INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team PRIMA

Perception, recognition and integration for interactive environments

Rhône-Alpes

THEME COG

Activity Report

2006

Table of contents

1. Team

Head of the team
James CROWLEY [Professeur CE INPG, HdR]

Assistante de projet
Caroline OUARI [Secretary (SAR) Inria]

Enseignants
Augustin LUX [Professeur INPG, HdR]
Patrick REIGNIER [MdC Universite Joseph Fourier]
Dominique VAUFREYDAZ [MdC Universite Pierre Mendes France]

Post Doc
Nicolas GOURIER [PostDoc INRIA]
Daneila HALL [PostDoc INPG through March 2006]

Engineers
Matthieu LANGET [Engineer CDD INPG]
Alba FERRER-BIOSCA [Engineer CDD INRIA]
Jean-Marie VALLET [Engineer CDD CNRS]

Doctoral Students
Olivier BERTRAND [Doctoral Student (ENS Cachan)]
Oliver BRDICZKA [ATER UPMF]
Suphot CHUNWIPHOT [Doctoral Student (Thailand)]
Rémi EMONET [Doctoral Student (Bourse MENSR)]
Julien LETESSIER [Doctoral Student (CDD UJF)]
Jérôme MAISONNASSE [Doctoral Student (CDD INPG)]
Sofia ZAIDENBERG [Doctoral Student (BDI CNRS)]
Matthieu ANNE [Doctoral Student (CIFRE France Telecom)]
Stansilaw BORKOWSKI [Doctoral Student (Egide) through June 2006]

Visitor
Marina PETTINARI [Doctoral Student Univ Bologne (Bourse Marie-Curie)]

Master Students
Jean-Pascal MERCIER [M2R IVR]
Rémi BARRAQUAND [M2R IVR]
John Alexander RUIZ HERNANDEZ [M2R IVR]

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 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 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 develops and employs machine perception techniques using acoustics, speech, computer vision and mechanical sensors.

An environment is said to be "active" when it is capable of changing its internal state. Trivial forms of state change include regulating ambient temperature and illumination. Automatic presentation of information and communication constitutes a challenging new class of actions with many practical applications. The use of multiple display surfaces coupled with location awareness of occupants offers the possibility of automatically adapting presentation 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 and interaction with information.

An environment may be considered as "interactive" when it is capable responding 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 or objects placed into projected interaction widgets. Richer forms of interaction require perceiving and modeling of the current task of users. 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 conceptual framework and theoretical foundation for context aware perception.
- The design of robust vision systems based on local appearance,
- A software architecture model for reactive control of multi-modal perceptual systems.

The experiments in project PRIMA are oriented towards perception of human activity. The project is particularly concerned with modeling the interaction between communicating individuals in order to provide video-conferencing and information services. Application domains include context aware video communications, new forms of man-machine interaction, visual surveillance, and new forms of information services and entertainment.

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. However, 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. 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.

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.

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.

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.

In collaboration with France Telecom, in 2006 we have adapted this technology to observing social activity in domestic environments. Our goal is to demonstrate new forms of context aware services for human to human communication, as well as monitoring services to allow the elderly to live independently without compromising access to immediate medical care or 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 [58]. 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 [53].

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, Auto-regulation, Auto-description, Auto-Monitoring and Auto-configuration. 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 [60]. 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 [53], [60].

Self-monitoring and self-regulating perceptual components are a typical of a new approach to system design known as "autonomic computing". Autonomic computing has recently emerged as an effort inspired by biological systems to render computing systems robust [58]. 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. In collaboration with the software engineering group of Walter Tichy at University of Karlsruhe, 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.

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

Auto-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.

Auto-configuration: The component reconfigures its own modules so as to respond to changes in the operating environment or quality requirements [70].

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 [49], [50], 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.

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

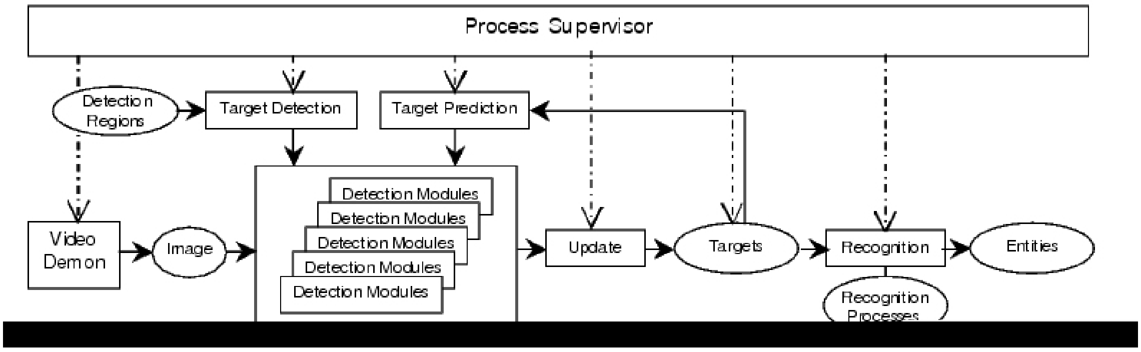

Figure 1. Architecture for an autonomic perceptual component

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 [65] 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.

3.3. Robust view-invariant Computer Vision

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

3.3.1. 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 during the period 1987 to 2002. The group is currently working on several innovations based on chromatic receptive fields and scale invariant ridges.

3.3.2. Detailed Description

The visual appearance of a neighbourhood can be described by a local Taylor series [62]. 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 [63] 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 [75]. 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 [63], and normalized in orientation using direction of the first derivatives [75].

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" [64]. 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" [52] of Gaussian derivatives. The problem is to determine the appropriate orientation. In earlier work by PRIMA members Colin de Verdiere [48], Schiele [75] and Hall [57], 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 [57]. 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.

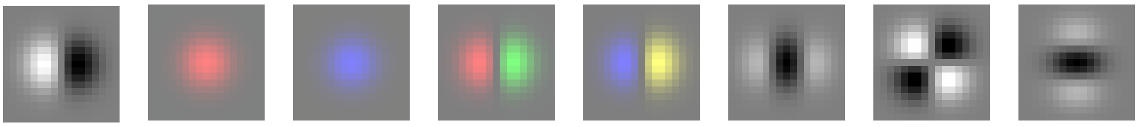

Figure 2. Chromatic Gaussian Receptive Fields ($G_x^L, G^{C_1}, G^{C_2}, G_x^{C_1}, G_x^{C_2}, G_{xx}^L, G_{xy}^L, G_{yy}^L$).

Key results in this area include

1. Fast, video rate, calculation of scale and orientation for image description with normalized chromatic receptive fields [51].
2. Real time indexing and recognition using a novel indexing tree to represent multi-dimensional receptive field histograms [71].
3. Robust visual features for face tracking [55], [54].
4. Affine invariant detection and tracking using natural interest lines [78].
5. Direct computation of time to collision over the entire visual field using rate of change of intrinsic scale [67].

We have recently achieved video rate calculation of intrinsic (characteristic) scale from interpolation within a Binomial Pyramid computed using an O(N) algorithm [51]. 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.

