a tightly integrated steerable camera-projector pair (SCP). The SCP described below, allows any surface to be used for interaction with information. It also offers a range new sensing techniques, including automatic surveillance of an environment to discover the environment topology, as well as the use of structured light for direct sensing of texture mapped 3D models.

4.2. Ambient Services for Assisted Living

Keywords: *Ambient Assisted Living*, *Monitoring Services*, *Presence awareness*.

Participants: Matthieu Langet, Frédéric Devernay, Amaury Negre, James Crowley.

The continued progress in extending life-span, coupled with declining birth rates have resulted in a growing number of elderly people with varying disabilities who are unable to conduct a normal life at home, thereby becoming more and more isolated from society. Governmental agencies including hospitals, healthcare institutions and social care institutions are increasingly overburdened with care of this growing population. Left unchecked, economic and man-power requirements for care of the elderly could well trigger a societal and economic crisis. There is an urgent societal need for technologies and services that allow elderly people to live autonomously in their own environments for longer periods. Smart environments provide a promising new enabling technology for such services.

Adapting smart environments to enhance the autonomy and quality of life for elderly require:

- Robust, plug-and-play sensor technologies monitor the activities and health of elderly in their own home environments.

- Easy to use communications services that allow people to maintain a sense of presence to avoid isolation without disrupting privacy or distracting attention from normal daily activities.

- Architectural frameworks that allow ad hoc composition of services from distributed heterogeneous components scattered throughout the environment.

- Distributed system architectures which allow the cooperation of independent emergency services to work together to provide emergency care,

- Technologies interpret activity to warn of loss of mobility or cognitive function.

- Engineering approaches for the customization/personalization/adaptation of living assistance systems at installation and run time,

- Social, privacy, ethical and legal safeguards for privacy and control of personal data.

The PRIMA Project works with SME company H2AD as well as France Telecom to develop ambient assisted living technologies in the context project ANR CASPER. Within CASPER, PRIMA is developing components for robust multi-modal sensing and monitoring of activity, as well as a plug-and-play middle ware architecture for dynamic composition of services, and new methods for end user configuration and adaptation of service behaviour.

4.3. The Steerable Camera Projector

Keywords: *Interactive Environments*, *Man-Machine Interaction*.

Participants: Stan Borkowski, Julien Letessier, James Crowley.

Surfaces dominate the physical world. Every object is confined in space by its surface. Surfaces are pervasive and play a predominant role in human perception of the environment. We believe that augmenting surfaces with information technology will proved an interaction modality that will be easily adopted by humans.

PRIMA has constructed a steerable video interaction device composed of a tightly coupled camera and video projector. This device, known as a Steerable Camera-Projector (or SCP) enables experiments in which any surface in the augmented meeting environment may be used as an interactive display for information [23]. With such a device, an interaction interface may follow a user, automatically selecting the most appropriate surface. The SCP provides a range of capabilities *(a)*The SCP can be used a sensor to discover the geometry of the environment, *(b)*The SCP can project interactive surfaces anywhere in the environment and *(c)* The SCP can be used to augment a mobile surface into a portable interactive display. *(d)* The SCP can be used to capture text and drawings from ordinary paper. *(e)* The SCP can be used as a structured light sensor to observe 3-D texture-mapped models of objects.

Current display technologies are based on planar surfaces. Recent work on augmented reality systems has assumed simultaneous use of multiple display surfaces [36], [52], [58]. Displays are usually treated as access points to a common information space, where users can manipulate vast amounts of information with a set of common controls. With the development of low-cost display technologies, the available interaction surface will continue to grow, and interfaces will migrate from a single, centralized screen to multiple, space-distributed interactive surfaces. New interaction tools that accommodate multiple distributed interaction surfaces will be required.

Projection is an ecological (i.e. non-intrusive) way of augmenting the environment. Projection does not change the augmented object itself, only its appearance. This change can be used to supplement the functionality of the object and henceforth its role in the world. However, the most common consequence of augmenting an object with projected images is transforming the object into an access point to the virtual information space. In [46] ordinary artifacts such as walls, shelves, and cups are transformed into informative surfaces. Though the superimposed projected image enables the user to take advantage of the information provided by the virtual world, the functionality of the object itself does not change. The object becomes a physical support for virtual functionalities. An example of enhancing the functionality of an object was presented in [25], where users could interact with both physical and virtual ink on an projection-augmented whiteboard.

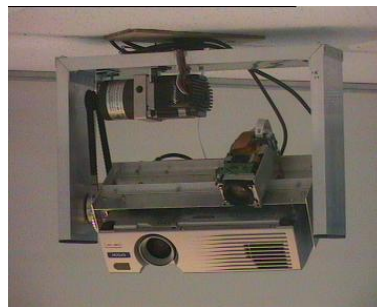*Figure 6. The Steerable Camera Projector*

The Steerable Camera Projector (SCP) (figure 6) platform is a device that provides a video-projector with two mechanical degrees of freedom: pan and tilt.

4.4. Video processing for 3-D displays

Keywords: *3-D display*, *Auto-calibration*, *Stereoscopy*, *View Interpolation*.

Participants: Frédéric Devernay, Sergi Pujades, Elise Mansilla, Martin Guillon.

Until recently, in order to view stereoscopic 3-D video, the user had to wear special glasses. Recent advances in 3-D displays provide true 3-D viewing experience without glasses. These screens use either a micro-lenticular network or a parallax barrier placed in front of a standard LCD, plasma, or LED display, so that different viewpoints provide different images. If the characteristics of the network and the screen are carefully chosen, the user will perceive two different images from the viewpoints of the left and right eyes. Such glasses-free 3-D screens usually display between 8 and a few dozen different viewpoints.

When the 3-D scene which has to be displayed is computer-generated, it is usually not a problem to generate a few dozen viewpoints. But when a real scene has to displayed, one would have to shoot it through the same number of synchronized cameras as there are viewpoints in order to display it properly. This makes 3-D shooting of real scenes for glasses-free 3-D displays mostly unpractical. For this reason, we are developing high-quality view-interpolation techniques, so that the many different viewpoints can be generated from only a few camera positions.

Another problem with 3-D displays is transmission bandwidth. There is much redundant information between the images to be displayed at different viewpoints. Transmitting video streams separately, even if each of these is compressed, would not take advantage of this redundancy. Most video compression techniques only take into account the temporal redundancy of video streams. For 3-D video compression we must also exploit the redundancy between viewpoints. We are approaching this problem using stereo-vision techniques that are currently used for view interpolation.

4.5. Context Aware Video Acquisition

Keywords: *Context Aware Systems*, *Intelligent Environments*, *Video Conferencing*.

Participants: Patrick Reignier, Dominique Vaufreydaz.

Video communication has long been seen as a potentially powerful tool for communications, teaching and collaborative work. Continued exponential decreases in the cost of communication and computation (for coding and compression) have eliminated the cost of bandwidth as an economic barrier for such technology. However, there is more to video communication than acquiring and transmitting an image. Video communications technology is generally found to be disruptive to the underlying task, and thus unusable. To avoid disruption, the video stream must be composed of the most appropriate targets, placed at an appropriate size and position in the image. Inappropriately composed video communications create distraction and ultimately degrades the ability to communicate and collaborate.

During a lecture or a collaborative work activity, the most appropriate targets, camera angle, and zoom and target position change continually. A human camera operator understands the interactions that are being filmed and adapts the camera angle and image composition accordingly. However, such human expertise is costly. The lack of an automatic video composition and camera control technology is the current fundamental obstacle to the widespread use of video communications for communication, teaching and collaborative work. One of the goals of project PRIMA is to create a technology that overcomes this obstacle.

