INRIA, Evaluation of Theme CogC

Project-team Prima

May 2006

Project-team title : Prima - Perception Recognition and Integration for Observation of Activity

Scientific leader : James L. Crowley, Professeur, I.N.P. Grenoble

Research center : INRIA Rhône Alpes

1 Personnel

Personnel (Project creation january 2003)

	Misc.	INRIA	CNRS	University	**Total**
DR (1) / Professors				2	**2**
CR (2) / Assistant Professors				1	**1**
Permanent Engineers (3)					**0**
Temporary Engineers (4)		2		1	**3**
PhD Students				4	**4**
Post-Doc.				1	**1**
Total		**2**	**0**	**9**	**11**
External Collaborators					
Visitors (> 1 month)					

(1) "Senior Research Scientist (Directeur de Recherche)"
(2) "Junior Research Scientist (Chargé de Recherche)"
(3) "Civil servant (CNRS, INRIA, ...)"
(4) "Associated with a contract (Ingénieur Expert or Ingénieur Associé)"

Personnel (May 2006)

	Misc.	INRIA	CNRS	University	**Total**
DR / Professors				2	**2**
CR / Assistant Professor				2	**2**
Permanent Engineer					**0**
Temporary Engineer		1		1	**2**
PhD Students		3	1	6	**10**
Post-Doc.					**0**
Total		**4**	**1**	**11**	**16**
External Collaborators					
Visitors (> 1 month)					

Changes in staff

DR / Professors CR / Assistant Professors	Misc.	INRIA	CNRS	University	total
Arrival				1 AP	
Leaving					

Current composition of the project-team : May 2006

- James L. Crowley [Professor INPG], head of the team
- Augustin Lux [Professor INPG]
- Patrick Reignier [Assistant professor UJF]
- Dominique Vaufreydaz [Assistant professor UPMF]
- Alba Ferrer-Biosca [Expert Engineer]
- Jean-Marie Vallet [Expert Engineer]
- Olivier Bertrand [PhD student, AMN, bourse MENRT]
- Stanislas Borkowski [PhD student, bourse EGIDE]
- Oliver Brdiczka [PhD student, Bourse INRIA]
- Suphot Chunwiphat [PhD student, Thai gouvernement]
- Remi Emonet [PhD student, Bourse MENRT]
- Nicolas Gourier [PhD student, Bourse INRIA]
- Julien Letessier [PhD student, Bourse INRIA]
- Jerome Maisonnasse [PhD student, INPG SA - Contrat France Telecom]
- Matthieu Anne [PhD student, Bourse CIFRE - France Telecom]
- Sofia Zaidenberg [PhD student, BDI CNRS]

Current position of former project-team members (including PhD students during the period 2003-2006):

- Fabien Pelisson (Phd january 2003) (Technical Director, Blue Eye Video)
- Christophe Legal (Phd january 2003), engineer ENIB
- Olivier Riff (Engineer, ID3Semiconductors, St. Egreve)
- Alban Caporossi engineer INSERM Grenoble
- Sébastien Pesnel (Engineer, Blue Eye Video)
- Daniela Hall (Project Team Leader, société ActiCM, Moirans)
- Thi-Thanh-HaiTran (Phd march 2006): postdoc at Irisa

Last INRIA enlistments

- NONE

Other comments :

Many of the innovations described below are the fruits of doctoral and post-doctoral researchers on temporary work contracts, self-financed by external research projects. Under French employment law, such employment is of limited duration (36 months) during which time the doctoral student is under heavy pressure to provide both project deliverables and a doctoral thesis. As these people leave, they take with them their acquired experience. The permanent staff of PRIMA is limited to 4 academic researchers. It is vital that the PRIMA project be reinforced with at least one permanent researcher in order to conserve and develop the knowledge and intellectual property developed in its funded research projects.

2 Work progress

2.1 Keywords

Interactive Environments, Computer Vision, Machine Perception, Man-Machine Interaction, Perceptual User Interfaces

2.2 Context and overall goal of the project

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.

A technology for human environments that are capable of perceiving, acting, communicating, and interacting with occupants is expected to provide a "mother technology", enabling new forms of man-machine interaction, context-aware information services, entertainment, collaborative work and meeting support, context aware video communications, as well as services and security based on visual surveillance.

2.3 Objectives for the evaluation period

Over the last four years, PRIMA has worked to develop:

- A conceptual framework for observing and modeling of human activity,

- A new modality for man-machine interaction based on interaction with projected information,
- Robust computer vision methods based on local appearance, and
- A component-based software architecture for integration of autonomic perceptual components.

Research towards these challenges has been tightly integrated, with each area providing problems, experiments, and technologies for the other areas. For example, work on context aware environments enables new forms of man-machine interaction. Both environments, and interaction technologies employ our robust autonomic architecture for real time multimodal perception.

2.4 Objective 1 Context aware interactive environments : Executive summary

Keywords: Situation Modeling, Activity modelling, Roles and Situations

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.

2.4.1 Personnel

James Crowley, Patrick Reignier, Dominique Vaufreydaz, Jerome Maisonasse, Olivier Brdiczka, Sophia Zaidenberg, Alba Ferrer, Suphot Chunwiphat

2.4.2 Project-team positioning

The project closest to PRIMA in this area is ORION. INRIA project ORION at Sophia Antipolis, uses 3D models for scene representation and composes models of scene activities based on hierarchies of video events. PRIMA explores a complementary approach that can by driven by both 2D and 3D perceptual components.

For activity modeling, PRIMA has taken inspiration from theater, to develop 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.

To be meaningful, the formal expression of a situation model must be grounded in procedures and actions for real systems. 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. Roles and entities specify perceptual components that detect and observe entities, assigning actors and objects to roles in order to provide interpret activity.

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.

2.4.3 Scientific and technological achievements

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 December 2004, incorporated a number of innovations, including 3D tracking of the lecturer and detection of face orientation and pointing gestures. This version is currently used to record the InTech lecture series a the INRIA amphitheater. Discussions are underway to commercialise this technology with the creation of a start up company.

2.4.4 Collaborations

European Project IST FAME (IST-2000-28323) - Facilitating Agent for Multicultural Communication

PRIMA, in strong collaboration with the group IIHM of Laboratoire CLIPS-IMAG, participated from Oct. 2001 to Fév. 2005 with Sony, Univ. Karlsruhe, IRST (Trento Italy), and UPC (Barcelone, Esp.) in project FAME. The goal of IST Project FAME has been to construct an intelligent agent to facilitate communication among people from different cultures who collaborate on solving a common problem. This agent will provides services: 1) facilitate human to human communication through multimodal interaction including vision, speech and object manipulation, 2) provide the appropriate information relevant to the context, and 3) make possible the production and manipulation of information blending both electronic and physical representations. The agent serves as an information butler to aid multicultural communication in a transparent way. The agent does not intervene in the conversation, but remains in the background to provide the appropriate support. A two week public demonstration was given at the Barcelona Cultural Fair in July 2004.

European Project IST CAVIAR (IST 2001- 37540) – Context Aware Vision for Image-based Analysis and Recognition.

In the project CAVIAR, PRIMA has worked with University of Edinburgh and University of Lisbon to develop context aware techniques for Computer Vision. The main objective of the CAVIAR is to address the scientific question: Can rich local image descriptions from foveal and other image sensors, selected by a hierarchal visual attention process and guided and processed using task, scene, function and object contextual knowledge improve image-based recognition processes? This is clearly addressing issues central to the cognitive vision approach.