Fabien Pelisson has demonstrated real time indexing and recognition using a novel indexing tree to represent multi-dimensional receptive field histograms [71]. This system has been used for content based indexing in very large image data bases. It has also been used for appearance based recognition of objects and people for video surveillance and for detecing publicity panels [71], [56].

Daniela Hall and Nicolas Gourier have developed machine learning techniques to statistically learn robust visual features for face tracking [55], [54].

Lux and Hai have recently developed a method with provides a direct measurement of affine invariant local features based on extending natural interest points to "natural interest ridges" [80], [79]. 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.

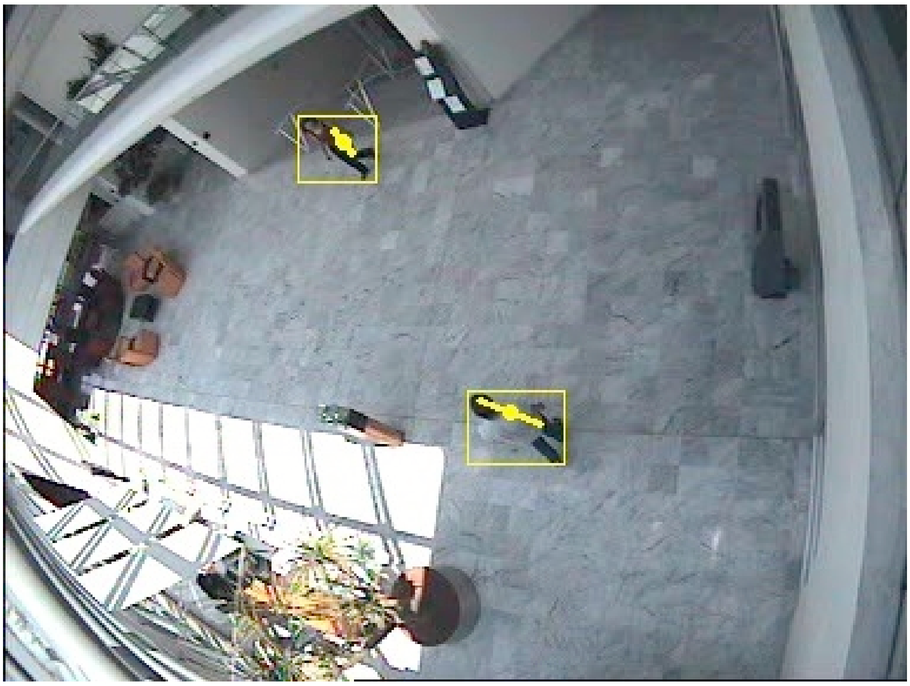

Figure 3. Dominant natural interest ridges for tracking people

Aumaury Negre has recently demonstrated direct computation of time to collision over the entire visual field using rate of change of intrinsic scale [67]. This approach is currently being adapted for use in visual navigation in joint work with project EMOTION.

3.4. 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 [69]. The original functionality of the objects does not change, only its appearance. An example of object enhancement is presented in [47], 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 [45] are increasingly used in augmented environment systems [68] [73] as an inexpensive means of making projected images interactive. Steerable projectors [45] [69] 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 [82]).

The PRIMA group has recently 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 [84], the Magic Table [47] or the Tele-Graffiti application [77].

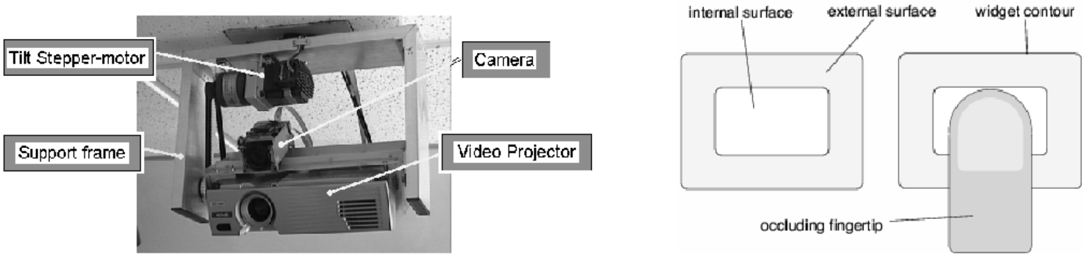

Figure 4. Steerable camera-projector pair (left) and surfaces defined to detect touch-like gestures over a widget (right)

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 posible 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, [45] 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 [61]. 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 [85].

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. Most notably, we are currently seeking to commercialise a projector-camera system for use in advertising in store windows and trade shows.

4. Application Domains

4.1. The Augmented Meeting Environment

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

Participants: Patrick Reignier, Dominique Vaufreydaz, Augustin Lux, Alba Ferrer-Biosca, Rémi Emonet, Matthieu Langet, Jerome Maisonnasse, Oliver Brdiczka, Sonia Zaidenberg, Nicolas Gourier, James L. Crowley.

In order to test and develop systems for observation of human activity, Project PRIMA has constructed an "Augmented Meeting Environment", show in figure 5. The PRIMA Augmented Meeting Environment is equipped with a microphone array, wireless lapel microphones, a wide angle surveillance camera, a panoramic camera, six steerable cameras, 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.

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. The Steerable Camera Projector

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

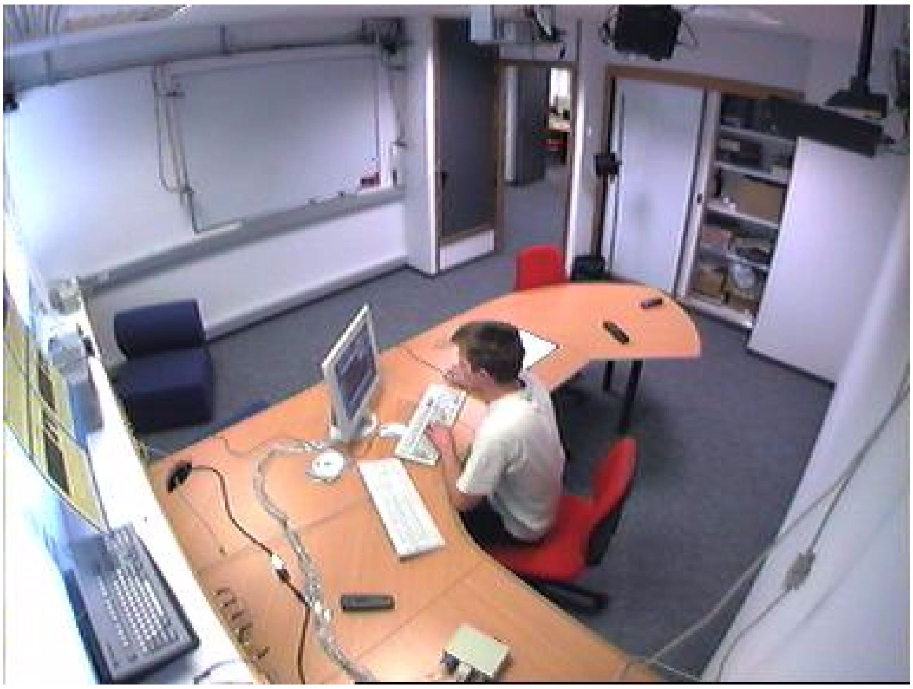

Figure 5. The augmented meeting environment is an office environment equipped with a microphone array, wireless lapel microphones, a wide angle surveillance camera, a panoramic camera, six steerable cameras, and two camera-projector video-interaction devices.