To provide a useful service for a communications, teaching and collaborative work, a video composition system must adapt the video composition to events in the scene. In common terms, we say that the system must be "aware of context". Computationally, such a technology requires that the video composition be determined by a model of the activity that is being observed. As a first approach, we propose to hand-craft such models as finite networks of states, where each state corresponds to a situation in the scene to be filmed and specifies a camera placement, camera target, image placement and zoom.

A finite state approach is feasible in cases where human behavior follows an established stereotypical "script". A lecture or class room presentation provides an example of such a case. Lecturers and audiences share a common stereotype about the context of a lecture. Successful video communications require structuring the actions and interactions of actors to a great extent. We recognize that there will always be some number of unpredictable cases where humans deviate from the script. However, the number of such cases should be sufficiently limited so as limit the disruption. Ultimately, we plan to investigate automatic techniques for "learning" new situations.

This system described above is based on an approach to context aware systems presented at UBICOMP in September 2002 [27]. The behavior of this system is specified as a situation graph that is automatically compiled into rules for a Java based supervisory process. The design process for compiling a situation graph into a rule based for the federation supervisors has been developed and refined within the last two years.

Since 2004, we have demonstrated a number of systems based on this model. In the FAME project, we demonstrated a context aware video acquisition system at the Barcelona Forum of Cultures during two weeks in July 2004. This system was also demonstrated publicly at "Fete de la science" in Grenoble in October 2004, and exhibited at the IST Conference in Den Haag in November 2004. A variation of this system has been integrated into the ContAct context aware presentation composition system developed with XRCE (Xerox European Research Centre), and is at the heart of the CHIL Collaborative Workspace Service used in the IP Project CHIL. A context aware interpretation system for video surveillance is currently under development for the IST project CAVIAR.

This system is under permanent installation in a meeting room and in an amphitheater of the LIG laboratory. This installation is a part of the LIG plateform demonstrators.

4.6. Context Aware Personnal Assistant

Keywords: *Context Aware Systems*, *Intelligent Environments*, *Personnal agent*.

Participants: Sofia Zaidenberg, Patrick Reignier, James Crowley, Rémi Barraquand.

As embedded computing matures, it is increasingly possible to build low-cost mobile devices that integrate sensing, display, computing, communications, and interaction. As cost have decreased and technologies matured, the number of such devices in ordinary human environments has doubled roughly every 2 years, in a progression driven by Moore's law. Wireless ad-hoc network technology allows such devices to be federated to create a new form of interconnected distributed computing environment [22]. In such environments, services are not limited to the resources of a single machine, but may be dynamically composed as an assembly of distributed components. An important challenge is to create "intelligent" services that exploit such environments to provide access to information and communications in a manner that is appropriate and non-disruptive.

To be considered as "intelligent", a service or system must incarnated, autonomous and situated. Services may be incarnated as an assembly of software and hardware components within a smart environment. The OMISCID middleware, developed by the PRIMA group, enables experiments in the dynamic construction of systems and services from software and hardware components. In order to maintain autonomy, a service or system must be able to monitor and reconfigure itself to continue robust operation in the presence of changes in operating conditions or support hardware. Autonomy is made possible by constructing components using autonomic programming techniques such as self-monitoring, auto-regulation and self repair.

Situated behaviour requires that the actions and reactions of the system be appropriate to the current context. In the PRIMA group we have developed situation models as a method to enable systems and services to model human activities and social contexts. Actions and reactions are made contingent on the current situation in order to provide services that are both appropriate and non-disruptive.

Within project PRIMA we are investigating the use of reinforcement learning to automatically construct a context aware personal agent. Rewards are given by the user when expressing his satisfaction of the system's actions. A default context model assures a consistent initial behavior. This default model is provided by the agent programmer and is neutral enough to satisfy a majority of users. This model is then adapted to each particular user in a way that maximizes its satisfaction. The learned system must also be able to explain all its actions to gain user's acceptance.

4.7. User localization in large-scale perceptive environment using multimodal heterogenic analysis

Keywords: *Context Aware Systems*, *Intelligent Environments*, *Multimodality*, *localization*.

Participants: Dominique Vaufreydaz, Han Yue, Matthieu Langet.

Ad-hoc assemblies of mobile devices embedding sensing, display, computing, communications, and interaction provide an enabling technology for smart environments. As described above, in the PRIMA project we have adopted a component oriented programming approach to compose smart services for such environments. Common services for smart environement include

- Services to manage energy in building, including regulating temperature, illumination, and acoustic noise,
- Ambient assisted living services to extend the autonomy of elderly and infirm,
- Logistics management for daily living,
- Communication services and tools for collaborative work,
- Services for commercial environments,
- Orientation and information services for public spaces, and
- Services for education and training.

We wish to develop the concept of "large-scale" perceptive space that is an intelligent environment which will be deployed on a large surface containing several buildings (as a university campus for example). We also define the "augmented man" concept as a human wearing one or many mobile intelligent wireless devices (telephone, Smartphone, pda, notebook). Using all these devices, one can use many different applications (read emails, browse the Internet, file exchange, etc.). By combining the concepts of large-scale perceptive environments and mobile computing, we can create intelligent spaces, it becomes possible to propose services adapted to individuals and their activities. We are currently focussing on two aspects of this problem: the user profile and the user location within a smart space.

A fundamental requirement for such services is the ability to perceive the current state of the environment. Depending on the nature of the service, environment state can require sensing and modeling the physical properties of the environment, the location, identity and activity of individuals within the environment, as well as the set of available computing devices and software components that compose the environment. All of these make up possible elements for context modeling.

Observing and tracking people in smart environments remains a challenging fundamental problem. Whether it is at the scale of a campus, of a building or more simply of a room, we can combine several additional localization levels (and several technologies) to allow a more accurate and reliable user perception system. Within the PRIMA project, we are currently experimenting with a multi-level localization system allowing variable granularity according to the available equipment and the precision required for the targeted service. This approach is in the same research area as the MagicMap system. However, contrary to MagicMap, we enrich location information from wireless technologies with information from 3D tracking using cameras, microphones, and other sensing devices.

5. Software

5.1. IMALAB

Keywords: *Computer Vision Systems*, *Software Development Environments*.

Participants: Augustin Lux [correspondant], Rémi Emonet.

The Imalab system represents a longstanding effort within the Prima project group (1) to capitalize on the work of successive generations of students, (2) to provide a coherent software framework for the development of new research, and (3) to supply a powerful toolbox for sophisticated applications. In its current form, ImaLab serves as a development environment for research in computer vision in the Prima team, and represents a considerable amount of effort (probably largely more than 10 man-years).

There are two major elements of the Imalab system: the PrimaVision library, which is a C++ based class library for the fundamental requirements of research in computer vision; and the Ravi system, which is an extensible system kernel providing an interactive programming language shell.

With respect to other well known computer vision systems, e.g. KHOROS [50] the most prominent features of Imalab are:

- A large choice of data structures and algorithms for the implementation of new algorithms.
- A subset of C++ statements as interaction language.
- Extensibility through dynamic loading.
- A multi language facility including C++, Scheme, Clips, Prolog.

The combination of these facilities is instrumental for achieving efficiency and generality in a large Artificial Intelligence system: efficiency is obtained through the use of C++ coding for all critical pieces of code; this code is seamlessly integrated with declarative programs that strive for generality.

Imalab's system kernel is built on the Ravi system first described in Bruno Zoppis's thesis [61]. The particular strength of this kernel comes from a combination of dynamic loading and automatic program generation within an interactive shell, in order to integrate new code, even new libraries, in a completely automatic way.

The Imalab system has, in particular, been used for the development of software in several European projects. The Imalab system has proven to be extremely efficient tool for the development of systems such as BrandDetect that need extensive performance evaluation as well as incremental design of of a complex user interface.