European Project IST IP CHIL (IST 2003 - 506909) - Computers in the Human Interaction Loop

The partners of IST CHIL include Fraunhofer Institut Karlsruhe, Universität Karlsruhe (TH), Daimler Chrysler AG, Stuttgart, ELDA, Paris, IBM Czech Republic, Resit , Athens, INPG, ITC Trento, KTH Stockholm, CNRS, TU Eindhoven, UPC Barcelona, Stanford University, Carnegie Mellon University, and INRIA Rhone Alpes. Within this group we work especially with University of Karlsruhe, IBM and University of Barcelona.

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
- Cognitive Infrastructure
- Context Aware Computing Devices
- Novel services
- New measures of Performance

RNTL/ProAct project : ContAct with XRCE and Univ. of Helsinki – Context Aware Activity Monitoring

The aim of Project RNTL/ProAct 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.

To achieve these results project ContAct has included four major activities:

- Definition of an ontology that describes context variables both at the user and at the sensor level.
- Definition of a platform providing formalism and an appropriate architecture to learn and combine context attributes.
- Definition of a library of context attributes, general enough to be reusable in support of different scenarios than the one used in the project.
- 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.

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

France Telecom Project HARP.

The object of France Telecom project HARP is to develop Software, Sensors, Models and Learning Methods for observing human activity in domestic environments. Such monitoring should serve to adapt communications and information services to domestic activity. In this project we work with divisions of FT R&D in Lannion and Meylan.

2.4.5 External support

- European Project IST FAME
- European Project IST CAVIAR (IST 2001- 37540)
- European Project IST IP CHIL
- RNTL/ ProAct project : ContAct
- France Telecom R&D contract HARP

2.4.6 Self assessment

The PRIMA approach to role based situation modeling has been found to provide a fundamental technology for a wide variety of domains. These include

- Video Conferencing and Communications
- Information services for meetings and collaborative work environments
- Domestic monitoring and surveillance of aged and infirm persons for health care
- Video Surveillance for security in public areas
- Surveillance and monitoring for commercial services

As a result we have found ourselves solicited to play a role in a number of European and national projects across a wide spectrum of applications areas ranging from Observing group interactions (FAME), meetings (CHIL), and collaborative work tools (ContAct), to security in public areas (CAVIAR) and health care monitoring (project proposal in preparation). The large number of projects have tended to stretch the human resources of our small group. Fortunately, administrative help of INRIA Rhône alpes has made it possible for us to hire young engineers on fixed duration contracts to meet these challenging projects. Recruting of permenant researchers in this area would allow us to consolidate and solidify our advance in this area.

An important challenge for this work is to move from hand-crafted situation models and role observation to automatic learning and adaptation. This is the subject of two current doctoral thesis in the group.

We have recently been solicited to work with people in the domain of speech understanding and dialog modeling to adapt our approach to develop a theory for socially aware, polite, man-machine interaction. We consider this to be an exciting new challenge for the period to come.

2.5 Objective 2 New forms of man-machine interaction based on perception : Executive summary

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

Context aware perception of human activity provides a foundation for a variety of new approaches to interaction between humans and machines. In this area we explore such opportunities. In particular, over the last few years we have explored a method for augmenting the surfaces in an environment with tools for interacting with information. An important tool for our investigation has been a device for projected video interaction created by the marriage of a video projector and a digital camera placed on a motorized steerable platform. This device projects rectified interaction widgets to any convenient surface. User interaction with porjected widgets are observed using the digital camera, yielding an extremely natural and intuitive form of man-machine interaction.

2.5.1 Personnel

James Crowley, Dominique Vaufredaz, Stan Borkowski, Julien Letessier

2.5.2 Project-team positioning

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 [33]. The original functionality of the objects does not change, only its appearance. An example of object enhancement is presented in [1], 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 [2] are increasingly used in augmented environment systems [32] [35] as an inexpensive means of making projected images interactive. Steerable projectors [2] [33] 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 [43]).

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 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.

The INRIA project closest to PRIMA in this activity is IN-SITU at INRIA Future. The IN-SITU project develops new interaction techniques, tools and methods for designing interactive systems. IN-SITU concentrates on situated interfaces, i.e. interfaces that are adapted (or adaptable) to their contexts of use, taking advantage of the complementary aspects of humans and computers.

The projected interaction widgets developed in PRIMA are designed to be adapted to different user situations in just the sense intended by IN-SITU. As a group with strong competence in computer vision and robotics, PRIMA is well equipped to invent new interaction technologies based on these technologies. The techniques demonstrated in this area have been developed in collaboration with the IIHM group of Joelle Coutaz at the IMAG CLIPS laboratory [1]. Similar work has also been demonstrated in the Tele-Graffiti system [39] in Japan inspired by the the DigitalDesk concept-demonstration [44].

2.5.3 Scientific achievements

The Steerable Camera Projector assembly has made possible experiments with projected interaction widgets that detect fingers dwelling over button-style UI elements.

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, [2] 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 [21]. 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 [45].

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.

2.5.4 Collaborations

- IIHM group at CLIPS-IMAG (Thesis of J. Letessier)
- University of Karlsruhe (Projects IST FAME and CHIL)
- IBM TJ Watson research center
- XRCE, Xerox European Research Centre (Project RNTL/ProAct ContAct)

2.5.5 External support

European Project IST IP CHIL - Computers in the Human Interaction Loop
RNTL/ ProAct project : ContAct with XRCE and Univ. of Helsinki.

2.5.6 Self assessment

When we first proposed to combine a projector and camera on a steerable platform in early 2003, we were skeptical about the potential for commercial applications for such a device. However, 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. Pierre Grandvionnet, PdG of the company GP Screen, based in Valence, France, has asked to commercialise system based on our hardware and software for use in advertising in store windows and trade shows. The company Apropos group has asked for licensing right for use of this technology in sterile environments such as operating theaters and IC manufacturing facilities. We have also discussed possible licencing to Philips Research for use in their "Kitchen of the future".

Unfortunately, the two doctoral students who have developed this approach will both complete their theses in 2006. We are currently arranging creation of a start-up to maintain the software tool kit for commercial applications. It is likely that this work will be complemented for other approachs to perceptual interaciton in the next few years.

2.6 Objective 3 Robust view-invariant Computer Vision : Executive summary

Keywords: Local Appearance, Affine Invariance, Receptive Fields, Scale Normalisation, Robust Computer Vision.

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. During the period 2003 - 2006, the group has demonstrated several new innovations based on chromatic receptive fields and scale invariant ridges.

Results in this area have been used for

- Real time detection and tracking of publicity panels in broadcast video of sports events in the BrandDetect software,
- Real time detection and tracking of faces for observation of activity in meetings and lectures in the IST FAME and CHIL projects
- Real time detection and tracking of people for context aware observation of activity in the IST CAVIAR project
- Visual navigation in the ROBEA ParkView project

2.6.1 Personnel

Current Personnel:
James L. Crowley, Augustin Lux, Olivier Bertrand, Nicolas Gourier, Amaury Negre (Common with E-Motion)

Personnel working in this area who have left the group during the period 2003 - 2006
Fabien Pelisson, Olivier Riff, Daniela Hall, Alban Caporossi, Thi Thanh Hai Tran

2.6.2 Project-team positioning

The visual appearance of a neighbourhood can be described by a local Taylor series [22]. 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 [23] 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 [37]. 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 [23], and normalized in orientation using direction of the first derivatives [37].