Participants: Stan Borkowski, Julien Letissier, James L. 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 [45]. 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 [59], [76], [83]. 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.

Video-projectors are increasingly used in augmented environment systems [73], [81]. Projecting images is a simple way of augmenting everyday objects and offers the possibility to change their appearance or their function. However, standard video-projectors have a fairly small projection area which significantly limits their spatial flexibility as output devices in an pervasive system. A certain degree of steerability can be achieved for a rigidly mounted projector: In particular, a sub window can be steered within the cone of projection for a fixed projector [82]. However, extending and/or moving the display surface requires augmenting the range of angles to which the projector beam may be directed. If using fixed projectors, this means increasing the number of projectors which is relatively expensive. A natural solution is to use a Steerable projector-camera assembly [66] and [69]. With a trend towards increasingly small and inexpensive video projectors and cameras, this approach will become increasingly attractive. Additionally having the ability to modify the scene with projected light, projector-camera systems can be exploited as sensors, thus enabling to collect data that can be used to build a model of the environment.

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 [69] 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 [46], where users could interact with both physical and virtual ink on an projection-augmented whiteboard.

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. The mechanical performance of the SCP is presented in Table 1. While somewhat bulky, our device anticipates the current trend of projectors to become portable devices, similar in shape to hand-held torch lamps [72].

Note that the SCP is not only a motorized video-projector, but a projector-camera pair. The camera is mounted in such a way that the projected beam overlaps with the camera-view. Equipping an SCP with a camera offers a number of interesting possibilities. User's actions can be observed within the field of view of the camera and interpreted as input information for the computer system. Additionally the system is able to provide visual

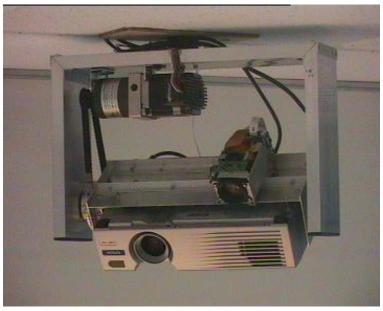

Figure 6. The Steerable Camera Projector

Table 1. Rotation platform mechanical performance

	Pan	Tilt
Rotation range	±177°	+90°
Angular resolution	0.11°	0.18°
Angular velocity	$146\frac{deg}{s}$	$80\frac{deg}{s}$
Response time	$\sim 2ms$	$\sim 3ms$

feedback in response to users action. In other words association of a camera to a projector creates a powerful actuator-sensor pair.

The SCP can be used as a steerable structured light sensor to automatically discover surfaces that are suitable for interaction. Figure 7 shows automatically discovered planar surfaces within the AME. described below.

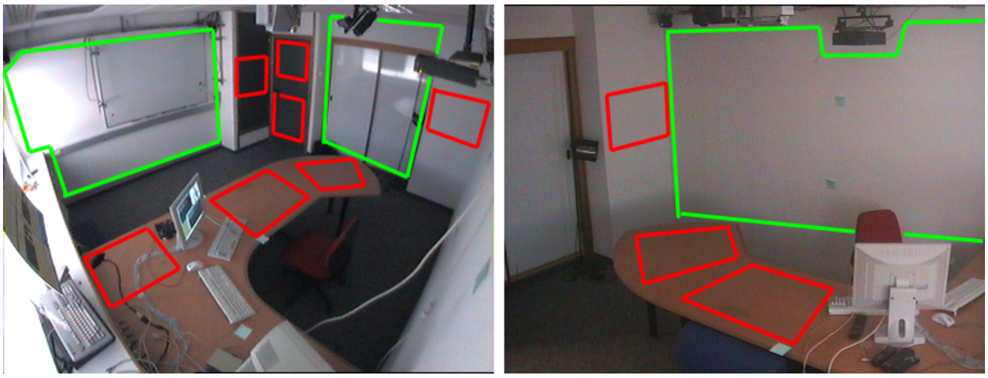

Figure 7. Planar surfaces in the environment

4.3. Context Aware Video Acquisition

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

Participants: Patrick Reignier, Dominique Vaufreydaz, Matthieu Langet, Jerome Maisonnasse, Oliver Brdiczka, Sonia Zaidenberg, Nicolas Gourier, Alba Ferrer-Biosca, James L. Crowley.

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 [49]. 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.

In 2004, we have demonstrated a number of reals 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.

5. Software

5.1. IMALAB

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

Participants: Augustin Lux, Dominique Vaufreydaz, Rémi Emonet.

The Imalab system represents a longstanding effort within the Prima team (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, it 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 [74] 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 [86]. 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. O3MiSCID Middleware for Distributed Multi-Modal Perception

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

Participants: Patrick Reignier, Dominique Vaufreydaz, Rémi Emonet.

O3MiSCID 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* [44]). 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. O3MiSCID 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.

5.3. CAR: Robust Real-Time Detection and Tracking

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

Participants: James L. Crowley, Justus Piater, Stephane Richetto.

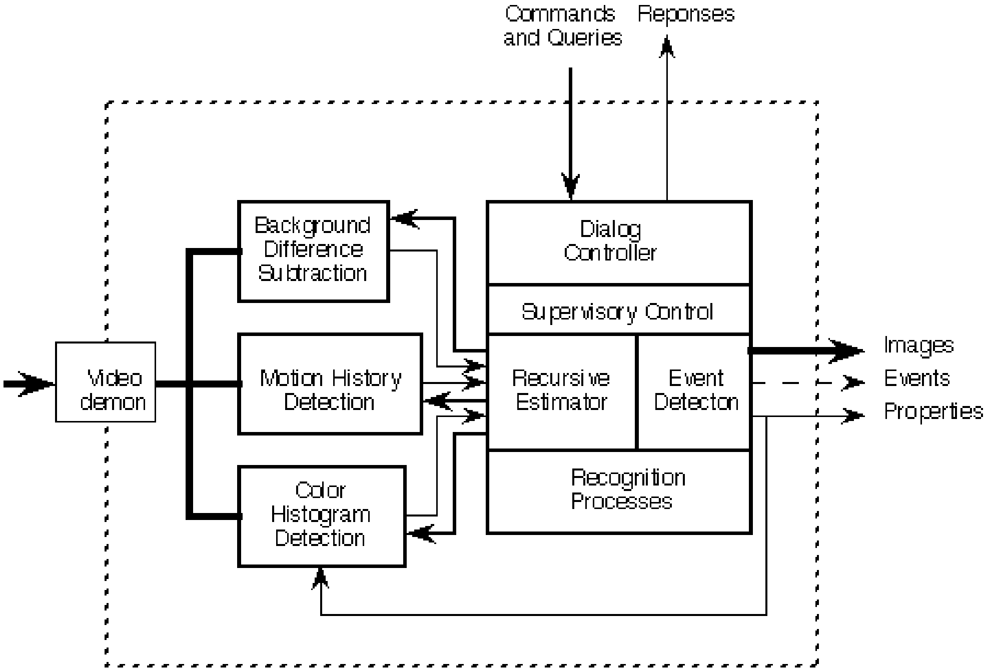

Figure 8. The CAR systems integrates several detection modules with a Kalman Filter for robust detection and tracking of entities

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 implemented a robust real time detection and tracking system (CAR). This system is 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 module architecture permits additional detection modes to be integrated into the process. A process supervisor adapts the parameters of tracking so as to minimize lost targets and to maintain real time response.