We currently are in the process of registering of ImaLab with the APP (Agence pour la Protection des Programmes). Imalab has been distributed as share ware to several research laboratories around Europe. Imalab has been installed and is in use at:

- XRCE - Xerox European Research Centre, Meylan France
- JOANNEUM RESEARCH Forschungsgesellschaft mbH, Austria
- HS-ART Digital Service GmbH, Austria
- VIDEOCATION Fernseh-Systeme GmbH, Germany
- Univ. of Edinburgh, Edinburgh, UK
- Instituto Superior Tecnico, Lisbon, Portugal
- Neural Networks Research Centre, Helsinki University of Technology (HUT), Finland
- Jaakko Poyry Consulting, Helsinki, Finland
- Université de Liège, Belgium
- France Télécom R&D, Meylan France

5.2. OMiSCID Middleware for Distributed Multi-Modal Perception

Keywords: *Distributed perceptual systems*, *Middleware*.

Participants: Patrick Reignier, Dominique Vaufreydaz [correspondant], Rémi Emonet, Jean-Pascal Mercier.

OMiSCID is new lightweight middleware for dynamic integration of perceptual services in interactive environments. This middleware abstracts network communications and provides service introspection and discovery using DNS-SD (*DNS-based Service Discovery* [21]). Services can declare simplex or duplex communication channels and variables. The middleware supports the low-latency, high-bandwidth communications required in interactive perceptual applications. It is designed to allow independently developed perceptual components to be integrated to construct user services. Thus our system has been designed to be cross-language, cross-platform, and easy to learn. It provides low latency communications suitable for audio and visual perception for interactive services.

OMiSCID has been designed to be easy to learn in order to stimulate software reuse in research teams and is revealing to have a high adoption rate. To maximize this adoption and have it usable in projects involving external partners, the OMiSCID middleware has been released under an open source licence. To maximize its target audience, OMiSCID is available from a wide variety of programming languages: C++, Java, Python and Matlab. A website containing informations and documentations about OMiSCID has been set up to improve the visibility and promote the use of this middleware.

The OMiSCID graphical user interface (GUI) is an extensible graphical application that facilitates analysis and debugging of service oriented applications. The core functionality of this GUI is to list running services, their communication channels and their variables. This GUI is highly extensible and many modules (i.e. plugins) have been created by different members of the team: figure 7 shows an example of some of these modules. OMiSCID GUI is based on the Netbeans platform and thus inherits from its dynamic installation and update of modules.
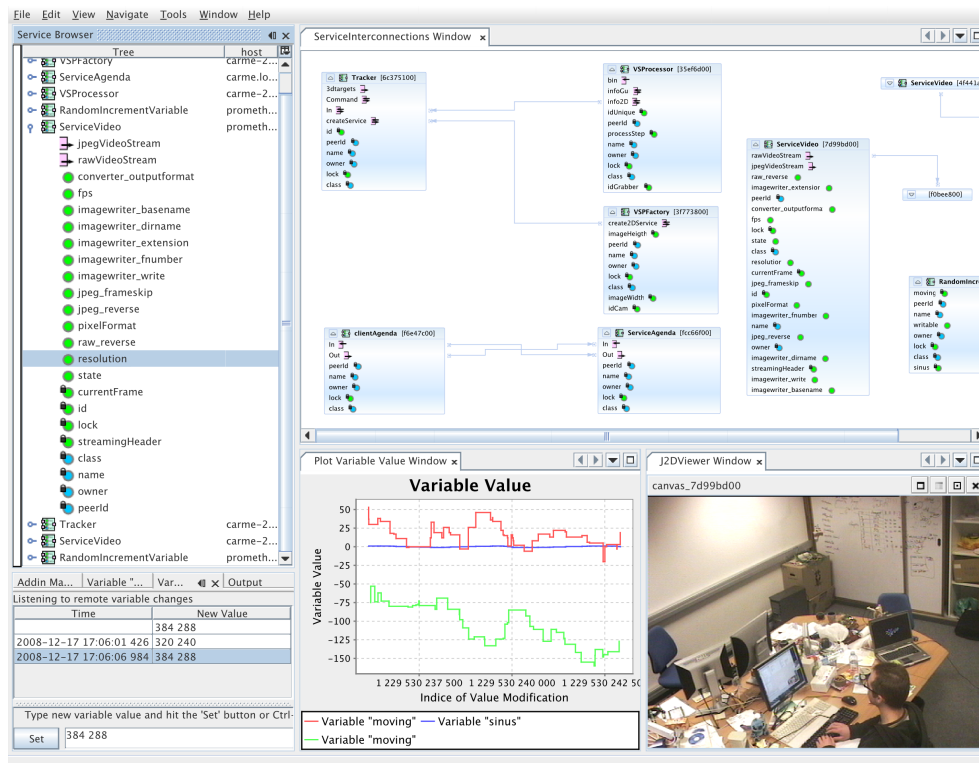

Figure 7. OMiSCID GUI showing a list of running services and some modules for service interconnections, variable plotting, live video stream display and variable control

5.3. Robust Real-Time Bayesian Tracking

Keywords: *Computer Vision Systems*, *Monitoring*, *Robust Tracking*, *Video Surveillance*.

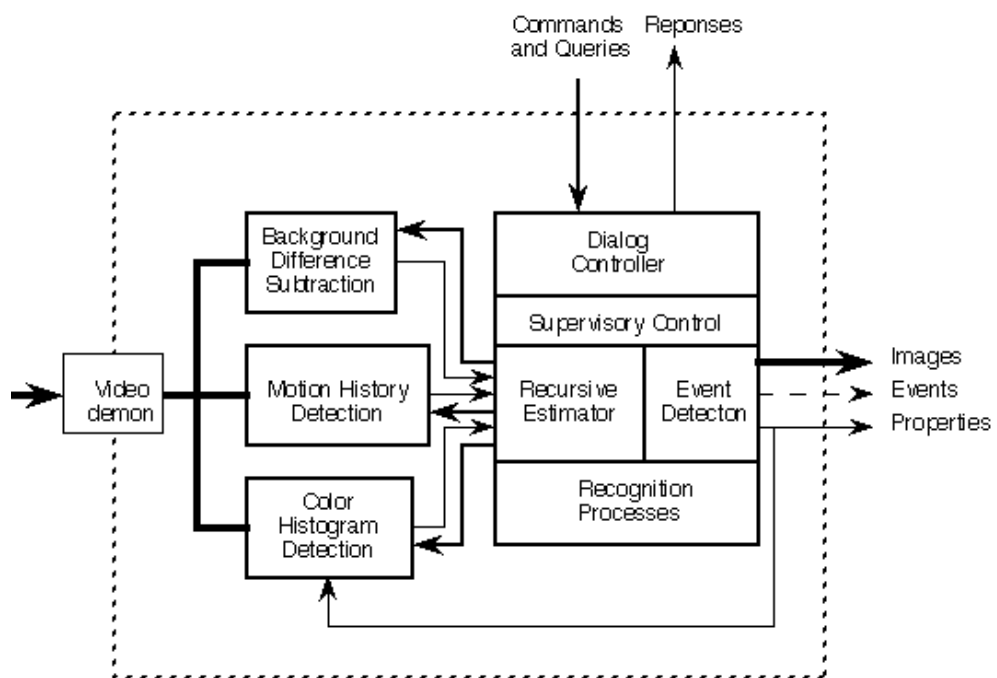**Participants:** James Crowley [correspondant], Justus Piater, Stephane Richetto.

Figure 8. The CAR systems integrates several detection modules with a Kalman Filter for robust detection and tracking of entities

Tracking is a basic enabling technology for observing and recognizing human actions. A tracking system integrates successive observations of targets so as to conserve information about a target and its history over a period of time. A tracking system makes it possible to recognize an object using off-line (non-video rate) processes and to associate the results of recognition with a target when it is available. A tracking system makes it possible to collect spatio-temporal image sequences for a target in order to recognize activity. A tracking system provides a prediction of the current location of a target which can improve the reliability, and reduce the computational cost of observation.