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" [24]. 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" [12] of Gaussian derivatives. The problem is to determine the appropriate orientation. In earlier work by PRIMA members Colin de Verdiere [4], Schiele [37] and Hall [15], 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 [15]. 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. Permutations of RGB lead to different opponent color spaces. The choice of the most appropriate space depends on the chromatic composition of the scene.

The LEAR project at INRIA Rhone-Alpes uses a similar method to explore techniques for object recognition and scene interpretation for static images and prerecorded video sequences. LEAR concentrates on off-line techniques and is particularly interested in image retrieval and video indexing. PRIMA has worked to apply such techniques to on-line, real time systems for interaction and control. Recently LEAR has gained considerable success with methods for affine invariant local image description. Their approach involves an iterative algorithm that searches for the appropriate local affine transformation. PRIMA has explored an alternative approach using natural interest ridges that directly extract invariant descriptions under changes in viewpoint without iteration.

PRIMA has worked closely with project E-MOTION at INRIA Rhone-Alpes to provide real-time visual information for navigation of the CyCab vehicle using invariant receptive field vectors. In particular, Doctoral student Amaury Negre is co-directed by Christian

Laugier and James Crowley has demonstrated direct calculation of time to contact from rate of change in characteristic scale as measured by natural interest points and ridges, as well as real time detection and tracking of pedestrians.

2.6.3 Scientific achievements

We have recently achieved video rate calculation of intrinsic (characteristic) scale from interpolation within a Binomial Pyramid computed using an O(N) algorithm [9]. 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 [31]. 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 [31], [17].

Daniela Hall and Nicolas Gourier have developed machine learning techniques to statistically learn robust visual features for face tracking [17], [13].

Lux and Hai have very recently developed a method with provides a direct measurement of affine invariant local features based on extending natural interest points to "natural interest ridges" [42],[41]. 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.

Aumaury Negre has recently demonstrated direct computation of time to collision over the entire visual field using rate of change of intrinsic scale [29]. This approach is currently being adapted for use in visual navigation in joint work with project EMOTION.

2.6.4 Collaborations

European Project IST DETECT, (IST-2001-32157) Real Time Detection of Motion Picture Content in Live Broadcasts,

From november 2001 to February 2004 with Joanneum Research (A), Taurus Media Technik (D), HS-ART Digital Services (A), Duvideo (P) and Videocation (D).

The goal of the DETECT project has been to implement a general platform for real time detection of semantic blocks and regions within digital video streams. These detected regions are then subjectd to further analysis and processing. The project has focused on the applications problem of detecting and delimiting predefined static and dynamic objects. This issue has currently a very large demand for both cultural and economic reasons.

Within IST project DETECT, we have worked Joanneum Research (Graz, Au) to develop a system to detects and tracks publicity panels in broadcast video of sports events. This systems uses histograms of scale and orientation normalized chromatique receptive fields to detect and track publicity that in the presences of exterieur lighting, motion blur, and rapid transformations in scale and orientation. This system has been licensed to a small Austrian company named HSArt and is currently undergoing evaluation for gathering publicity statistics by the Neilsen Media Metrics company. The methods used in this system were based on the thesis work of Fabien Pelisson. These techniques are described in [9], [16], [17], [31].

European Project IST CAVIAR (IST 2001- 37540)

In the context of IST project CAVIAR, we have worked closely with University of Edinburgh and University of Lisbon to develop robust view invariant methods computer vision techniques for context aware observation of activity in public areas. The main objective of the CAVIAR were to address the scientific question: Can rich local image descriptions from foveal and other image sensors, selected by a hierarchal visual attention process and guided and processed using task, scene, function and object contextual knowledge improve image-based recognition processes?

CAVIAR has addressed problems of recognizing and describing human activity in a variety of scenarios drawn from city centre surveillance. In this project, we have demonstrated detection and tracking using receptive field histograms. We have also extended the concept of "natural interest point" to describe elongated structures with "natural interest lines". We have experimentally compared use of such structural descriptions to SIFT and other appearance based methods for detection, tracking and recognition of entities in video surveillance. Publications include [40], [18], [14]. Techniques are described in the doctoral theses of Fabien Pelisson [31], and Hai Tran [41].

In the IST FAME and CHIL projects, described below, we have worked with the Univeristy of Karlsruhe to develop techniques for robust detection and tracking of faces. The objective is to determine the object of attention for participants in meetings, lectures, and social groupings. This work has been reported in [13], [18].

Results in this area have been used for
- Real time detection and tracking of publicity panels in broadcast video of sports events in the BrandDetect software,
- Real time detection and tracking of faces for observation of activity in meetings and lectures in the IST FAME and CHIL projects
- Real time detection and tracking of people for context aware observation of activity in the IST CAVIAR project
- Visual navigation in the ROBEA ParkView project.

2.6.5 External support

European Project IST DETECT, IST-2001-32157
European Project IST FAME (IST-2000-28323) - Facilitating Agent for Multi-cultural Communication
European Project IST CAVIAR (IST 2001- 37540)
Projet Européen IST IP CHIL - Computers in the Human Interaction Loop

2.6.6 Self assessment

As noted above, the PRIMA group has a long history of fundamental contributions to robust, appearance based computer vision techniques using scale and orientation normalized Gaussian derivative fields, with a number of fundamental contributions predating incorporation as an INRIA project. In particular, papers by Schiele (ECCV 96, ECCV 98, ICCV 98, and IJCV 2000), Colin de Verdier (RFIA 98, ECCV 98, ECCV 2000), Chomat (CVPR 1999, ECCV 2000) and Hall (ECCV 2000) documented the potential of this approach well before its recent popularity.

In the last few years, the rising popularity of Lowe's "SIFT" [25], and the results of Schmid et al in Affine Invariant Descriptors [28] developed in the INRIA LEAR project, has helped trigger a rise in popularity of this approach. However, those approaches are limited to local neighborhoods around natural interest points. Such points ignore many informative image structures, and can become unstable in the presence of elongated image structures. Our recent extension of this approach to ”Natural Interest Ridges” provides natural solution to this problem.

We are currently expanding our efforts for use of natural interest ridges for use in real time tracking of people and landmarks, as well as for view invariant recognition.

2.7 Objective 4 Robust architectures for multi-modal perception : Executive summary

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

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 are applying recent progress in autonomic computing to the problem of producing reliable, robust perceptual components.

In PRIMA, we have developed a data-flow process architecture for components for perception and action. Component based architectures, as described in Shaw and Garlan [38], are composed of auto-descriptive functional components joined by connectors. Such an architecture is well adapted to interoperability of components, and thus provides a framework by which multiple partners can explore design of specific components without having to rebuild the entire system.

We have constructed a software architectural model based on dynamically assembled perceptual components to form federatiuons [10]. Our model builds on previous work on process-based architectures for machine perception and computer vision [36], [6], as well as on data flow models for software architecture.

2.7.1 Personnel

Current Personnel:

James Crowley, Augustin Lux, Patrick Reignier, Remi Emonet, Dominique Vaufreydaz, Alba Ferrer-Biosca

Personnel who have worked in this area and have left the project:

Daniela Hall, Sebastien Pesnel, Alban Caparossi

2.7.2 Project-team positioning

Perception and action operate at a higher level of abstraction than sensors and actuators. Perception interprets sensor signals by recognizing and observing entities. Abstract tasks are expressed in terms of a desired result rather than actions to be blindly executed.

To separate sensor-actuator control, perception-action behaviours, and user services, PRIMA has proposed a layered architectural model for services based on observing human activity. At the lowest layer, the service’s view of the world is provided by a collection of physical sensors and actuators. This corresponds to the sensor-actuator layer. This layer depends on the technology and encapsulates the diversity of sensors and actuators by which the system interacts with the world. Information at this layer is expressed in terms of sensor signals and device commands.

