Activity Report 2012

Project-Team PRIMA

Perception, recognition and integration for observation of activity

Table of contents

<div align="center">**Project-Team PRIMA**</div>

Keywords: Computer Vision, Interaction, Perception, Artificial Intelligence, Human-Robot Interaction

Creation of the Project-Team: January 01, 2003 .

1. Members

Research Scientists
Frédéric Devernay [CR1 Inria]
Thierry Fraichard [CR1 Inria, HdR]
Sabine Coquillart [DR2 Inria, HdR]

Faculty Members
James Crowley [Team Leader, Professeur CE2 Grenoble INP, HdR]
Augustin Lux [Professeur Grenoble INP, HdR]
Patrick Reignier [Professeur Grenoble INP, HdR]
Dominique Vaufreydaz [MdC Université Pierre Mendès-France]
Alexandre Demeure [MdC Université Joseph Fourier]

Engineers
Amaury Nègre [Engineer CNRS (Shared with E-Motion)]
Claudine Combe [Engineer CDD Inria]
Matthieu Volat [Engineer CDD Inria]
Sergi Pujades-Rocamora [Engineer CDD Inria]
Rémi Barraquand [Engineer CDD Inria]
Lukas Rummelhard [Engineer CDD Inria]
Lucas Nacsa [Engineer CDD Inria]

PhD Students
John Alexander Ruiz Hernandez [Doctoral Student (Bourse Inria)]
Antoine Meler [Doctoral Student (Bourse MENSR)]
Mathieu Guillame-Bert [Doctoral Student (Bourse MENSR)]
Marion Decrouez [Doctoral Student (Bourse CEA)]
Dimitri Masson [Doctoral Student (Bourse MENSR)]
Evanthia Mavridou [Doctoral Student (Bourse MESR)]
Julian Quiroga [Doctoral Student (Bourse Columbian)]
Varun Jain [Doctoral Student (Bourse Inria)]

Administrative Assistant
Marion Ponsot [Secretary (SAR) Inria]

Other
Wafa Benkaouar [Master's Student]

2. Overall Objectives

2.1. Perception, Recognition and Multimodal Interaction for Smart Spaces.

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

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

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

An environment may be considered as "interactive" when it is capable of interacting with humans using tightly coupled perception and action. Simple forms of interaction may be based on observing the manipulation of physical objects, or on visual sensing of fingers, hands or arms. Richer forms of interaction require perception and understanding of human activity and context. PRIMA has developed a novel theory for situation modeling for machine understanding of human activity, based on techniques used in Cognitive Psychology [51]. PRIMA explores multiple forms of interaction, including projected interaction widgets, observation of manipulation of objects, fusion of acoustic and visual information, and systems that model interaction context in order to predict appropriate action and services by the environment.

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

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

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

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

2.2. Highlights of the Year

Publication of a special issue on motion safety in the Autonomous Robot journal edited by Thierry Fraichard and James Kuffner [30].

3. Scientific Foundations

3.1. Context Aware Smart Spaces

Situation Models for Context Aware Systems and Services

3.1.1. *Summary*

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.

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. 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. Situations are organized into networks, with transition probabilities, so that possible next situations may be predicted from the current situation.

Current technology allows us to handcraft real-time systems for a specific services. The current hard challenge is to create a technology to automatically learn and adapt situation models with minimal or no disruption of human activity. An important current problem for the PRIMA group is the adaptation of Machine Learning techniques for learning situation models for describing the context of human activity.

3.1.2. *Detailed Description*

Context Aware Systems and Services require a model for how humans think and interact with each other and their environment. Relevant theories may be found in the field of cognitive science. Since the 1980's, Philippe Johnson-Laird and his colleagues have developed an extensive theoretical framework for human mental models [52], [53]. Johnson Laird's "situation models", provide a simple and elegant framework for predicting and explaining human abilities for spatial reasoning, game playing strategies, understanding spoken narration, understanding text and literature, social interaction and controlling behavior. While these theories are primarily used to provide models of human cognitive abilities, they are easily implemented in programmable systems [37], [36].