Individuals are tracked using a recursive estimation process. Predicted position and spatial extent are used to recalculate estimates for position and size using the first and second moments. Detection confidence is based on the detection energy. Tracking confidence is based on a confidence factor maintained for each target.

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: Daniela Hall, 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 microphone is 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.

Rights to this system are jointly owned by INP Grenoble, UJF and INRIA. The system is currently undergoing Deposit at APP and will shortly be available for licensing. We are investigating plans to create a start up to commercially exploit this system.

5.5. APTE: Automatic Parameter Tuning and Error Recovery

Keywords: *Monitoring*, *Process optimization*, *Self-adaptive system*, *Video Surveillance*.

Participants: Daniela Hall, Rémi Emonet.

Perceptual systems observe the world, interpret the observations in form of input signals such as images, audio data, or laser range data and communicates the result as numeric, vectorial or symbolic events. There is a wide range of different perceptual systems such as tracking systems in video surveillance, expert systems in traffic control or image segmentation systems. Stability, reliability and robustness are required for perceptual systems to be exploited widely in commercial applications. To meet these constraints, developers commonly design simple systems whose parameters can be adapted manually. Such systems perform well as long as the environment stays constant. Unfortunately, in most real applications the environmental conditions perceived by the sensors frequently change, which often breaks the system and requires reinitialisation and new hand tuning of the parameters. This software provides a solution to this problem by enabling a system to automatically adapt its parameters to the environmental changes that would degrade the system performance.

Some error types can not be solved by parameter tuning. Examples are local illumination changes and person fragmentation. To cope with such kinds of errors, the software contains a system for automatic error detection, error classification and error repair. Error detection is performed by evaluating a probabilistic measure with respect to a model of “correct” system output. Classification requires an (incremental) acquisition of error classes. This acquisition requires a minimum amount of human supervision. The system provides a GUI that allows to manually process previously unclassified errors. The knowledge is incrementally incorporated in the system’s knowledge base. First tests showed that manual interaction of 15 min gives a usable error classification model.

This software has been 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. New Results

6.1. 3D Bayesian Tracker

Keywords: *Multi-modal tracking*.

Participants: Alba Ferrer-Biosca, Augustin Lux, Rémi Emonet, James L. Crowley.

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.

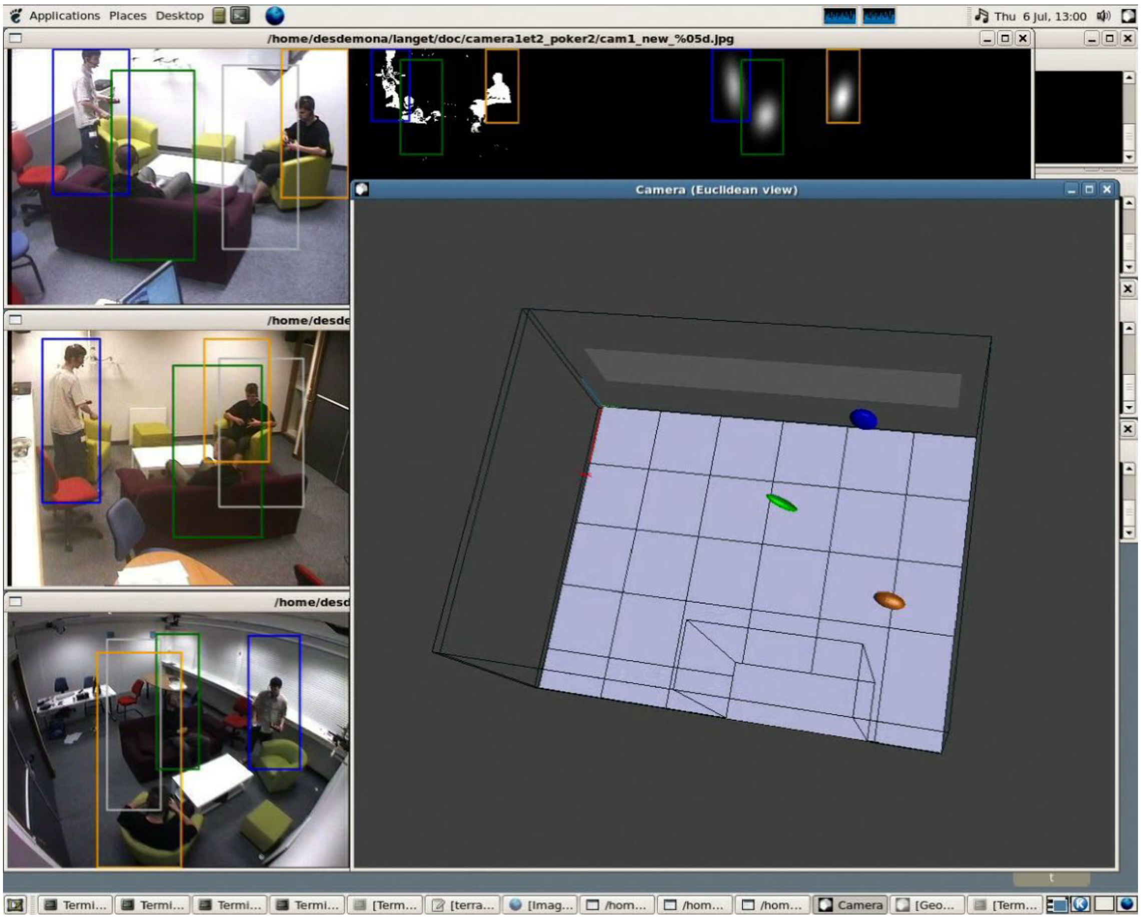

Figure 9. The 3D Bayesian tracker integrates observations from multiple sensors

The 3D body tracker is a Bayesian estimator. It maintains a list of 3D targets in the form of a position and 2nd moment matrix. For each target, the Gaussian blob makes it possible to calculate a 3D region of interest. This 3D region of interest is projected onto each of the current images using the camera calibration matrix, to obtain a 2D "Region of Interest" (or ROI). For each 2D ROI, a detection process is run to locate the current target position in image coordinates that is to be used to update the 3D target locations. Updating operates as follows: For each target, each voxel in the 3D ROI is projected into the detection image from each of the cameras, and likelihood of a target voxel is updated by combining the current voxel likelihood with the detected likelihood in each of the images using Bayesian estimation. The parameters of the 3D Gaussian blob is then recalculated from the updated voxel likelihoods.