Components based porgramming 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 [20]. 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 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 [11], [20].

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 [19]. 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, and our colleagues a IBM TJ Watson research center, PRIMA group has taken a leading role in introducing autonomic system approachs to programming perceptual systems.

2.7.3 Scientific achievements

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 [34].

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 [7], [8], uses 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 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 [26] 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.

2.7.4 Collaborations

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.

In particular, our layered software model for interactive services, and our components based approach for perceptual components have been adopted as standards in the IST CHIL project. This project has allowed us to develop collaboration with IBM TJ Watson research center and with Walter Tichy (Univ of Karlsruhe) in the area of Autonomic Computing methods for perceptual systems and services.

2.7.5 External support

European Project IST CAVIAR (IST 2001- 37540)

European Project IST IP CHIL - Computers in the Human Interaction Loop

2.7.6 Self assessment

Members of the PRIMA group have a long established reputation as leaders in the area of architectures and evaluation methods for computer vision systems [5], [3]. For many years we have pioneered techniques for building self monitoring, self regulating and self describing perceptual components using computer vision. The recent rise popularity of autonomic computing and component based programming has both provided validation or our approach, and has made available many new tools and techniques for our work.

Collaboration with sotware engineering experts at Univ Kalsruhe and IBM have shown us that computer vision and acoustic perception are ideal domains for developing methods for autonomic computing. These collaborations have also shown that statistical learning, widely used in computer vision, but but relatively unknown in software engineering, have the potential to provide a very high impact on the field of autonomic systems design.

This area constitutes an important priority for the PRIMA project in the coming years, particularly with the thesis work of Remi Emonet (started in 2005), and the adaptation or our BIP components based programming model by partners in the CLIPS and LSR laboratories of IMAG in Grenoble.

3 Knowledge dissemination

3.0.7 Publications

	year1 2003	year2 2004	year3 2005	year4 2006
PhD Thesis	2			2
H.D.R (*)				
Journal		3	1	2
Journal (edited)		1		
Conference proceedings (**)	9	8	10	12
Book chapter			1	3
Book (written)				
Book (edited)				
Patent			1	
Technical report				
Deliverable	5	5	10	3

(*) HDR Habilitation à diriger des Recherches

(**) Conference and major workshops with a program committee

3.1 Software

Imalab 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 all researchers in the Prima team, and represents a considerable amount of effort (probably largely more than 10 man-years).

BrandDetect BrandDetect is a commercial software system developed by Prima in the context of European IST project DETECT (2002-2004). BrandDetect is a system for detection, tracking and recognition of corporate logos, commercial trademarks and other publicity panels in broadcast television video streams. BrandDetect collects statistics on the frequency of occurrence, size, appearance and duration of presentation of the publicity. It is especially designed for use in the production of broadcast video of sports events such as football matches and formula one racing.

CAR CAR is a robust real-time detection and tracking system for observing the actions of individuals in a commercial or public environment, 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].

AAVR 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.

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.

APTE 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.

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.

3.1.1 Valorization and technology transfert

Creation of the company Blue Eye Video

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 doubling roughly every 6 months. As of September 2005, Blue Eye Video has 9 employees, over 100 systems installed and a capital reserve of roughly 400 000 Euros.

3.2 Teaching

As academic personnel, James Crowley, Augustin Lux, Patrick Reignier and Dominique Vaufreydaz carry a full academic teaching load of 192 hours per year.

Professor Augustin Lux directs the Doctoral Program "Image-Vision-Robotics" of the INPG/UJF.

James Crowley has been on Sabbatical during the fall of 2004. From January 2005 through January 2006 his academic service has been "funded" by INRIA allowing him to devote his full time to research and project management.

3.3 Visibility

Television Reports

- National Television Station M6: Coup de Flash, Special report at 19h50, Les Technologies d'Interaction, broadcast in July 2004.
- National Television Station FR2 : Savoir Plus Science, Les Cameras Sont Devenues Intelligentes, broadcast samedi 10 september 2005 ay 13h50. (This was the lead story of their opening show. The story devoted 14 minutes to PRIMA and the creation of Blue Eye Video.
- Portugese Television : A 10 minute interview on the DETECT project broadcast in July 2003.
- Radio : A 2 minute interview on measuring the number of participants size for demonstrations, broadcast on "Nostalgie" in April 2006.
- Print: Numerous articles related to the creation of Blue Eye Video in the national and regional press.

James L. Crowley

Scientific Activities

1. Conference Co-Chairman: Soc-EUSAI'05: Smart Objects and Ambient Intelligence, Grenoble, October 2005.
2. Program Co-Chair, International Conference on Multimodal Interaction, ICMI '05, Trento Italy, Oct 05.
3. Program Co-chair: EUSAI '04 European Symposium on Ambient Intelligence, Eindhoven, Nov 2004.
4. Program Chairman, International Conference on Vision Systems, ICVS 2003, Graz Austria, April 2003
5. Member "Board of Editors" Robotics and Autonomous Systems, North Holland.
6. Member of conference program commitees: IROS 2006, ICPR 2006, CVPR 2006, ECCV 2006, ICVS 2006, ICCV 2005, IROS 2005, CVPR 2005,ICRA 2005, EUSAI 2004, ICMI 2004, CVPR 2004, ICPR 2004, FG 2004,ECCV 2004, IAS 2004, RFIA 2004, ICMI 2003, AMFG03, ICCV 2003, ICIP'03, sOc 03, ScaleSpace 2003, UBICOMP 2003 and many workshops.

Invited Presentations

1. Perception of human activity for Interactive Environments, OFTA, 13 April 2006.
2. "Context Aware Observation of Human Activity", Invited Presentation at A*STAR Cognitive Science Symposium, Agency for Science, Technology and Research (A*STAR) of Singapore, Sept 2005.
3. "Context Aware Observation of Human Activity", keynote Speaker at ASCI Symposium on Image Analysis, Boxmeer, Netherlands, June 2005.

4. "What does it mean to see? Recent progress in Computer Vision", Plenary presentation at the Polish National Robotics Conference, Poland, June 2004.

5. "Context Driven Observation of Human Activity", Plenary Presentation at PSIPS Processing Sensory Information for Proactive Systems, Oulu June 2004.

6. "Context Driven Observation of Human Activity", Invited presentation at the France-Taiwan Workshop on Information Technology, Ecole Polytechnique, Avril 2004.

7. "Context Driven Observation of Human Activity", Keynote presentation at EUSAI '03 - European Symposium on Ambient Intelligence, Amsterdam, 3-5 November 2003.

8. "Things that See: Context-Aware Multi-modal Interaction", Dagstuhl Workshop on Cognitive Vision, October 2003.

9. "A Research Roadmap for Cognitive Vision", Dagstuhl Workshop on Cognitive Vision, October 2003.

10. "Dynamic composition of process federations for context aware perception of human activity", Keynote presentation at KIMAS '03, Boston october 2003.