In Johnson-Laird's Situation Models, a situation is defined as a configuration of relations over entities. Relations are formalized as N-ary predicates such as beside or above. Entities are objects, actors, or phenomena that can be reliably observed by a perceptual system. Situation models provide a structure for organizing assemblies of entities and relations into a network of situations. For cognitive scientists, such models provide a tool to explain and predict the abilities and limitations of human perception. For machine perception systems, situation models provide the foundation for assimilation, prediction and control of perception. A situation model identifies the entities and relations that are relevant to a context, allowing the perception system to focus limited computing and sensing resources. The situation model can provide default information about the identities of entities and the configuration of relations, allowing a system to continue to operate when perception systems fail or become unreliable. The network of situations provides a mechanism to predict possible changes in entities or their relations. Finally, the situation model provides an interface between perception and human centered systems and services. On the one hand, changes in situations can provide events that drive service behavior. At the same time, the situation model can provide a default description of the environment that allows human-centered services to operate asynchronously from perceptual systems.

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

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

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

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

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

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

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

Currently situation models are constructed by hand. Our current challenge is to provide a technology by which situation models may be adapted and extended by explicit and implicit interaction with the user. An important aspect of taking services to the real world is an ability to adapt and extend service behaviour to accommodate individual preferences and interaction styles. Our approach is to adapt and extend an explicit model of user activity. While such adaptation requires feedback from users, it must avoid or at least minimize disruption. We are curently exploring reinforcement learning approaches to solve this problem.

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

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

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

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

3.2. Service Oriented Architectures for Intelligent Environments

Software Architecture, Service Oriented Computing, Service Composition, Service Factories, Semantic Description of Functionalities

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

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

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

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

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

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

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

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

The knowledge that a service exposes a functionality like a "Timer" functionality for a service emitting message at a regular frequency.

The knowledge that a kind of functionality can be converted to another one. For example, a "Metronome" functionality issued from a music centered application can be seen as a "Timer" functionality.

The knowledge that a particular service is a factory and can instantiate other services on demand. A TimerFactory can for example start a new service with a "Timer" functionality with any desired frequency. Factories greatly helps in the deployment of service based applications. UFCL factories can also express the fact that they can compose existing functionalities to provide another one.

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

3.3. Robust view-invariant Computer Vision

Local Appearance, Affine Invariance, Receptive Fields

3.3.1. *Summary*

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

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

3.3.2. *Detailed Description*

The visual appearance of a neighbourhood can be described by a local Taylor series [54]. 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 [56] 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 [62]. 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 [56], and normalized in orientation using direction of the first derivatives [62].

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" [57]. 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" [47] of Gaussian derivatives. The problem is to determine the appropriate orientation. In earlier work by PRIMA members Colin de Verdiere [34], Schiele [62] and Hall [50], 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 [50]. These filters transform an (R,G,B), into a cartesian representation for luminance and chrominance (L,C1,C2). Chromatic Gaussian receptive fields are computed by applying the Gaussian derivatives independently to each of the three components, (L, C1, C2). The components C1 and C2 encodes the chromatic information in a Cartesian representation, while L is the luminance direction. Chromatic Gaussian receptive fields are computed by applying the Gaussian derivatives independently to each of the three components, (L, C1, C2). Permutations of RGB lead to different opponent color spaces. The choice of the most appropriate space depends on the chromatic composition of the scene. An example of a second order steerable chromatic basis is the set of color opponent filters shown in figure 1.
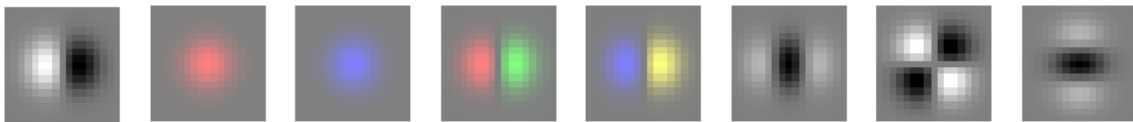
Figure 1. Chromatic Gaussian Receptive Fields ($G_x^L, G^{C_1}, G^{C_2}, G_x^{C_1}, G_x^{C_2}, G_{xx}^L, G_{xy}^L, G_{yy}^L$).

Key results in this area include

Fast, video rate, calculation of scale and orientation for image description with normalized chromatic receptive fields [37].

Real time indexing and recognition using a novel indexing tree to represent multi-dimensional receptive field histograms [60].

Robust visual features for face tracking [49], [48].

Affine invariant detection and tracking using natural interest lines [63].

Direct computation of time to collision over the entire visual field using rate of change of intrinsic scale [58].

We have achieved video rate calculation of scale and orientation normalised Gaussian receptive fields using an O(N) pyramid algorithm [37]. This algorithm has been used to propose an embedded system that provides real time detection and recognition of faces and objects in mobile computing devices.