In order to operate properly, the 3D tracker requires access to the current image from at least 2 cameras that have been calibrated to a common reference frame. In general, adding additional cameras improves both the precision and the reliability of tracking. Cameras may be added or removed from the system while it is operating provided that at any time at least 2 cameras are available. The 3D tracker does not currently contain control logic to search for new cameras. The action of assigning new cameras to the tracker must be taken from outside the system, either by an automatic process (service supervisor) or by a human programmer.

The 3D tracker produces an event stream containing the target ID (index number), 3D position, and 3D spatial extent (covariance) for each target in real time. The output stream can be fed directly to the situation model, or alternatively can be combined with other 3D estimation and tracking processes (such as acoustic localization) to produce a multi-modal target position from situation modeling.

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.

6.2. Robust Real-Time Detection and Tracking System

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

Participants: James L. Crowley, Alba Ferrer-Biosca, Daniela Hall, Patrick Reignier.

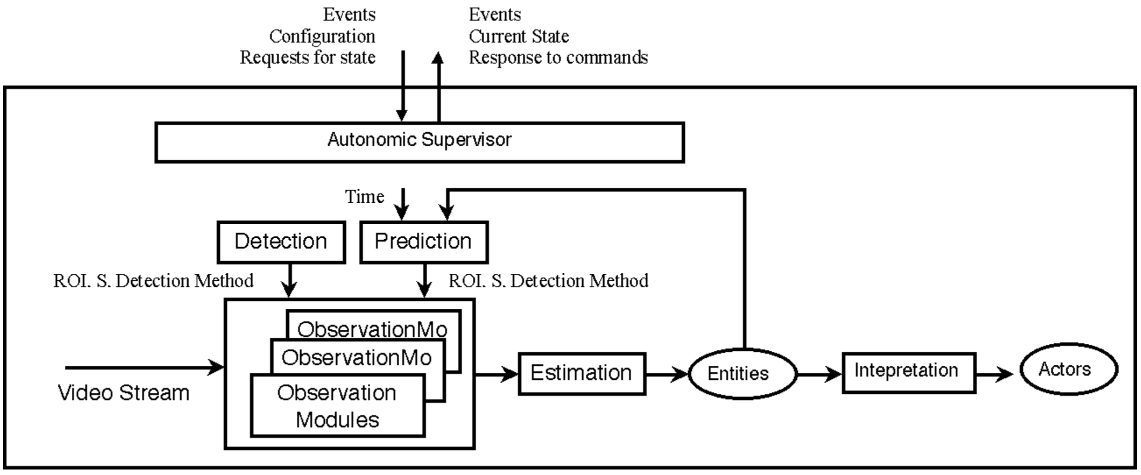

Figure 10. The components and architecture for the new agent detection and tracking process.

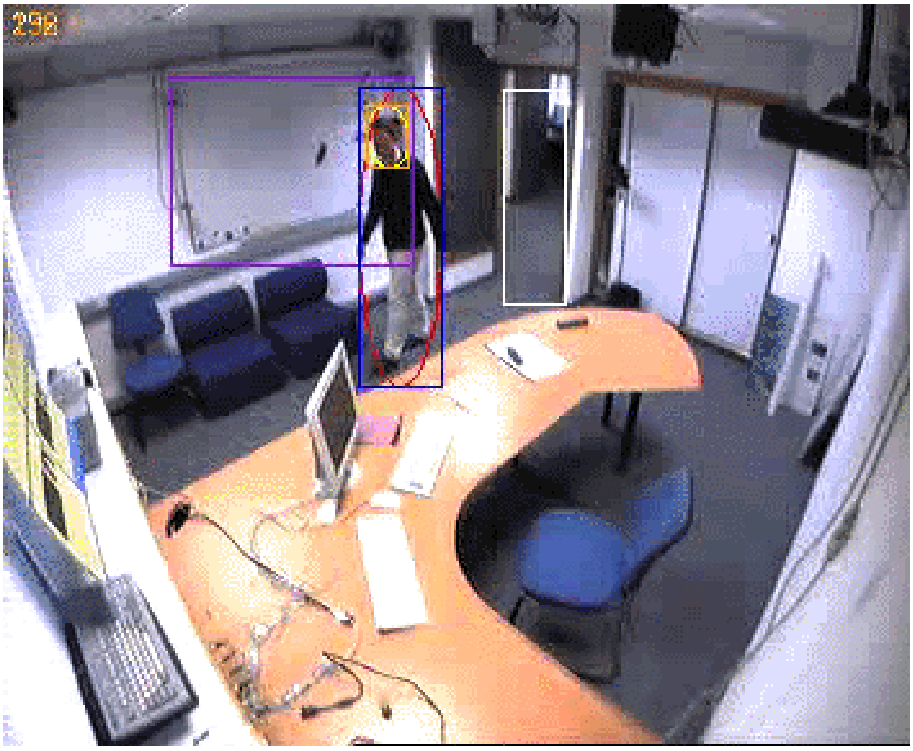

Figure 11. The new programmable robust tracker makes it possible to observe composite entities

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 O3MiSCID 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.

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 (depot Numero..

6.3. Tracking Focus of Attention

Keywords: *Face orientation estimation.*

Participants: Nicolas Gourier, Jerome Maisonasse, Oliver Brdiczka, James L. Crowley.

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^o in yaw (pan) angle and 12^o 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 experiments 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.

7. Contracts and Grants with Industry

7.1. European and National Projects

7.1.1. IST 506909 CHIL: Computers in the Human Interaction Loop

European Commission project IST 506909 (Framework VI - Call 1)

Strategic Objective: Multi-modal Interaction

Start Date 1 January 2004.

Duration 36 months (renewable).

CHIL is an Integrated Project in the new Framework VI programme.

Participants

- Fraunhofer Institut fuer Informations- und Datenverabeitung, Karlsruhe, Germany
- Universitaet Karlsruhe (TH), Interactive Systems Laboratories, Germany
- Daimler Chrysler AG, Stuttgart, Germany
- ELDA, Paris, France
- IBM Czech Republic, Prague, Czech Republic
- Research and Education Society in Information Systems, Athens, Greece
- Insitut National Polytechnique de Grenoble, France
- Insituto Trentino di Cultura, Trento, Italy
- Kungl Tekniska Hogskolan (KTH), Stockholm, Sweden
- Centre National de la Recherche Scientifique, Orsay, France
- Technische Universiteit Eindhoven, Eindhoven, Netherlands
- Universitaet Karlsruhe (TH), IPD, Karlsruhe, Germany
- Universitat Politecnica de Catalunya, Barcelona, Spain
- Stanford University, Stanford, USA
- Carnegie Mellon University, Pittsburgh, USA

The theme of project IP CHIL is to put Computers in the loop of humans interacting with humans. To achieve this goal of Computers in the Human Interaction Loop (CHIL), the computer must engage and act on perceived human needs and intrude as little as possible with only relevant information or on explicit request. The computer must also learn from its interaction with the environment and people. Finally, the computing devices must allow for a dynamically networked and self-healing hardware and software infrastructure. The CHIL consortium will build prototypical, integrated environments providing:

Perceptually Aware Interfaces: Perceptually aware interfaces can gather all relevant information (speech, faces, people, writing, and emotion) to model and interpret human activity, behaviour, and actions. To achieve this task we need a variety of core technologies that have progressed individually over the years: speech recognition and synthesis, people identification and tracking, computer vision, automatic categorization and retrieval, to name a few. Perceptually aware interfaces differ dramatically from past and present approaches, since the machine now observes human interaction rather than being directly addressed. This requires considerably more robust and integrated perceptual technology, since perspectives, styles and recording conditions are less controlled and less predictable, leading to dramatically higher error rates.