4 External Funding

(k euros)	year1	year2	year3	year4
National initiatives				
RNTL/Proact ContAct	52	52	52	0
Robea	23	18	18	0
European projects				
IST DETECT	129	99	56	0
IST FAME	155	159	90	20
IST CAVIAR	148	148	130	38
IST FGNet	21	22	48	0
IST ECVision	10	10	10	9
IST Chil	0	65	121	177
Industrial contracts				
FT HARP	0	0	40	90
FT R&D	0	0	0	24
Scholarships				
MENSR R.Emonet	0	0	8	30
BDI(CNRS) S.Zaidenberg	0	0	8	30
Post Doc*				
AI+				
ODL#				
Total	538	572	579	418

† INRIA Cooperative Research Initiatives
‡ Large-scale Initiative Actions
∗ other than those supported by one of the above projects
+ junior engineer supported by INRIA
# engineer supported by INRIA

ARCs

National initiatives

FrancoFinish ProAct project : ContAct with XRCE and Univ. of Helsinki.

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.

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

European projects

Project IST CAVIAR (IST 2001- 37540)

(nov. 2002 – oct. 2005) (With U. Edinburgh and Univ. de Lisbon)

The main objective of the CAVIAR is to address the scientific question: Can rich local image descriptions from foveal and other image sensors, selected by a hierarchal visual attention process and guided and processed using task, scene, function and object contextual knowledge improve image-based recognition processes? This is clearly addressing issues central to the cognitive vision approach.

Project IST FAME (IST-2000-28323) - Facilitating Agent for Multi-cultural Communication

PRIMA, in strong collaboration with the group IIHM of Laboratoire CLIPS-IMAG, participated from Oct. 2001 to Fév. 2005 with Sony, Univ. Karlsruhe, IRST (Trento Italie), and UPC (Barcelone, Esp.) in project FAME. The goal of IST Project FAME has been to construct an intelligent agent to facilitate communication among people from different cultures who collaborate on solving a common problem. This agent will provides services: 1) facilitate human to human communication through multimodal interaction including vision, speech and object manipulation, 2) provide the appropriate information relevant to the context, and 3) make possible the production and manipulation of information blending both electronic and physical representations. The agent serves as an information butler to aid multicultural communication in a transparent way. The agent does not intervene in the conversation, but remains in the background to provide the appropriate support. A two week public demonstration was given at the Barcelona Cultural Fair in July 2004.

Project IST DETECT, IST-2001-32157

"Real Time Detection of Motion Picture Content in Live Broadcasts", (november 2001 to February 2004) with Joanneum Research (A), Taurus Media Technik (D), HS-ART Digital Services (A), Laboratoire GRAVIR, UMR CNRS 5527, Duvideo (P) and Videocation (D).

The goal of the DETECT project has been to implement a general platform for real time detection of semantic blocks and regions within digital video streams. These detected regions are then subjectd to further analysis and processing. The project has focused on the applications problem of detecting and delimiting predefined static and dynamic objects. This issue has currently a very large demand for both cultural and economic reasons.

DETECT was an industrial driven project, although its nature is R&D. The project produced a prototype system, which has since been converted into a product for monitoring commercial broadcast video to collect statistics on publicity.

Associated teams and other international projects

Industrial contracts

France Telecom Project HARP.

Duration: May 2005 - December 2006

The object of France Telecom project HARP is to develop Software, Sensors, Models and Learning Methods for observing human activity in domestic environments. Such monitoring should serve to adapt communications and information services to domestic activity. In this project we work with divisions of FT R&D in Lannion and Meylan.

5 Objectives for the next four years

5.1 Context aware interactive environments

Our current methods use hand crafted networks of situations to represent the situation model. An important challenge for the coming years is create methods to automatically acquire and develop situation models, as well as to create methods for such models to adapt in order to provide robust invariant services under variations in operating conditions.

Adaptation allows a system to maintain consistent behaviour across variations in operating environments. The environment denotes the physical world (e.g., in the street, lighting conditions), the user (identification, location, goals and activities), social settings, and computational, communicational and interactive resources. Development refers to the automatic acquisition of situation and context, and ultimately the acquisition of the entities, roles, relations from which situations and contexts emerge.

Adaptation and development are fundamental to providing useful and usable services to a variety of users in the presence of large variations in resources and activities. Context is too complex to be pre-programmed as a fixed set of stable variables: worse, the contract itself, which defines "correct behaviour," is not always precisely specifiable in advance. Thus, the context model, contract and adaptation process must develop through observation and interaction with the environment. At the same time, context models must not be disruptive. This creates a dilemma: how can context models evolve and develop without introducing disruption? The challenge is to find the appropriate balance between implicit and explicit interaction for providing the feedback required for development. We must determine the appropriate degree of autonomy, and this problem can impact every level of abstraction.

One attractive solution to these constraints is to observe that the correct selection of elements - including adaptation strategies, interfaces, devices, information, etc. - can only be made in the infrastructure. This suggests replacing explicitly-coded responses to situations and contexts - which can only accommodate a fixed range of predicated compositions - with a higher-level, more knowledge-intensive, use of machine-readable strategies coupled with reasoning and learning. This goes beyond the usual distinction of closed- versus open- adaptive systems.

Current learning technologies require large sets of training data - something that is difficult to obtain for an extensible environment. Non-disruptive development of context models will require new ways of looking at learning, and may ultimately require a new class of minimally-supervised learning algorithms that will need to be studied explicitly as part of semi-autonomous systems.

Context is key in the development of new services that will impact social inclusion for the emerging information society. For this to come true, we need to find the right balance between contradictory features. If context is redefined continually and ubiquitously, then how users can form an accurate model of a constantly evolving digital world? If system adaptation is negotiated, then how to avoid disruption in human activities? We believe that clear architecture and a well-founded, explicit relationship between environment and adaptation are the critical factors - the key that will unlock context-aware computing at a global scale.

5.2 New forms of man-machine interaction based on perception

In the coming years, PRIMA will continue parallel exploration of two complementary paradigms for man machine interaction: 1) The machine as tool, and 2) The machine as collaborator.

The machine as tool. The SCP (steerable camera projector pair described above), is an example of the "machine as tool" paradigm. With this approach, perception of human action and projection of information are coupled to augment ordinary objects with information technology. We will continue development of this approach, using steerable projectors and dynamic service configuration to augment the human environment, and to use any surface as a focus for man-machine interaction. Notably, we are currently completing a tool-kit for visual interaction. This tool-kit is designed to provide an easy to use foundation for experimenting with all forms of projective widgets based on observing human activity.

The machine as collaborator. A new challenge for PRIMA involves developing the foundations for a technology for social interaction. Current information technology is autistic. The creation of a technology for artificial systems that are socially aware constitutes a grand challenge for humanity. Meeting this challenge will provide a new class of technology with profound social and economic impact. The emergence of socially aware systems can potentially mark a rupture point in the evolution of informatics and the start in an exponential growth in new applications across a broad spectrum.

We believe that social interaction between man and machine requires two core abilities. 1) The ability for machine to learn to sense and learn to evoke human emotions, and 2) The ability for machines to adopt social roles that correspond to the implicit expectations for users.

In the area of sensing and evoking emotions, we propose to collaborate with colleagues in the CHIL project who have developed artificial agent technologies and speech understanding to develop a technology for abilities to perceive and evoke interest and pleasure in human partners. Perception of the affect of a human partner can provide feedback for learning, both to refine the ability to evoke attention and pleasure, and to develop other cognitive abilities. We propose to use these abilities exploring scripting behaviours for virtual agents in the form of roles. For this work we propose to extend the situation models and roles that we have developed in the FAME and CHIL projects.

Note that we are not proposing, in PRIMA, to develop virtual agents or speech recognition. We will use virtual agent took kits developed by our partners IRST and KTH to provide a graphical display technology. We propose to adapt acoustic perception and speech recognition technologies developed by our partners IBM, UKA, KTH, and Univ of Erlangen to complete our work with computer vision.