Project PRIMA has created and licensed several systems for robust real-time visual tracking based Baysian estimation. These include CAR (2003), 2DBT (2006), 3DBT (2006), SuiviDeCiblesCouleur (2007) and FaceStabilisationSystems (2007).

The CAR system was designed for observing the actions of individuals in a commercial or public environment, and is designed to be general so as to be easily integrated into other applications. This system has been filed with the APP "Agence pour la Protection des Programmes" and has Interdeposit Digital number of IDDN.FR.001.350009.000.R.P.2002.0000.00000. The basic component for the CAR systems is a method for robust detection and tracking of individuals [Schwerdt 00]. The system is robust in the sense that it uses multiple, complementary detection methods are used to ensure reliable detection. Targets are detected by pixel level detection processes based on back-ground subtraction, motion patterns and color statistics.

The CAR system uses techniques based on statistical estimation theory and robust statistics to predict, locate and track multiple targets. The location of targets are determined by calculating the center of gravity of detected regions. The spatial extent of a targets are estimated by computing the second moment (covariance) of detected regions. A form or recursive estimator (or Kalman filter) is used to integrate information from the multiple detection modes. All targets, and all detections are labeled with a confidence factor. The confidence factor is used to control the tracking process and the selection of detection mode.

In 2003, with the assistance by INRIA Transfert and the GRAIN, the PRIMA group has founded a small enterprise, Blue Eye Video to develop commercial applications based on the CAR system. Blue Eye Video has been awarded an exclusive license for commercial application of the CAR tracker. In June 2003, Blue Eye Video was named Laureat of the national competition for the creation of enterprises. Since mid 2004, the number of system installed by Blue Eye Video has been growing rapidly. During 2006 Blue Eye Video employed 9 persons, with annual sales of nearly 400 K Euros.

5.4. PRIMA Automatic Audio-Visual Recording System

Keywords: *audio-visual recording system.*

Participants: Patrick Reignier [correspondant], Dominique Vaufreydaz.

The PRIMA automatic audio-visual recording system controls a battery of cameras and microphones to record and transmit the most relevant audio and video events in a meeting or lecture. The system uses a can employ both steerable and fixed cameras, as well as a variety of microphones to record synchronized audio-video streams. Steerable cameras automatically oriented and zoomed to record faces, gestures or documents. At each moment the most appropriate camera and microphones are automatically selected for recording. System behaviour is specified by a context model. This model, and the resulting system behaviour, can be easily edited using a graphical user interface.

In video-conferencing mode, this system can be used to support collaborative interaction of geographically distributed groups of individuals. In this mode, the system records a streaming video, selecting the most appropriate camera and microphone to record speaking inviduals, workspaces, recorded documents, or an entire group. In meeting minute mode, the system records a audio-visual record of "who" said "what".

The system is appropriate for business, academic and governmental organizations in which geographically remote groups must collaborate, or in which important meetings are to be recorded for future reference.

The primary innovative features are:

1. Dynamic real time detection and tracking of individuals and workspace tools
2. Dynamic real time 3D modeling of the scene layout.
3. Dynamic recognition and modeling of human activity using stereotypical graphs of situations.

This system can dramatically improve the effectiveness of group video conferencing. It can eliminate the need for human camera operators and editors for recording public meetings. The product reduces energy consumption by allowing video conferencing to serve as an effective alternative to airline travel.

Market for this system is determined by:

1. The number of "business meetings" between remote collaborators
2. The number of important meetings for which an audio-visual record should be maintained.

This system is actually part of our demonstration environment. It is under deployement at one of the main conference room on the campus: the Jean Kuntzman amphitheater. Another plan is to use this automatic audio-visual recording system at the ENSIMAG school in order to permit students in two 15kms-separated buildings to follow the same curses at the same time sharing and interacting with the same professor.

5.5. Stereo Viewfinder

Keywords: *Auto-calibration*, *Feature matching*, *Real-time video processing*, *Stereoscopy*.

Participants: Frédéric Devernay [correspondant], Elise Mansilla, Loic Lefort, Sergi Pujades.

This software has been filed with the APP "Agence pour la Protection des Programmes" under the Interdeposit Digital number IDDN.FR.001.370083.000.S.P.2007.000.10000

6. New Results

6.1. 3D Bayesian Tracker

Keywords: *Multi-modal tracking*.

Participants: Augustin Lux [correspondant], Rémi Emonet, James Crowley, Matthieu Langet, Jean-Pascal Mercier.

The 2DBT and 3DBT tracking systems are autonomic perceptual components created for the IST CAVIAR project and the IST CHIL project. Both systems are autonomous perceptual components managed by a autonomic supervisor. The Autonomic supervisor provides self monitoring, self repair, self configuration, auto-regulation of parameters and self-description.

The INRIA 3D Bayesian body tracker is used to detect, locate and track multiple 3D entities in a CHIL room in real time. It is configured and optimized for detecting and tracking people within CHIL rooms using multiple calibrated cameras. In theory the camera set can include an overhead panoramic camera, but as of this writing, the system has only been tested with wall mounted cameras, typically found in the upper corners of CHIL rooms. Cameras may be connected and disconnected while the component is running, but they must be pre-calibrated to a common room reference frame. The calibration data is obtained by reading a file obtained from the CHIL KBS.

This perceptual component can be configured to monitor and track the activity within a CHIL room. The tracker receives its observations from 2D detection process that can use any available pixel level detection algorithm. The tracker currently integrates information from adaptive background subtraction, motion detection, skin color detection, and local appearance using scale normalised Gaussian derivatives. A common scenario is to use the motion to detect and initialise tracking, adaptive background subtraction to track 3D bodies, and skin color to track hands and faces. Cameras may be connected dynamically.
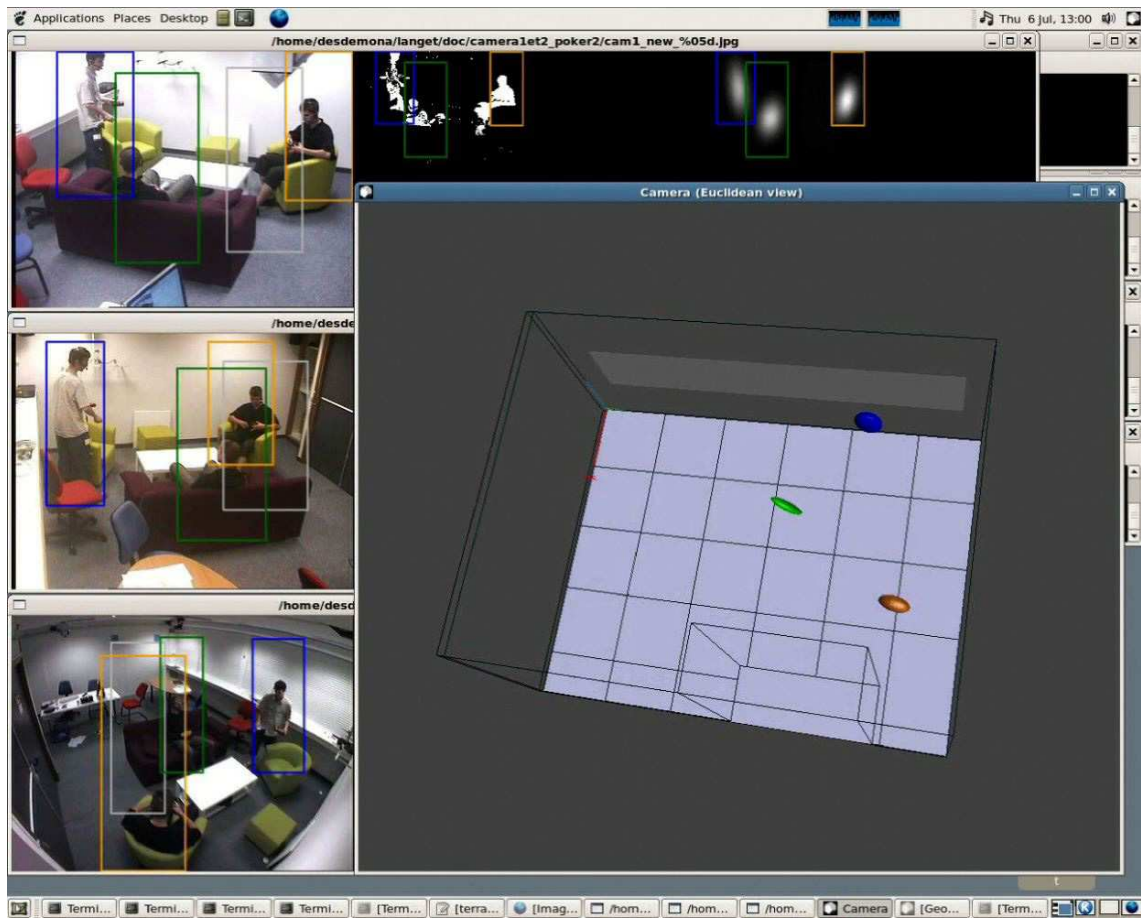
Figure 9. The 3D Bayesian tracker integrates observations from multiple sensors