Applications have been demonstrated for detection, tracking and recognition at video rates. This method has been used in the MinImage project to provide real time detection, tracking, and identification of faces. It has also been used to provide techniques for estimating age and gender of people from their faces

3.4. Perception for Social Interaction

Affective Computing, Perception for social interaction.

Current research on perception for interaction primarily focuses on recognition and communication of linguistic signals. However, most human-to-human interaction is non-verbal and highly dependent on social context. A technology for natural interaction will require abilities to perceive and assimilate non-verbal social signals, to understand and predict social situations, and to acquire and develop social interaction skills.

The overall goal of this research program is to provide the scientific and technological foundations for systems that observe and interact with people in a polite, socially appropriate manner. We address these objectives with research activities in three interrelated areas:

>   Multimodal perception for social interactions.

>   Learning models for context aware social interaction, and

>   Context aware systems and services.

Our approach to each of these areas is to draw on models and theories from the cognitive and social sciences, human factors, and software architectures to develop new theories and models for computer vision and multi-modal interaction. Results will be developed, demonstrated and evaluated through the construction of systems and services for polite, socially aware interaction in the context of smart habitats.

3.4.1. Detailed Description

First part of our work on perception for social interaction has concentrated on measuring the physiological parameters of Valence, Arousal and Dominance using visual observation form environmental sensors as well as observation of facial expressions.

People express and feel emotions with their face. Because the face is the both externally visible and the seat of emotional expression, facial expression of emotion plays a central role in social interaction between humans. Thus visual recognition of emotions from facial expressions is a core enabling technology for any effort to adapt ICT for social interaction.

Constructing a technology for automatic visual recognition of emotions requires solutions to a number of hard challenges. Emotions are expressed by coordinated temporal activations of 21 different facial muscles assisted by a number of additional muscles. Activations of these muscles are visible through subtle deformations in the surface structure of the face. Unfortunately, this facial structure can be masked by facial markings, makeup, facial hair, glasses and other obstructions. The exact facial geometry, as well as the coordinated expression of muscles is unique to each individual. In additions, these deformations must be observed and measured under a large variety of illumination conditions as well as a variety of observation angles. Thus the visual recognition of emotions from facial expression remains a challenging open problem in computer vision.

Despite the difficulty of this challenge, important progress has been made in the area of automatic recognition of emotions from face expressions. The systematic cataloging of facial muscle groups as facial action units by Ekman [45] has let a number of research groups to develop libraries of techniques for recognizing the elements of the FACS coding system [33]. Unfortunately, experiments with that system have revealed that the system is very sensitive to both illumination and viewing conditions, as well as the difficulty in interpreting the resulting activation levels as emotions. In particular, this approach requires a high-resolution image with a high signal-to-noise ratio obtained under strong ambient illumination. Such restrictions are not compatible with the mobile imaging system used on tablet computers and mobile phones that are the target of this effort.

As an alternative to detecting activation of facial action units by tracking individual face muscles, we propose to measure physiological parameters that underlie emotions with a global approach. Most human emotions can be expressed as trajectories in a three dimensional space whose features are the physiological parameters of Pleasure-Displeasure, Arousal-Passivity and Dominance-Submission. These three physiological parameters can be measured in a variety of manners including on-body accelerometers, prosody, heart-rate, head movement and global face expression.

In our work, we address the recognition of social behaviors multimodal information. These are unconscious inmate cognitive processes that are vital to human communication and interaction. Recognition of social behaviors enables anticipation and improves the quality of interaction between humans. Among social behaviors, we have focused on engagement, the expression of intention for interaction. During the engagement phase, many non-verbal signals are used to communicate the intention to engage to the partner [65]. These include posture, gaze, spatial information, gestures, and vocal cues.

For example , within the context of frail or elderly people at home, a companion robot must also be able to detect the engagement of humans in order to adapt their responses during interaction with humans to increase their acceptability. Classical approaches for engagement with robots use spatial information such as human position and speed, human-robot distance and the angle of arrival. Our believe that, while such uni-modal methods may be suitable for static display [66] or robots in wide space area [55] they are not sufficient for home environments. In an apartment, relative spatial information of people and robot are not as discriminative as in an open space. Passing by the robot in a corridor should not lead to an engagement detection, and possible socially inappropriate behavior by the robot.