5.3 View-Invariant Computer Vision

Natural Interest points are local extremal points in a Laplacian scale space as described by Crowley [Cro87] and Lindeberg [23]. Recent papers by Lowe [24] and Schmid [27] have led to rapid growth of interest in the use of such points as scale invariant keypoints on the part of the computer vision community. Schmid has recently extended the detection of such points to correct for affine approximations to local projective transformations using an iterative search method [28] Nonetheless, interest points suffer from a serious problem in the case of elongated structures. In such a case, many local peaks occurs along a ridge in the Laplacian scale space, and the number and position of keypoints becomes highly unstable. This phenomena is very common in real world scenes.

In doctoral work directed by Augustin Lux, Hai Tranh has recently revisited the problem of description of ridges in a Laplacian Scale Space. Tran and Lux have developed simple, robust detection technique based on differential invariants using the Hessian matrix of the Laplacian. They have shown that such ridges, referred to as "natural interest ridges" provide a natural complement to "natural interest points". Natural interest ridges provide scale and rotational invariant skeleton to organize detection and recognition [40]. In his Masters Research project this year, Amaury Negre has extended the work of Tran and Lux to provide a simple and robust algorithm for detecting and tracking natural interest ridges [29] across scale.

The PRIMA group is building on this approach to demonstrate methods for use of natural interest ridges and points for 1) Detecting and segmenting printed text under projective transformations 2) Estimating parameters for an analytic expression of optical flow in motion sequences. 3) Direct, real time calculation of "time to collision" measures over the entire visual field, 4) Real time detection and description of human posture from low-resolution surveillance images. 5) Real time-detection and description of face orientation in video sequences.

5.4 Robust architectures for multi-modal perception

In order to exploit autonomic components, user services must be designed with autonomic capabilities. Most important is the ability to launch and configure components, and to automatically reconfigure communications between components. This ability makes it possible to efficiently incorporate redundant components in the system, with the ability to respond to failure or degradation by shutting down the failed component and launching an alternate. Thus the user service must incorporate a form of monitor in order to assure detection and response to degradation. Such a configuration monitoring agent may constructed using agent technologies currently used to provide services.

It is conceivable to program a configuration monitoring agent with hard wired procedures for reconfiguration. However, a proper, robust, design can be obtained using an ontological model. Modern tools for software ontologies can be used as a component registry to record index available sensors and components according to their capabilities. Such a tool can also provide a registry for the data types and events consumed and produced by components.

An important current effort concerns dynamic service composition. Dynamic service composition will add flexibility and robustness to services. Dynamic service composition enables a system to create new services based on their semantic description by discovering appropriate services and combining them [30]. Service composition should be semantics-based so that a service is requested and composed not by its syntax but by its semantics. In order to enable semantics-based dynamic service composition, both the modeling of components as well as the service composition mechanism must support semantics.

Auto-criticism will be addressed by applying statistical learning techniques to learn a likelihood function for output reliability. Such a function would provide a numerical measure of the likelihood that the current component output is reliable. Auto-regulation can be provided by a separate monitoring process that detects when output of a component has become unreliable, and reprocesses the most recent data with variations of processing parameters and proposing the parameter set that produces the most reliable result. Rather than exhaustively explore the parameter space, such a monitoring process can be designed to provide a small incremental correction for each data record, thus gradually hill climbing to a local maximum. Auto-description requires that the component be programmed to respond with a description of its capabilities. Auto-monitoring requires that the component continually estimate a set of internal performance metrics that characterize its internal state as well as observe the availability of resources Auto-configuration within a perceptual component requires that the component be built with a possibility to vary its service levels or to enable and disable modules. We propose to equip the component with a scheduling mechanism driven by an explicit list of modules and parameters. Configuration will involve modifications to the execution schedule and service levels.

For most human activities, there are a potentially infinite number of entities that could be detected and an infinite number of possible relations for any set of entities. The appropriate entities and relations must be determined with respect to a task or service to be provided. This is the role of the situation model. Situation Models allow focusing attention and computing resources to determine the information required to provide services.

For a given state of a service, the situation model acts as a filter for events and streams from perceptual components. In certain well defined cases, arrival of event or interpretation of a stream can result in an event being sent to the service components. In this sense, the situation model acts as a bottom-up filter for events from perceptual components.

Services specify a context. The context determines the appropriate entities, roles and relations top observe with perceptual components. Information flow is inverted when a service changes state. In reaction to the user command, or to a change in situation, the service may send events to the situation model forcing a change in the situation graph, and possibly forcing a change in the configuration of perceptual components. In this case, the situation model acts in a top-down manner, elaborating and expanding the service event into commands to the situation model as well as to the perceptual component.

6 Bibliography of the project-team

6.1 Books and Monographs

6.2 Doctoral dissertations and "Habilitation" theses

1. C. Le Gal. *Intégration et contrôle de processus de vision répartis pour les environnements intelligents*. PhD thesis, INPG, january 2003.

2. F. Pelisson. *Reconnaissance et indexation basees sur l'apparence*. PhD thesis, INPG, january 2003.

3. Thi Thanh Hai Tran. *Etude de lignes d' intérêt naturelles pour la representation d'objets en vision par ordinateur*, PdD thesis, INPG, march 31 2006.

4. Stanislaw Borkowski *Steerable interfaces for Interactive Environments*, PdD thesis, INPG, June 26 2006.

6.3 Articles in refereed journals and book chapters

Journal articles:

1. J. Coutaz, J.L. Crowley, S. Dobson, and D. Garlan. Context is key. *Communications of the ACM, Special issue on the Disappearing Computer*, March 2005.
2. J.L. Crowley and J. Piater. Introduction to the special issue: International conference on vision systems. *Machine Vision and Applications*, 16(1):4–5, 2004.
3. D. Hall, F. Pélisson, O. Riff, and J.L. Crowley. Brand identification using gaussian derivative histograms. *Machine Vision and Applications*, 16(1):41–46, 2004.
4. A. Lux. The imalab method for vision systems. *Machine Vision and Applications*, 16(1):21–26, 2004.
5. J. Tisseau, M. Parenthoen, C. Buche, and P. Reignier. Comportements perceptifs d'acteurs virtuels autonomes : Une application des cartes cognitives floues. *Technique et Science Informatique*, 24(10):1259–1293
6. James L. Crowley. Situation Models for Observing Human Activity *ACM Queue Magazine*, to appear, 2006.

Book chapters:

1. J.L. Crowley, P. Reignier and J. Coutaz. Designing context aware services for ambient informatics. In W. Weber, J.M. Rabaey and E. Aarts, eds, *Ambient Intelligence*. Springer-Verlag Berlin and Heidelberg GmbH & Co, April, 2005.
2. James L. Crowley Things that see: Context-aware multi-modal interaction. In H.-H. Nagel and H.I. Christensen, editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, chapter Recognition and Categorization. Springer Verlag, Heidelberg, 2006.
3. D. Hall. A system for object class detection. In H.-H. Nagel and H.I. Christensen, editors, *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, chapter Recognition and Categorization. Springer Verlag, Heidelberg, March 2006.
4. James L. Crowley. Situated observation of human activity. In *Computer Vision for Intelligent Environments*. IEEE Press, to appear, 2006.