The system 3DBT has been declared with the APP "Agence pour la Protection des Programmes" under the Interdeposit Digital number IDDN.FR.001.490023.000.S.P.2006.000.10000

6.2. Robust Real-Time Detection and Tracking System

Keywords: *Computer Vision Systems*, *Event Detection*, *Monitoring*, *Robust Tracking*, *Video Surveillance*.

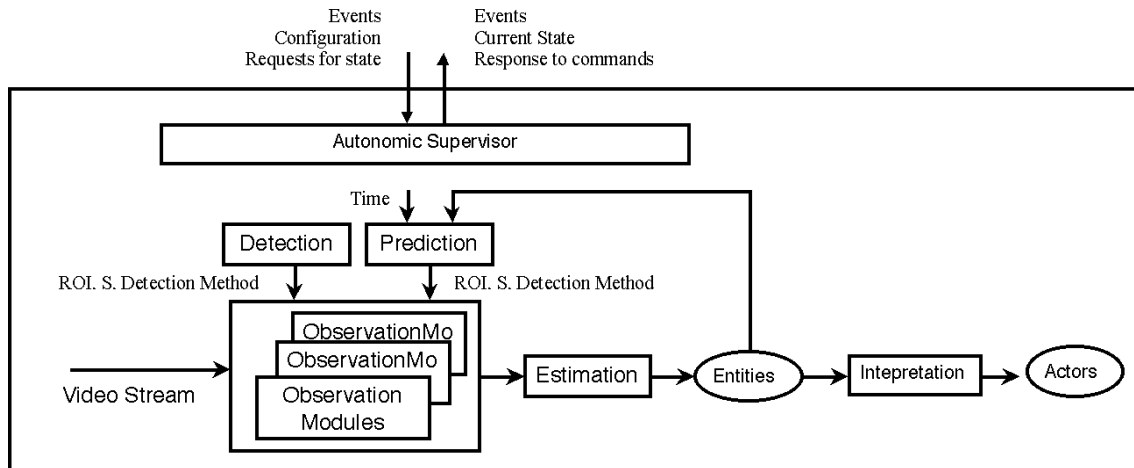**Participants:** James Crowley [correspondant], Daniela Hall, Patrick Reignier.

Figure 10. The components and architecture for the new agent detection and tracking process.

The PRIMA robust tracker is a "stand-alone" software components that can be used to detect and recognize entities, measure properties or detect events. Version 3.4 of the tracker has been designed to be launched and configured remotely using a distributed computing middleware such as the CHIL layered service architecture or the OMiSCID middleware.

The robust tracker implements a cyclic process of recursive estimation. based on a Kalman filter. Tracking provides a number of fundamentally important functions for a perception system. Tracking conserves information about over time, thus provides object constancy. Object constancy assures that a label applied to a blob at time T1 can be used at time T2. Tracking enables the system focus attention, applying the appropriate detection processes only to the region of an image where a target is likely to be detected. Also the information about position and speed provided by tracking can be very important for describing situations.

Tracking is classically composed of four phases: Predict, observe, detect, and update. The prediction phase updates the previously estimated attributes for a set of entities to a value predicted for a specified time. The observation phase applies the prediction to the current data to update the state of each target. The detect phase detects new targets. The update phase updates the list of targets to account for new and lost targets.

To this set of phases, the PRIMA robust tracker adds a recognition phase, an auto-regulation phase, and a communication phase. In the recognition phases, the tracker interprets recognition methods that have been downloaded to the process by a configuration tool. These methods are bits of code that may be expressed in scheme, CLIPS or C++. They are interpreted by a RAVI interpreter and may result in the generation of events or the output to a stream. The auto-regulation phase determines the quality of service metric, such as total cycle time and adapts the list of targets as well as the target parameters to maintain a desired quality. During the communication phase, the supervisor responds to requests from other processes, the PFT or a federation supervisor. These requests may ask for descriptions of process state, or capabilities, or may provide specification of new recognition methods.
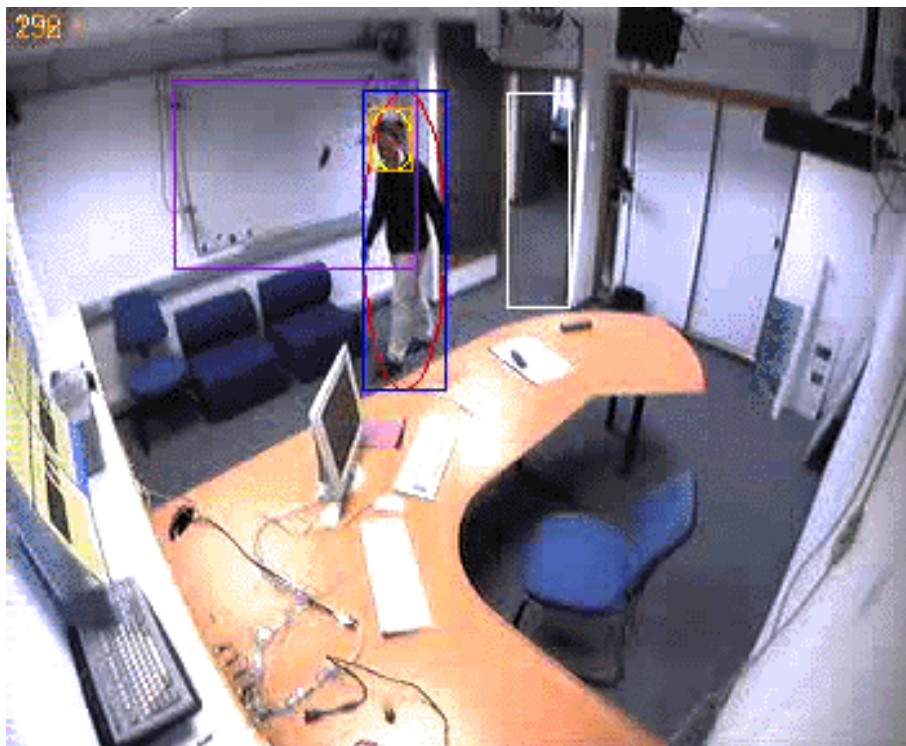
Figure 11. The new programmable robust tracker makes it possible to observe composite entities

The supervisory controller provides four fundamental functions: command interpretation, execution scheduling, parameter regulation, and reflexive description. The supervisor acts as a programmable interpreter, receiving snippets of code script that determine the composition and nature of the process execution cycle and the manner in which the process reacts to events. The supervisor acts as a scheduler, invoking execution of modules in a synchronous manner. The supervisor handles event dispatching to other processes, and reacts to events from other processes. The supervisor regulates module parameters based on the execution results. Auto-critical reports from modules permit the supervisor to dynamically adapt processing. Finally, the supervisor responds to external queries with a description of the current state and capabilities.