Cognitive Infrastructure: The supporting infrastructure that will allow the perceptual interfaces to provide real services to the uses needs to be dramatically advanced. Cognitive and Social modeling to understand human activities, model human workload, infer and predict human needs has to be included in the agent and middleware technology that supports CHIL. Further, the network infrastructure has to be dynamic and reconfigurable to accommodate the integration of a variety of platforms, components, and sensory systems to collaborate seamlessly and on-demand to satisfy user needs.

Context Aware Computing Devices: CHIL aims to change present desktop computer systems to context aware computing devices that provide services implicitly and autonomously. Devices will be able to utilize the advanced perceptual interfaces developed and the infrastructure in CHIL to free the user and allow him instead of serving the device to be served and supported in the tasks and human-to-human interactions he needs to focus. Further, human centered design, where the artistic value, appeal, and look & feel, become important in taking computing devices and human environments to the next level.

Novel services: The above innovations and advances in perceptual interfaces, cognitive infrastructure and context aware computing devices are integrated and showcased in novel services that aim at radically changing the way humans interact with computers to achieve their tasks in a more productive and less stressful way. These services are based on a thorough understanding of the social setting, the task situation, and the optimal interaction that maximizes human control while minimizing workload. Furthermore, some issues of privacy and security are to be addressed since the change human-computer interaction introduced by CHIL also touches a lot of the ways information in which is shared and communicated.

New measures of Performance: The resulting systems should reduce workload in measurable ways. To achieve these breakthroughs in a number of component technologies, the integrated system and a better understanding of its new use in human spaces are needed. Evaluation must be carried out both, in terms of performance and effectiveness to assess and track progress of each component, and the "end to end" integrated system(s). This will be carried out by an independent infrastructure that would also allow any third party to benchmark its findings against the project results after the end of the project.

7.1.2. RNTL/Proact: ContAct Context management for pro-Active computing

Start Date February 2003.

Duration: 36 months

The consortium consists of five partners:

- Xerox Research Centre Europe (Project coordinator)
- Project PRIMA, Laboratoire GRAVIR, INRIA Rhone Alpes
- Neural Networks Research Centre, Helsinki University of Technology (HUT), Finland
- Jaakko Pyry Consulting, Helsinki, Finland
- Ellipse, Helsinki, Finland

Project Contact is one of three RNTL projects that have been included in the French-Finland scientific program: ProAct.

The aim of Project RNTL CONTACT was to explore novel approaches to the detection and manipulation of contextual information to support proactive computing applications, and to make the results available as part of a more extensive toolkit for ubiquitous and proactive computing. In particular the project has addressed address two levels of context manipulation:

- Support for developing and adapting methods to compute context variables.
- The construction of example classifiers for context and situation.

To achieve these results project CONTACT has included:

1. Definition of an ontology that describes context variables both at the user and at the sensor level.
2. Definition of a platform providing formalism and an appropriate architecture to learn and combine context attributes.
3. Definition of a library of context attributes, general enough to be reusable in support of different scenarios than the one used in the project.
4. Validation of the contextual middleware on a pilot case. The chosen application of personal time management will help guide the development of the middleware and also to conduct an evaluation of our technology using a real-world problem.

7.2. Industrial Contracts

7.2.1. France Telecom: Project HARP

The HARP project - Human Activity Recognition and Prediction - started in may 2005 as a collaboration between INRIA and France Telecom R&D. The project required two persons: 18 months for a PhD student and 12 months for engineer funded by France Telecom R&D.

The HARP Project realized in collaboration with France Telecom R&D The purpose of this project is the development of a web service component in order to send information about current activity of users in a perceptual environment in direction to France Telecom's architectures services. The extracted information belongs to two different details levels. The lowest level represents users actions. The highest level represents the scenario which is playing by users. A global representation of hieratical structure of our approach is presented in figure 13. In order to extract this information, HARP component is based on computer vision techniques with three dimensions tracker component and acoustic processing techniques with a voice detector component. The experimental environment is a smart environment equipped with video cameras and microphones.

The 12 recognized users actions are a combination of movements, postures and relatives distances between users and elements which inhabit the environment. From 3D target properties, 3 postures are recognized with SVM classifier: standing, sitting and sleeping. The strong interest of 3D tracker is the occlusion robustness when more than one user is present in the scene (figure 12). A heuristic approach presented in figure 14, allows the roles recognition by combining the movements of entities, the distances with objects and recognized posture.

Based on the combination of actions and acoustics events such as speech or noise detection, learning techniques are used in order to recognize scenario in on-line and off-line situation. From the derived multimodal observations, different situations: aperitif, presentation, siesta, individual work and game are learned and detected using statistical models (HMMs) trained for each type of scenario. The experimental results shows excellent results situated between 66% and 8% scenario recognition rate when an observation window size is used with respectively 250 and 1500 frames. We present a web service interface in figure 15 which illustrates actions (icons) and scenario (colours) outputs in real-time application.