6.4 Publications in Conferences and Workshops

1. A. Lux. The imalab method for vision systems. In *International Conference on Vision Systems*, Graz, Austria, April 2003.
2. F. Pélisson, D. Hall, O. Riff, and J.L. Crowley. Brand identification using gaussian derivative histograms. In *International Conference on Vision Systems*, pages 492–501, Graz, Austria, April 2003.
3. J.L. Crowley and P. Reignier. An architecture for context aware observation of human activity. In *Workshop on Computer Vision System Control Architectures (VSCA 2003)*, 2003.
4. J.L. Crowley and O. Riff. Fast computation of scale normalised gaussian receptive fields. In *Scale Space Methods in Computer Vision*, pages 584–598, Skye, UK, June 2003.

5. D. Hall and J.L. Crowley. Computation of generic features for object classification. In *Scale Space Methods in Computer Vision*, pages 744–756, Skye, UK, June 2003.

6. S. Borkowski, O. Riff, and J.L. Crowley. Controle de l'affichage dans un environnement augmente. In *Conference Francophone sur l'Interaction Homme-Machine, IHM'03*, 2003.

7. S. Borkowski, O. Riff, and J.L. Crowley. Projecting rectified images in an augmented environment. In *ProCams Workshop*, Nice, October 2003.

8. J.L. Crowley and P. Reignier. Dynamic composition of process federations for context aware perception of human activity. In *International Conference on Integration of Knowledge Intensive Multi-Agent Systems, KIMAS 03*, Cambridge, USA, October 2003.

9. J.L. Crowley. Context driven observation of human activity. In *European Symposium on Ambient Intelligence*, Amsterdam, The Netherlands, November 2003.

10. D. Hall and J.L. Crowley. Détection du visage par caractéristiques génériques calculées à partir des images de luminance. In *Reconnaissance des formes et intelligence artificelle*, Toulouse, France, January 2004.

11. T.H.T. Tran and A. Lux. A method for ridge extraction. In *Asian Conference on Computer Vision*, January 2004.

12. A. Caporossi, D. Hall, P. Reignier, and J.L. Crowley. Robust visual tracking from dynamic control of processing. In *PETS '04, Performance Evaluation of Tracking and Surveillance*, pages 23–31, Prague, Czech Republic, May 2004.

13. S. Borkowski, S. Sabry, and J.L. Crowley. Camera-projector sensing and actuating abilities. In *Polish National Robotics Conference*, June 2004.

14. J. Letessier S. Borkowski and J.L. Crowley. Spatial control of interactive surfaces in an augmented environment. In *European Conference on Human Computer Interaction, EHCI 04*, July 2004.

15. N. Gourier, D. Hall, and J.L. Crowley. Estimating face orientation from robust detection of salient facial structures. In *POINTING '04, Visual Observation of Deictic Gestures*, August 2004.

16. N. Gourier, D. Hall, and J.L. Crowley. Facial feature detection robust to pose, illumination, and identity. In *SMC'04, International Conference on Systems, Man and Cybernetics*, October 2004.

17. Hoang Lan NGUYEN THI Thi Thanh Hai TRAN, Augustin LUX. 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.

18. Oliver Brdiczka, Patrick Reignier, and James L. Crowley. Automatic development of an abstract context model for an intelligent environment. In *PERCOM '05, International Conference on Pervasive Computing and Communications*, 2005.

19. D.Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R.B. Fisher, J. Santos Victor, and J.L. Crowley. Comparison of target detection algorithms using adaptive background models. In *PETS '05, Performance Evaluation in Tracking and Surveillance*, 2005.

20. J. Maisonnasse and O. Brdiczka. Détection automatique des groupes d'interactions. In *Deuxièmes Journées Francophones: Mobilité et Ubiquité (UbiMob)*, june 2005.

21. D. Hall. Automatic parameter regulation for a tracking system with an auto-critical function. In *CAMP '05, Computer Architecture for Machine Perception*, pages 39–45, Palermo, Italy, July 2005.

22. F. Metze, P. Gieselmann, H. Holzapfel, T. Kluge, I. Rogina, A. Waibel, M. Wolfel, J. Crowley, P. Reignier, D. Vaufreydaz, F. Berard, B. Cohen, J. Coutaz, S. Rouillard, V. Arranz, M. Bertran, and H. Rodriguez. the "FAME" interactive space. In *2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Edinburgh, UK, July 2005.

23. H. L. Nguyen Thi T. Tran, A. Lux and A. Boucher. A novel approach for text detection in images using structural features. In *International Conference on Advances in Pattern Recognition (ICAPR)*, Bath, United Kingdom, 21-25 August 2005.

24. Joëlle Coutaz, Stanislaw Borkowski, and Nicolas Barralon. Coupling interaction resources: an analytical model. In J. L. Crowley G. Bailly and G. Privat, editors, *Joint conference on Smart Objects and Ambient Intelligence sOc-EUSAI*, pages 183–188, oct 2005.

25. Oliver Brdiczka, Patrick Reignier, and James L. Crowley. Supervised learning of an abstract context model for an intelligent environment. In *Proceedings of Smart Object and Ambient Intelligence Conference (sOc-EUSAI)*, october 2005.

26. Oliver Brdiczka, Patrick Reignier, and Jérôme Maisonnasse. Unsupervised segmentation of small group meetings using speech activity detection. In *ICMI Workshop Proceedings.* International Workshop on Multimodal Multiparty Meeting Processing (MMMP), october 2005.

27. Oliver Brdiczka, Jérôme Maisonnasse, and Patrick Reignier. Automatic detection of interaction groups. In *ICMI'05, Proceedings of International Conference on Multimodal Interfaces.* ACM, october 2005.

28. Stanislaw Borkowski, Julien Letessier, François Bérard, and James L. Crowley. User-centric design of a vision system for interactive applications. In *2006 IEEE International Conference on Computer Vision Systems, (ICVS'06)*, pages 9–16. IEEE Computer Society Press, Jan 2006.

29. Daniela Hall, Rémi Emonet, and James L. Crowley. An automatic approach for parameter selection in self-adaptive tracking. In *International Conference on Computer Vision Theory and Applications (VISAPP '06)*, Setubal, Portugal, Feb. 2006.

30. O. Brdiczka, P. Reignier, J.L. Crowley, D. Vaufreydaz, and J. Maisonnasse. Deterministic and probabilistic implementation of context. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications*, March 2006.

31. Stanislaw Borkowski, Jérôme Maisonnasse, Julien Letessier, and James L. Crowley. Exploiter des interfaces mobiles dans le cadre dún travail collaborative co-present. In *To appear in 18e Conférence Francophone sur lÍnteraction Homme-Machine, IHM0́6*, Montreal, Canada, April 2006.

32. N. Gourier, J. Maisonnasse, D. Hall, J.L. Crowley Head Pose Estimation on Low Resolution Images In *CLEAR '06, Classification of Events, Activities and Relationships*, South Hampton, UK, April 2006.

33. J.L. Crowley, O. Brdiczka, and P. Reignier. Learning Situation Models for Understanding Activity In *The 5th International Conference on Development and Learning 2006 (ICDL06)*, Bloomington, Il., USA, June 2006.

34. R. Emonet, D. Vaufreydaz, P. Reignier, J. Letessier O3MiSCID: an Object Oriented Opensource Middleware for Service Connection, Introspection and Discovery. In *SIPE '06, IEEE Workshop on Services Integration in Pervasive Environments*, June 2006.

35. O. Brdiczka, D. Vaufreydaz, J. Maisonnasse, and P. Reignier. Unsupervised segmentation of meeting configurations and activities using speech activity detection. In *3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI) 2006, Athens, Greece*, June 2006.

36. J. Maisonnasse, N. Gourier, O. Brdiczka, and P. Reignier. Attentional model for perceiving social context in intelligent environments. In *3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI) 2006, Athens, Greece*, June 2006.