Quality of service metrics such as cycle time, number of targets can be maintained by dropping targets based on a priority assignment or by reducing resolution for processing of some targets (for example based on size). Requests are serial messages that arrive from the federation supervisor or from the PFT.

The robust tracker has been declared to the APP under the name 2DBT. This system uses the APTE parameter tuning system filed with the APP "Agence pour la Protection des Programmes" under the Interdeposit Digital number IDDN.FR.001.480025.000.S.P.2005.000.10000.

6.3. Tracking Focus of Attention

Keywords: *Face orientation estimation.*

Participants: Nicolas Gourier, Jérôme Maisonnasse, James Crowley [correspondant].

Project PRIMA has developed a method for estimating the head orientation of previously unseen subjects from images obtained under natural, unconstrained conditions in real time. This method uyses a three-stage approach in which global appearance is first used to provide a low-resolution, coarse estimate of orientation. This coarse estimate is then used as the starting point for a higher-resolution, refined estimate based on local appearance. The high resolution estimation is then used to drive an attentional model based on models of human to human interaction. When applied to Pointing'04 benchmark, this method provides an accuracy of 10° in yaw (pan) angle and 12° in pitch (tilt) angle.

Knowing the head orientation of a person provides information about visual focus of attention. The task of estimating and tracking focus of attention can serve as an important component for systems for man-machine interaction, video conferencing, lecture recording, driver monitoring, video surveillance and meeting analysis. To be useful, such applications require a method that is unobtrusive to avoid distraction. In general, this means estimating orientation of arbitrary subjects from a relatively low resolution imagette, extracted from an image taken from an unconstrained viewing angle under unconstrained illumination. This problem is more difficult than estimating face orientation from high-resolution mug-shot images.

In our system we use a robust video rate face tracker to focus processing on face regions, although any reliable face detection process. Our tracker uses pixel level detection of skin colored regions based on probability density function of chrominance, and provides estimates of the first and second moments of the probability image of skin. From these, we compute an affine transformation that is used to warp the face onto a standard size imagette, while normalising position, width, height and orientation. Experiments have shown imagettes of size 23x30 pixels provide reasonably good input for head pose estimation.

In 2007, software based on this system system has been licensed to the startup company TechnoSens under the names SuiviDeCiblesCouleur and FaceStabilsationSystem. These systems work together to provide automatic video composition for hands-free video communications.

SuiviDeCiblesCouleur locates individuals in a scene for video communications. FaceStabilsationSystem renormalises the position and scale of images to provide a stabilised video stream. SuiviDeCiblesCouleur has been declared with the APP "Agence pour la Protection des Programmes" under the Interdeposit Digital number IDDN.FR.001.370003.000.S.P.2007.000.21000.

7. Contracts and Grants with Industry

7.1. European and National Projects

7.1.1. *ANR Project CASPER: Communication, Activity Analysis and Ambient Assistance for Senior PERsons*

Start Date: 1 Jan 2007

Duration: 36 months

The consortium consists of INRIA Rhone Alpes, France Telecom R&D and H2AD.

The CASPER project will develop new technologies to respond directly to the increasingly urgent social problem of care for elderly persons. A complementary target group of persons with cognitive deficiencies, including persons with early forms of Alzheimer disease and brain-injured people will also be addressed.

We propose to create a family of devices that provide monitoring services as well as a sense of "presence" between elderly people or people with cognitive deficiencies on the one hand, their family or friends, volunteer helpers of health care professionals on the other hand. The proposed family of devices will consist of a "mother station" assisted by a number of satellite devices for specialised sensing services. The mother station and satellite sensors form a wireless sensor network that uses machine-learning techniques to collectively construct and maintain a model of daily activity. The devices will use "ambient" communication technologies (lights, sounds, movements, vibration) to interact in a familiar, simple and unobtrusive way with their users, and to communicate a sense of presence between elderly people, their family and health care service providers.

The devices will embed and combine visual, acoustic and tactile sensor modalities for observing the daily patterns of activity. Mother stations will embed a panoramic camera, microphones, tactile sensor and biometric sensors to note movements, activities and physiological parameters of its owner. The wireless satellite devices will provide specialised sensing to recognize specific classes of domestic activity. For example they may use acoustic sensing to recognize "water sounds" (e.g. shower, bath, basin etc), kitchen sounds (cooking, washing or manipulating food packages), and living room sounds (telephone, TV, Radio or visitors). Incremental machine learning will be used to refine and adapt pre-learned recognition procedures to ambient sounds.

The system will not record images or sound (unless instructed to), but will use embedded real-time machine perception and pattern recognition to recognize classes of activities and organise an activity log. The mother station will maintain a journal of daily activities organised according to location and time, where location will be in the form of topologically associated spatio-temporal map that devices infer from sensor observations. Activity will be recorded as a series of situations organised using 24 hour and 7 day cycles. The activity log can be used to provide un-intrusive "presence" information for family members, as well as more detailed records for health care professionals. Comparison of current activity to a statistical summary of the activity log can be used to sense unusual situations requiring intervention.

The result of this project will be a new technological approach to assist elderly and brain-injured people in maintaining independent living and social interaction. The overall aim of this project is to demonstrate the required technology, and to understand its user requirements and its social implications.

The project will make key contributions to progress beyond the state of the art. These include:

- A new approach to provision of care, assistance, social interaction and social connectedness to older people living independently, combining traditional monitoring with high-level detection of situations, activities, and activity patterns

- Advances in perceptual technologies, affective computing, ambient system design, and sensor networks that will be generic and relevant beyond the use for healthcare monitoring.

The impact of the proposed technologies goes beyond the problem of independent living, health and well being of the elderly. These technologies can stimulate emergence of new commercial industries in a number of areas. For example, affective interaction is expected to provide a major impact on the way that people interact with information technology. Such technology may mark a rupture point in the evolution of informatics and the start in an exponential growth in new applications across a broad spectrum.

7.1.2. *OSEO Project MinImage: Embedded Integrated Vision Systems*

Start Date: 1 march 2008

Duration: 60 months

The consortium consists of - STMicroelectronics - Saint-Gobain Recherche - CEA-LETI and LIST - Varioptic - INRIA Grenoble Rhone-Alpes Research Centre - DxO

The goal of the MinImage project is to develop integrated micro-cameras for portable telephones. This is a 141 Million Euro development program provided with 70 Million Euros of Aide by OSEO/AII. The program includes major development efforts in micro-electronics, optics, image processing, and image analysis.

8. Dissemination

8.1. Contribution to the Scientific Community

8.1.1. *Participation on Conference Program Committees*

James L. Crowley served as program co-chair for the European Conference on Ambient Intelligence, AMI '08 in Nuremberg Germany, 19 - 22 November 2008.

James L. Crowley will serves as conference co-chairman for the 2009 ACM conference on MultiModal Interaction to be held at MIT Media Lab in Cambridge Massachusets in Oct. 2009.

James L. Crowley serves as programme committee area-chair for the 2009 IEEE International Conference on Robotics and Automation (ICRA2009) in Kobe, Japan, May 12 - 17, 2009.