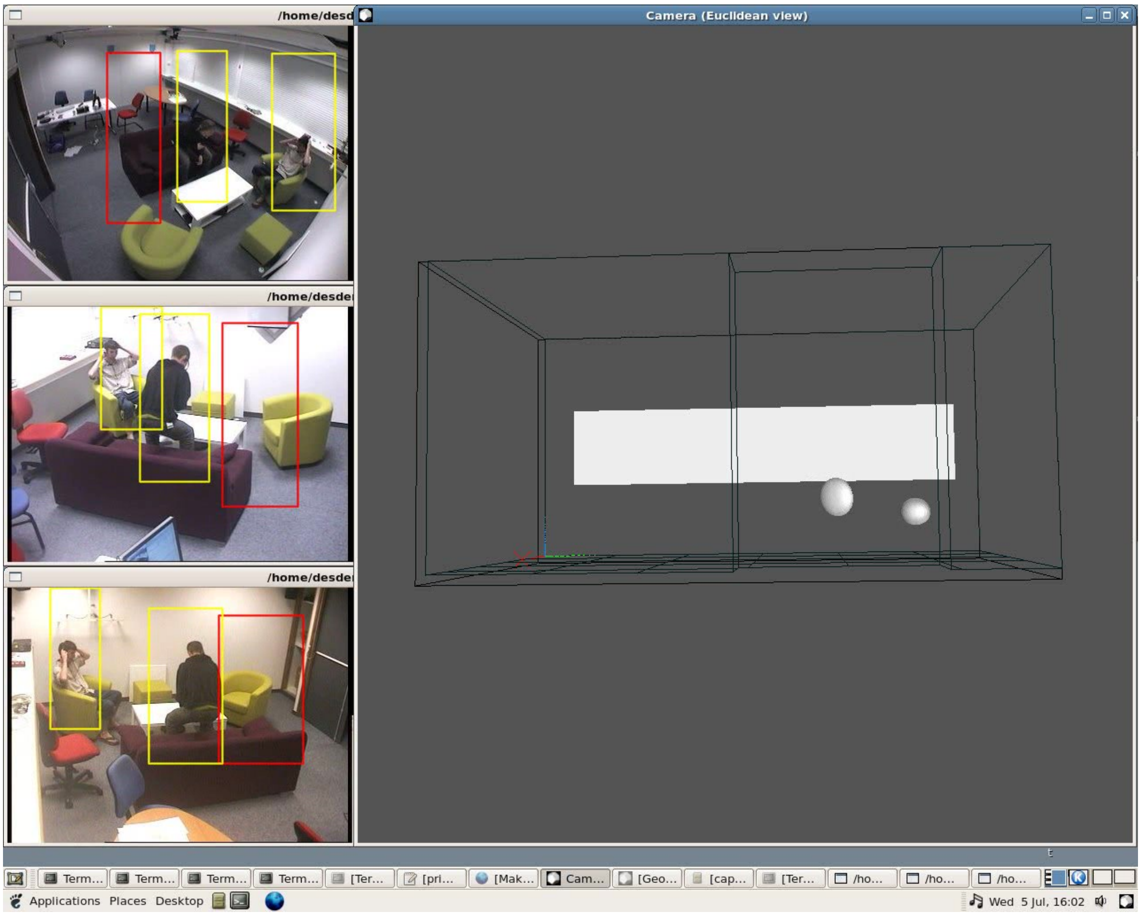

Figure 12. 3D video tracking system with 3D representation.

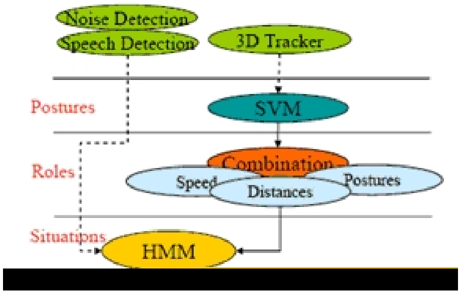

Figure 13. Overview of the different parts and methods of our approach.

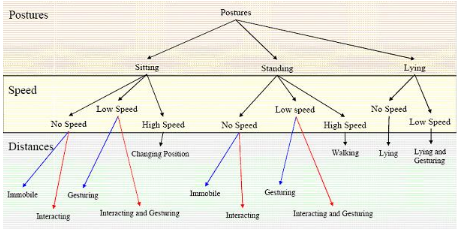

Figure 14. Schema describing the combination of posture, speed and distance values (blue arrows refer to 'no interaction distance with table', red arrows refer to 'interaction distance with table').

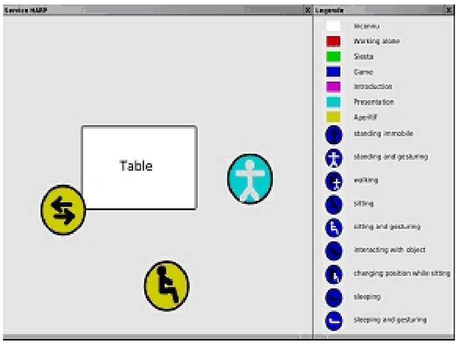

Figure 15. Web interface visualizing detected roles and situations (per entity) in real-time.

8. Dissemination

8.1. Contribution to the Scientific Community

8.1.1. Participation on Conference Program Committees

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

- ICVS 07, International Conference on Vision Systems, Beilefeld, March 2007
- ICRA 07, International Conference on Robotics and Automation, Rome, April 2007
- Percom 07, International Conference on Pervasive Computing, New York, March 2007
- IROS 2006, IEEE Conference on Intelligent Robotics and Systems, July, 2006
- ICPR 2006, International Conference on Pattern Recognition, Hong Kong, August 2006
- CVPR 2006, IEEE Conf. on Computer Vision and Pattern Recognition, juin 2006
- ECCV 2006, European Conference on Computer Vision, Graz Au, may 2006
- ICVS 2006, 4th International Conference on Vision Systems, New York, Jan 2006

8.1.2. Invited Presentations by James L. Crowley

1. Perception of human activity for Interactive Environments, OFTA, 13 April 2006.
2. Situated Observation of Human Activity, IBM T.J. Watson Research Center, White Plains NY, 22 June 2006.
3. Perception of human activity for Interactive Environments, Idea's Lab, CENG, 12 October 2006.

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^o 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^o 3, March 2000, p. 54-64.

[6] J. 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, in Machine Vision and Applications", vol. 16, n^o 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. CROWLEY. *Recognition without Correspondence using Multidimensional Receptive Field Histograms*, in "International Journal of Computer Vision", vol. 36, n^o 1, January 2000, p. 31–50.

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

Year Publications

Doctoral dissertations and Habilitation theses

[12] M. ANNE. *Integration de Services Perceptuels dans une Infrastructure de Communication Ambiante*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, december 2006.

[13] S. BORKOWSKI. *Steerable Interfaces for Interactive Environments*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, June 2006.

[14] N. GOURIER. *Machine Observation of the Direction of Human Focus of Attention*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, October 2006.

[15] 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.

Articles in refereed journals and book chapters

[16] O. BRDICZKA, P. REIGNIER, J. L. CROWLEY. *Modéliser et faire évoluer le contexte dans des environnements intelligents*, in "Ingénierie des Systèmes d'Information (ISI)", 2006.

[17] O. BRDICZKA, P. YUEN, J. L. CROWLEY, P. REIGNIER. *ASiMo: Automatic Acquisition of Situation Models*, in "eMinds: International Journal on Human-Computer Interaction", 2006.

[18] J. L. CROWLEY. *Situated Observation of Human Activity*, in "Computer Vision for Interactive and Intelligent Environments", C. JAYNES, R. COLLINS (editors). , IEEE Press, April 2006, p. 97–108.

[19] J. L. CROWLEY. *Situation Models for Observing Human Activity*, in "ACM Queue Magazine", May 2006.

[20] J. L. CROWLEY. *Things that See: Context-Aware Multi-modal Interaction*, in "Cognitive Vision Systems: Sampling the Spectrum of Approaches", H.-H. NAGEL, H. CHRISTENSEN (editors). , chap. Recognition and Categorization, Springer Verlag, Heidelberg, 2006.

[21] J. L. CROWLEY, P. REIGNIER, J. COUTAZ. *Designing Context Aware Services for Ambient Informatics*, in "True Vision in Ambient Intelligence", E. AARTS (editor). , Springer Verlag, March 2006, p. 231–244.

[22] D. HALL. *A system for object class detection*, in "Cognitive Vision Systems: Sampling the Spectrum of Approaches", H.-H. NAGEL, H. CHRISTENSEN (editors). , chap. Recognition and Categorization, Springer Verlag, Heidelberg, 2006.