37. S. Zaidenberg, O. Brdiczka, P. Reignier, and J.L. Crowley. Learning context models for the recognition of scenarios. In *3rd IFIP Conference on Artificial Intelligence Applications and Innovations (AIAI) 2006, Athens, Greece*, June 2006.

38. O. Brdiczka, J. Maisonnasse, P. Reignier, J.L. Crowley "Learning Individual Roles from Video in a Smart Home". In *IE'06, Second International Confernence on Intelligent Environments*, Athens Greece, July 2006.

39. O. Brdiczka, J. Maisonnasse, P. Reignier, J.L. Crowley Extracting Activities from Multimodal Observation In *KES2006 10th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, Bournemouth, UK., October 2006.

7 Bibliography

[1] F. Bérard. The magic table: computer-vision based augmentation of a whiteboard for creative meetings. In *Workshop on Projector-Camera Systems*, 2003.

[2] S. Borkowski, O. Riff, and J. L. Crowley. Projecting rectified images in an augmented environment. In *PROCAMS'03 Workshop*, 2003.

[3] H. I. Christensen and Eds J. L. Crowley. *Experimental Environments for Computer Vision and Image Processing.* World Scientific Press, 1993.

[4] V. Colin de Verdière and J.L. Crowley. Visual recognition using local appearance. In *ECCV98*, pages 640–654, Freiburg, June 1998.

[5] J. L. Crowley and H. I Christensen, editors. *Vision as Process.* Springer Verlag, 1993.

[6] J.L. Crowley. Integration and control of reactive visual processes. *Robotics and Autonomous Systems*, 15(1), 1995.

[7] J.L. Crowley, J. Coutaz, G. Rey, and P. Reignier. Using context to structure perceptual processes for observing activity. In *UBICOMP*, Sweden, September 2002.

[8] J.L. Crowley and P. Reignier. An architecture for context aware observation of human activity. In *Workshop on Computer Vision System Control Architectures (VSCA 2003)*, 2003.

[9] J.L. Crowley and O. Riff. Fast computation of scale normalised gaussian receptive fields. In *International Conference on Scalespace theories in Computer vision*, pages 584–598, Skye, UK, June 2003.

[10] J. Estublier, P. Y. Cunin, and N. Belkhatir. Architectures for process support interoperability. In *ICSP5*, 1997.

[11] D. Garlan et al. Rainbows: Architecture-based, self-adaptation with reusable infrastructure. *IEEE Computer*, 2004.

[12] W.T. Freeman and E.H. Adelson. The design and use of steerable filters. *Pattern Analysis and Machine Intelligence*, 13(9):891–906, September 1991.

[13] N. Gourier, D. Hall, and J.L. Crowley. Facial feature detection robust to pose, illumination, and identity. In *International Conference on Systems, Man and Cybernetics, Special track on Automatic Facial Expression Analysis*, pages 617–622, October 2004.

[14] D. Hall. Automatic parameter regulation for a tracking system with an auto-critical function. In *International Workshop on Computer Architecture for Machine Perception*, pages 39–45, Palermo, Italy, July 2005.

[15] D. Hall, V. Colin de Verdière, and J.L. Crowley. Object recognition using coloured receptive fields. In *European Conference on Computer Vision*, pages I 164–177, Dublin, Ireland, June 2000.

[16] D. Hall and J.L. Crowley. Computation of generic features for object classification. In *Scale Space Methods in Computer Vision*, pages 744–756, Skye, UK, June 2003.

[17] D. Hall and J.L. Crowley. Détection du visage par caractéristiques génériques calculées à partir des images de luminance. In *Reconnaissance des formes et intelligence artificelle*, Toulouse, France, 2004. to appear.

[18] D. Hall and J.L. Crowley. Détection du visage par caractéristiques génériques calculées à partir des images de luminance. In *Reconnaissance des formes et intelligence artificelle*, Toulouse, France, 2004. to appear.

[19] P. Horn. Autonomic computing; ibm's perspective on the state of information technology. http://researchweb.watson.ibm.com/autonomic, October 2001.

[20] J.O. Kephart and D.M. Chess. The vision of autonomic computing. *IEEE Computer*, 2003.

[21] R. Kjeldsen, A. Levas, and C. S. Pinhanez. Dynamically reconfigurable visionbased user interfaces. In *2003 International Conference on Vision Systems, ICVS03*, pages 257–267, Graz, Austria, April 2003.

[22] J. J. Koenderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[23] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[24] D.G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.

[25] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 2004.

[26] A. Lux. The imalab method for vision systems. In *International Conference on Vision Systems*, Graz, Austria, April 2003.

[27] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *International Conference on Computer Vision*, pages 525–531, Vancouver, Canada, July 2001.

[28] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *European Conference on Computer Vision*, Copenhague, 2002.

[29] A. Negre, C. Braillon, and J.L. Crowley. Visual navigation from variation of intrinsic scale. In *Submitted to ICRA*, 2006.

[30] A. Paar and W.F Tichy. Semantic software engineering approaches for automatic servicd lookup and integration. In 5^o *Workshop on Active Middleware Services*, pages 103–111, June 2003.

[31] F. Pélisson, D. Hall, O. Riff, and J.L. Crowley. Brand identification using gaussian derivative histograms. In *International Conference on Vision Systems*, pages 492–501, Graz, Austria, April 2003.

[32] G. Pingali, C. Pinhanez, A. Levas, and R. Kjeldsen. Steerable interfaces for pervasive computing spaces. In *IEEE PerCom*, March 2003.

[33] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Proceedings of Ubiquitous Computing 2001 Conference*, September 2001.

[34] V. Poladian. Dynamic configuration of resource-aware services. In *Int. Conf. on Software Engineering*, September 2001.

[35] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and 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.

[36] J. Rasure and S. Kubica. The khoros application development environment. In J.L. Crowley and H.I. Christensen, editors, *Experimental Environments for Computer Vision and Image Processing*, number 1(11) in Machine Perception Artificial Intelligence Series, pages 1–32. World Scientific Press, 1994.

[37] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, January 2000.

[38] M. Shaw and D. Garlan. *Software Architecture: Perspectives on an Emerging Disciplines*. Prentice Hall, 1996.

[39] N. Takao, J. Shi, and S. Baker. Tele graffiti: a camera projector based remote sketching system with hand based user interface and automatic session summarization. *Int. Journal of Computer Vision*, 53:115–133, 2003.

[40] T.H. Tran Thi and A. Lux. A method for ridge extraction. In *Asian Conference on Computer Vision*, pages 960–966, Jeju, Korea, 2004.

[41] Thi Thanh Hai TRAN. *Etude des lignes d'intérêt naturelles pour la représentation d'objets en vision par ordinateur*. Ph.d. thesis in computer sciences, INP Grenoble, Grenoble (France), March 2006.

[42] Thi Thanh Hai TRAN, Augustin LUX, and Hoang Lan 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.

[43] F. Vernier, N. Lesh, and C. Shen. Visualization techniques for circular tabletop interfaces. In *Advanced Visual Interfaces*, 2002.

[44] P. Wellner. The digitaldesk calculator: Tactile manipulation on a desk top display. In *ACM Symposium on User Interface Software and Technology (UIST '91)*, pages 27–33, November 1991.

[45] G. Ye, J. J. Corso, D. Burschka, and G. D. Hager. Vics: A modular vision-based hci framework. In *2003 International Conference on Vision Systems - ICVS03*, pages 257–267, Graz, Austria, April 2003.