James L. Crowley has served as a member of the program committee for the following conferences.

- AMI 2008, European Conference on Ambient Intelligence, Darnstadt, Nov 2008
- ICMI 2008, International Conference on Multimodal Intelligence, Chania, Greece, Oct 2008
- ECAI 2008, European Conference on Artificial Intelligence, July 2008.
- ECOOP 2008, European Conference on Object Oriented Programming, July 2008.
- CVPR 2008, IEEE Conf. on Computer Vision and Pattern Recognition, June 2008
- ICVS 2008, 6th International Conference on Computer Vision Systems, May 2008.
- ICRA 2008, IEEE International Conference on Robotics and Automation, april 2008

8.1.2. *Participation on Advisory Panels*

- James L. Crowley served as Rapporteur for the FET (Future and Emerging Technologies) unit of the EC ICT program "Human Computer Confluence" panel. The panel has defined the scope and objectives for an FET Proactive Research initiative on Human Computer Confluence published in 2009 as part of FP7 ICT program.
- James L. Crowley hasq served as co-chairman (with Joelle Coutaz) of the DGRI A3/GCS3 working group on Intelligence ambiante, organised by the French MESR (Ministère de l'enseignement supérieur et de la Recherche).
- James L. Crowley participates as member of the Minister of Research panel SNRI: National Research and Innovation Strategy, subpanel "Quality of Life". The SNRI panel is tasked to prepare a research plan to govern French national research for the period 2010 through 2014

8.1.3. Invited Presentations by James L. Crowley

1. "Perception of Human Activity for Augmented Environments", SIBGRAPI 2008, XXI Brazilian Symposium on Computer Graphics and Image Processing, Campo Grande, Brazil, Oct 2008.

2. "Intelligence equals Embodied, Autonomous and Situated", Canadian Vision and Robotics Conference, Windsor, July 2008.

8.2. Teaching

1. Augustin Lux and James Crowley co-direct the joint INPG/UJF Informatics Research Masters Degree.

2. Augustin Lux is correspondant to the INPG Doctoral College for the Informatics and Applied Mathematics.

3. James Crowley co-direct the Master of Science in Informatics at Grenoble.

4. Patrick Reignier directs the Licence Professionnel of Web Development

9. Bibliography

Major publications by the team in recent years

[1] S. BORKOWSKI, J. LETESSIER, F. BÉRARD, J. L. CROWLEY. *User centric design of a vision system for interactive applications*, in "2006 IEEE International Conference on Computer Vision Systems, ICVS'06, New York", January 2006, p. 9-16.

[2] S. BORKOWSKI, J. LETESSIER, J. L. CROWLEY. *Spatial Control of Interactive Surfaces in an Augmented Environment*, in "EHCI '04, Engineering Human Computer Interaction and Interactive Systems, New York", july 2004, p. 228-244.

[3] J. COUTAZ, J. L. CROWLEY, S. DOBSON, D. GARLAN. *Context is key*, in "Communications of the ACM", vol. 48, n° 3, March 2005, p. 49-53.

[4] J. L. CROWLEY, O. BRDICZKA, P. REIGNIER. *Learning Situation Models for Understanding Activity*, in "The 5th International Conference on Development and Learning 2006 (ICDL06), Bloomington, Il., USA", June 2006.

[5] J. L. CROWLEY, J. COUTAZ, F. BÉRARD. *Things that See: Machine Perception for Human Computer Interaction*, in "Communications of the A.C.M.", vol. 43, n° 3, March 2000, p. 54-64.

[6] J. L. CROWLEY, J. COUTAZ, G. REY, P. REIGNIER. *Using Context to Structure Perceptual Processes for Observing Activity*, in "UBICOMP, Sweden", September 2002.

[7] D. HALL, F. PÉLISSON, O. RIFF, J. L. CROWLEY. *Brand Identification Using Gaussian Derivative Histograms*, in "Machine Vision and Applications", vol. 16, n° 1, 2004, p. 41-46.

[8] C. LE GAL, J. MARTIN, A. LUX, J. L. CROWLEY. *Smart Office: An Intelligent Interactive Environment*, in "IEEE Intelligent Systems", July/August 2001.

[9] A. LUX. *The Imalab Method for Vision Systems*, in "Machine Vision and Applications (MVA)", vol. 16, 2004, p. 21–26.

[10] B. SCHIELE, J. L. CROWLEY. *Recognition without Correspondence using Multidimensional Receptive Field Histograms*, in "International Journal of Computer Vision", vol. 36, n° 1, January 2000, p. 31–50.

[11] K. SCHWERDT, J. L. CROWLEY. *Robust Face Tracking using Color*, in "International Conference on Automatic Face and Gesture Recognition, Grenoble, France", March 2000, p. 90–95.

Year Publications

Articles in International Peer-Reviewed Journal

[12] O. BRDICZKA, J. L. CROWLEY, P. REIGNIER. *Learning situation models for providing context-aware services*, in "IEEE Transactions on Man, Systems and Cybernetics, Part B", to appear, 2008.

[13] O. BRDICZKA, M. LANGET, J. MAISONNASSE, J. L. CROWLEY. *Detecting human behavior models from multimodal observation in a smart home*, in "In IEEE Transactions on Automation Science and Engineering", to appear, 2008.

[14] S. ZAIDENBERG, P. REIGNIER, J. L. CROWLEY. *An Architecture for Ubiquitous Applications*, in "Ubiquitous Computing and Communication Journal (UBiCC)", to appear, vol. Special Issue IWUC07, 2008, http://www-prima.imag.fr/prima/pub/Publications/2008/ZRC08b/.

International Peer-Reviewed Conference/Proceedings

[15] R. EMONET, D. VAUFREYDAZ. *Usable developer-oriented Functionality Composition Language (UFCL): a Proposal for Semantic Description and Dynamic Composition of Services and Service Factories*, in "4th IET International Conference on Intelligent Environments", 2008.

[16] J. A. RUIZ HERNANDEZ, A. LUX, J. L. CROWLEY. *Face Detection by Cascade of Gaussian Derivatives Classifiers Calculated with a Half-Octave Pyramid*, in "IEEE Conference on Automatic Face and Gesture Recognition", Amsterdam, Sep 2008.

[17] S. ZAIDENBERG, P. REIGNIER, J. L. CROWLEY. *Reinforced learning of context models for ubiquitous computing: a ubiquitous personal assistant*, in "Proceedings of the 6th ICEIS Doctoral Consortium - DCEIS 2008", J. CARDOSO (editor), vol. 6, INSTICC Press, june 2008, p. 36–48.

Scientific Books (or Scientific Book chapters)

[18] O. BRDICZKA, J. L. CROWLEY, J. CURIN, J. KLEINDIENST. *Situation modeling*, in "Computers in the Human Interaction Loop", n° 12, Springer, 2008, p. 121-132.

[19] J. KLEINDIENST, J. CURIN, O. BRDICZKA, N. DIMAKIS. *Situation modeling layer*, in "Computers in the Human Interaction Loop", vol. 26, Springer, 2008, p. 317-326.

[20] S. ZAIDENBERG, P. REIGNIER, J. L. CROWLEY. *Reinforcement Learning of Context Models for a Ubiquitous Personal Assistant*, Advances in Soft Computing, vol. Volume 51/2009, Springer Berlin / Heidelberg, september 2008, p. 254–264.

References in notes

[21] *DNS-SD : DNS Service Discovery*, http://www.dns-sd.org.

[22] G. BELL. *Bell's law for the birth and death of computer classes*, in "Communications of the ACM", vol. 51, n° 1, 2008, p. 86–94.

[23] S. BORKOWSKI, O. RIFF, J. L. CROWLEY. *Projecting Rectified Images in an Augmented Environment*, in "PROCAMS'03 Workshop", 2003.