[23] D. HALL. *Automatic parameter regulation of perceptual systems*, in "Image and Vision Computing", vol. 24, n^{o} 8, August 2006, p. 870-881.

Publications in Conferences and Workshops

[24] O. BERTRAND, A. LUX, T. T. H. TRAN. *From Ridges in Scale-Space to Hierarchical Shape Representation*, in "1st Int. Workshop on Shapes and Semantics, Matsushima, Japan", Area della Ricerca of the Consiglio Nazionale delle Ricerche, Genova, Italy, June 2006, p. 13–21, http://www.ge.cnr.it.

[25] 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", IEEE Computer Society Press, january 2006, p. 9–16.

[26] S. BORKOWSKI, J. MAISONNASSE, J. LETESSIER, J. L. CROWLEY. *Exploiter des interfaces mobiles dans le cadre d'un travail collaborative co-present*, in "18e Conférence Francophone sur l'Interaction Homme-Machine, IHM'06, Montreal, Canada", April 2006, http://www-prima.imag.fr/prima/pub/Publications/2006/BMLC06/.

[27] O. BRDICZKA, J. MAISONNASSE, P. REIGNIER, J. L. CROWLEY. *Extracting Activities from Multimodal Observation*, in "10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, Bournemouth, UK", October 2006, p. 162–170.

[28] O. BRDICZKA, J. MAISONNASSE, P. REIGNIER, J. L. CROWLEY. *Learning Individual Roles from Video in a Smart Home*, in "2nd IEE International Conference on Intelligent Environments, Athens, Greece", July 2006, p. 61–69.

[29] O. BRDICZKA, P. REIGNIER, J. L. CROWLEY, D. VAUFREYDAZ, J. MAISONNASSE. *Deterministic and Probabilistic Implementation of Context*, in "Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops", March 2006, p. 46–50.

[30] O. BRDICZKA, D. VAUFREYDAZ, J. MAISONNASSE, P. REIGNIER. *Unsupervised Segmentation of Meeting Configurations and Activities using Speech Activity Detection*, in "3rd IFIP Conference on Artificial Intelligence Applications & Innovations (AIAI) 2006, Athens, Greece", June 2006, p. 47–52.

[31] O. BRDICZKA, P. YUEN, S. ZAIDENBERG, P. REIGNIER, J. L. CROWLEY. *Automatic Acquisition of Context Models and its Application to Video Surveillance*, in "18th International Conference on Pattern Recognition, ICPR'06, Hong Kong", August 2006, p. 1175–1178.

[32] J. L. CROWLEY, O. BRDICZKA, P. REIGNIER. *Learning Situation Models for Understanding Activity*, in "5th International Conference on Development and Learning, ICDL'06, Bloomington, USA", May 2006.

[33] R. EMONET, D. VAUFREYDAZ, P. REIGNIER, J. LETESSIER. *O3MiSCID: an Object Oriented Opensource Middleware for Service Connection, Introspection and Discovery*, in "1st IEEE International Workshop on Services Integration in Pervasive Environments, Lyon - France", June 2006.

[34] N. GOURIER, J. MAISONNASSE, D. HALL, J. L. CROWLEY. *Head Pose Estimation on Low Resolution Images*, in "CLEAR Workshop, In Conjunction with Face and Gesture", Springer Verlag, April 2006.

[35] D. HALL, R. EMONET, J. L. CROWLEY. *An automatic approach for parameter selection in self-adaptive tracking*, in "International Conference on Computer Vision Theory and Applications (VISAPP)", 2006.

[36] J. MAISONNASSE, O. BRDICZKA, N. GOURIER, P. REIGNIER. *Attentional Model for Perceiving Social Context in Intelligent Environments*, in "3rd IFIP Conference on Artificial Intelligence Apllications and Innovations (AIAI)", I. MAGLOGIANNIS (editor). , Springer, IFIP, june 2006, p. 171–178.

[37] J. MAISONNASSE, N. GOURIER, O. BRDICZKA, P. REIGNIER, J. L. CROWLEY. *Detecting Privacy in Attention Aware System*, in "2nd International Conference on Intelligent Environments", IET, july 2006, p. 231–239.

[38] J. MAISONNASSE, N. GOURIER, O. BRDICZKA, P. REIGNIER, J. L. CROWLEY. *Gestion d'applications confidentielles sur la base d'un modèle attentionnel*, in "Troisiémes Journées Francophones: Mobilité et Ubiquité (UbiMob)", ACM, ACM, IEEE France, september 2006, p. 143–146.

[39] A. NEGRE, T. T. H. TRAN, N. GOURIER. *Comparative Study of People Detection in Surveillance Scene*, in "Structural and Syntactic Pattern Recognition, in Conjunction with ICPR, Hong Kong", Springer Verlag, August 2006.

[40] P. REIGNIER, S. ZAIDENBERG, R. EMONET, D. VAUFREYDAZ, J. LETESSIER. *O3MiSCID, un intergiciel sous OSGi pour l'informatique ubiquitaire*, in "Atelier OSGi, 3e Journées Francophones Mobilité et Ubiquité, Paris, France", September 2006.

[41] P. T. TON, A. LUX, T. T. H. TRAN. *Graph Based Model for Object Recognition*, in "First International Conference on Theories and Applications of Computer Science (ICTACS2006)", August 2006.

[42] D. VAUFREYDAZ, R. EMONET, P. REIGNIER. *A Lightweight Speech Detection System for Perceptive Environments*, in "3rd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, Washington DC, USA", May 2006.

[43] S. ZAIDENBERG, O. BRDICZKA, P. REIGNIER, J. L. CROWLEY. *Learning context models for the recognition of scenarios*, in "3rd IFIP Conference on Artificial Intelligence Applications &amp; Innovations (AIAI) 2006", IFIP International Federation for Information Processing, vol. Volume 204/2006, Springer Boston, IFIP, june 2006, p. 86-97.

References in notes

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

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

[46] 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.

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

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

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

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

[51] J. 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.

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

[53] 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.

[54] N. GOURIER, D. HALL, J. 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.

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

[56] D. HALL, F. PÉLISSON, O. RIFF, J. CROWLEY. *Brand identification using Gaussian derivative histograms*, in "Machine Vision and Applications", vol. 16, n^{o} 1, 2004, p. 41–46.

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

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

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

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

[61] 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.

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

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

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

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

[66] N. NAKAMURA, R. HIRAIKE. *Active Projector: Image correction for moving image over uneven screens*, in "Companion of the 15th Annual ACM Symposium on User Interface Software and Technology", October 2002, p. 1–2.

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

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

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

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

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

[72] R. RASKAR. *iLamps: Geometrically Aware and Self-Configuring Projectors*, in "ACM SIGGRAPH 2003 Conference Proceedings", 2003.

[73] 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.

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

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

[76] 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.

[77] 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.

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

[79] 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.

[80] 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.

[81] J. UNDERKOFFLERAND, B. ULLMER, H. ISHII. *Emancipated Pixels: Real-World Graphics in the Luminous Room*, in "Proceedings of ACM SIGGRAPH", 1999, p. 385-392.

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

[83] 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.

[84] 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.

[85] G. YE, J. 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.

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