[24] F. BÉRARD. *The magic table: computer-vision based augmentation of a whiteboard for creative meetings*, in "Workshop on Projector-Camera Systems", 2003.

[25] F. BÉRARD. *The Magic Table: Computer-Vision Based Augmentation of a Whiteboard for Creative Meetings*, in "Proceedings of the ICCV Workshop on Projector-Camera Systems", IEEE Computer Society Press, 2003.

[26] V. COLIN DE VERDIÈRE, J. L. CROWLEY. *Visual Recognition using Local Appearance*, in "ECCV98, Freiburg", June 1998, p. 640–654.

[27] J. L. CROWLEY, J. COUTAZ, G. REY, P. REIGNIER. *Using Context to Structure Perceptual Processes for Observing Activity*, in "UBICOMP, Sweden", September 2002.

[28] J. L. CROWLEY, P. REIGNIER. *An Architecture for Context Aware Observation of Human Activity*, in "Workshop on Computer Vision System Control Architectures (VSCA 2003)", 2003.

[29] J. L. CROWLEY, O. RIFF. *Fast Computation of Scale Normalised Gaussian Receptive Fields*, in "International Conference on Scalespace theories in Computer vision, Skye, UK", June 2003, p. 584–598.

[30] W. FREEMAN, E. ADELSON. *The Design and Use of Steerable Filters*, in "Pattern Analysis and Machine Intelligence", vol. 13, n° 9, September 1991, p. 891–906.

[31] D. GARLAN, S. CHENG, A. HUANG, B. SCHMERL, P. STEENKISTE. *Rainbow: Architecture-based, self-adaptation with reusable infrastructure*, in "IEEE Computer", October 2004, p. 46-54.

[32] N. GOURIER, D. HALL, J. L. CROWLEY. *Facial feature detection robust to pose, illumination, and identity*, in "International Conference on Systems, Man and Cybernetics, Special track on Automatic Facial Expression Analysis", October 2004, p. 617-622.

[33] D. HALL, J. L. CROWLEY. *Détection du visage par caractéristiques génériques calculées à partir des images de luminance*, in "Reconnaissance des formes et intelligence artificielle, Toulouse, France", to appear, 2004.

[34] D. HALL, V. COLIN DE VERDIÈRE, J. L. CROWLEY. *Object Recognition using Coloured Receptive Fields*, in "European Conference on Computer Vision, Dublin, Ireland", June 2000, p. I 164–177.

[35] P. HORN. *Autonomic Computing; IBM's Perspective on the state of information technology*, October 2001, http://researchweb.watson.ibm.com/autonomic.

[36] B. JOHANSON, G. HUTCHINS, T. WINOGRAD, M. STONE. *PointRight: Experience with Flexible Input Redirection in Interactive Workspaces*, in "Proceedings of UIST-2002", 2002.

[37] P. JOHNSON-LAIRD. *How we reason*, Oxford University Press, 2006.

[38] J. KEPHART, D. CHESS. *The vision of autonomic computing*, in "IEEE Computer", vol. 36(1), 2003, p. 41–50.

[39] R. KJELDSEN, A. LEVAS, C. S. PINHANEZ. *Dynamically reconfigurable visionbased user interfaces*, in "2003 International Conference on Vision Systems, ICVS03, Graz, Austria", April 2003, p. 257-267.

[40] J. J. KOENDERINK, A. J. VAN DOORN. *Representation of Local Geometry in the Visual System*, in "Biological Cybernetics", n° 55, 1987, p. 367-375.

[41] T. LINDEBERG. *Feature Detection with Automatic Scale Selection*, in "International Journal of Computer Vision", vol. 30, n° 2, 1998, p. 79–116.

[42] D. LOWE. *Object Recognition from Local Scale-Invariant Features*, in "ICCV, Corfu, Greece", September 1999, p. 1150–1157.

[43] A. LUX. *The Imalab Method for Vision Systems*, in "ICVS03, Graz, Austria", April 2003.

[44] A. NEGRE, C. BRAILLON, J. L. CROWLEY. *Visual navigation from variation of intrinsic scale*, in "under review", 2006.

[45] G. PINGALI, C. S. PINHANEZ, A. LEVAS, R. KJELDSEN. *Steerable Interfaces for pervasive computing spaces*, in "IEEE PerCom", March 2003.

[46] C. S. PINHANEZ. *The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces*, in "Proceedings of Ubiquitous Computing 2001 Conference", September 2001.

[47] V. POLADIAN. *Dynamic Configuration of Resource-aware services*, in "Int. Conf. on Software Engineering", September 2001.

[48] F. PÉLISSON, D. HALL, O. RIFF, J. L. CROWLEY. *Brand identification using Gaussian derivative histograms*, in "International Conference on Vision Systems, Graz, Austria", April 2003, p. 492–501.

[49] R. RASKAR, G. WELCH, M. CUTTS, A. LAKE, L. STESIN, H. FUCHS. *The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays*, in "Proceedings of the ACM SIGGRAPH'98 Conference", 1998.

[50] J. RASURE, S. KUBICA. *The Khoros Application Development Environment*, in "Experimental Environments for Computer Vision and Image Processing", J. L. CROWLEY, H. CHRISTENSEN (editors), Machine Perception Artificial Intelligence Series, vol. 11, n° 1, World Scientific Press, 1994, p. 1-32.

[51] B. SCHIELE, J. L. CROWLEY. *Recognition without Correspondence using Multidimensional Receptive Field Histograms*, in "International Journal of Computer Vision", vol. 36, n° 1, January 2000, p. 31–50.

[52] N. A. STREITZ, J. GEISSLER, T. HOLMER, S. KONOMI, C. MÜLLER-TOMFELDE, W. REISCHL, P. REXROTH, P. SEITZ, R. STEINMETZ. *i-LAND: An interactive Landscape for Creativitiy and Innovation*, in "ACM Conference on Human Factors in Computing Systems", 1999.

[53] N. TAKAO, J. SHI, S. BAKER. *Tele Graffiti: a camera projector based remote sketching system with hand based user interface and automatic session summarization*, in "Int. Journal of Computer Vision", vol. 53, 2003, p. 115-133.

[54] T. T. H. TRAN, A. LUX. *A method for ridge extraction*, in "Asian Conference on Computer Vision, Jeju, Korea", 2004, p. 960–966.

[55] T. T. H. TRAN. *Etude de lignes d'interet naturelles pour la representation d'objets en vision par ordinateur*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, March 2006.

[56] T. T. H. TRAN, A. LUX, H. L. NGUYEN THI. *Towards a ridge and peak base symbolic representation for object recognition*, in "The 3rd International Conference in Computer Science, Can Tho University, VietNam", 21-24 Feb 2005.

[57] F. VERNIER, N. LESH, C. SHEN. *Visualization Techniques for Circular Tabletop Interfaces*, in "Advanced Visual Interfaces", 2002.

[58] S. VOIDA, E. MYNATT, B. MACINTYRE, G. CORSO. *Integrating virtual and physical context to support knowledge workers*, in "Proceedings of Pervasive Computing Conference", IEEE Computer Society Press, 2002.

[59] P. WELLNER. *The DigitalDesk Calculator: Tactile Manipulation on a Desk Top Display*, in "ACM Symposium on User Interface Software and Technology (UIST '91)", November 1991, p. 27–33.

[60] G. YE, J. CORSO, D. BURSCHKA, G. D. HAGER. *Vics: A modular vision-based HCI framework*, in "2003 International Conference on Vision Systems - ICVS03, Graz, Austria", April 2003, p. 257-267.

[61] B. ZOPPIS. *Outils pour l'Intégration et le Contrôle en Vision et Robotique Mobile*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, June 1997.