In our experiments, we use a kompai robot from Robosoft [32]. As an alternative to wearable physiological sensors (such as pulse bracelet Cardiocam, etc.) we inegrate multimodal features using a Kinect sensor (see figure 5). In addition of the spatial cues from the laser telemeter, one can use new multimodal features based on persons and skeletons tracking, sound localization, etc. Some of these new features are inspired from results in cognitive science domain [61].

Our multimodal approach has been confronted to a robot centered dataset for multimodal social signal processing recorded in a home-like environment [24]. The evaluation on our corpus highlights its robustness and validates use of such technique in real environment. Experimental validation shows that the use of multimodal sensors gives better results than only spatial features (50% of error reduction). Our experimentations also confirm results from [61]: relative shoulder rotation, speed and facing visage are among crucial features for engagement detection.

3.5. End Users control over Smart Environment

Missing keywords.

Ubiquitous computing promises unprecedented empowerment from the flexible and robust combination of software services with the physical world. Software researchers assimilate this promise as system autonomy where users are conveniently kept out of the loop. Their hypothesis is that services, such as music playback and calendars, are developed by service providers and pre-assembled by software designers to form new service frontends. Their scientific challenge is then to develop secure, multiscale, multi-layered, virtualized infrastructures that guarantee service front-end continuity. Although service continuity is desirable in many circumstances, end users, with this interpretation of ubiquitous computing, are doomed to behave as mere consumers, just like with conventional desktop computing.

Another interpretation of the promises of ubiquitous computing, is the empowerment of end users with tools that allow them to create and reshape their own interactive spaces. Our hypothesis is that end users are willing to shape their own interactive spaces by coupling smart artifacts, building imaginative new functionalities that were not anticipated by system designers. A number of tools and techniques have been developed to support this view such as CAMP [64] or iCAP [44].

We adopt a End-User Programming (EUP) approach to give the control back to the inhabitants. In our vision, smart Homes will be incrementally equiped with sensors, actuators and services by inhabitants themselves. Our research programm therefore focus on tools and languages to enable inhabitants in activities related to EUP for Smart Homes :

> Installation and maintenance of devices and services. This may imply having facilities to attribute names.
> Visualizing and controling of the Smart Habitat.

Programming and testing. This imply one or more programming languages and programming environment which could rely on the previous point. The programming language is especially important. Indeed, in the context of the Smart Homes, End-User Programms are most likely to be routines in the sens of [38] than procedure in the sens of traditionnal programming languages.

Detecting and solving conflicts related to contradictory programs or goals.

4. Application Domains

4.1. The Smart Spaces Research Plateform

Participants: Rémi Barraquand, James Crowley [correspondant], Augustin Lux, Amaury Nègre, Patrick Reignier, Lukas Rummelhard, Dominique Vaufreydaz.

Smart Spaces, Observation of human activity, context aware systems and services.

Project PRIMA has recently moved to a new Smart Spaces Research Plateform in order to develop and test components and services for context aware human centered services. The Smart Spaces Research Plateform is a 50 Square Meter space equipped with a microphone array, wireless lapel microphones, wide angle surveillance cameras, panoramic cameras, steerable cameras, scanning range sensors and two camera-projector video-interaction devices, and a KNX smart electrical system. The microphone array is used as an acoustic sensor to detect, locate and classify acoustic signals for recognizing human activities. The wide-angle and panoramic cameras provide fields of view that cover the entire room, and allows detection and tracking of individuals. Steerable cameras are used to acquire video of activities from any viewing direction.

Context aware human centered services may categorized as tools, advisors, or media. Tool services are designed to perform a specific task or function as robustly as possible. If any adaptation is involved, it should serve to adapt the function to a changing environment. The user interface, and any interaction with users should be perfectly predictable. The degree to which the operation of a tool should be transparent, visible or hidden from the user is an open research question. Advisor services observe the users actions and environment in order to propose information on possible courses of actions. Advisors should be completely obedient and non-disruptive. They should not take initiatives or actions that cannot be overridden or controlled by the user. Media services provide interpersonal communications, entertainment or sensorial extension.

Examples of human centered tool services include:

An activity log recording system that records the events and activities of an individual's daily activities.

A service that integrates control of heating, air-conditioning, lighting, windows, window-shades, exterior awnings, etc to provide an optimum comfort level defined in terms of temperature, humidity, CO_2, acoustic noise and ambient light level.

A service that manages the available stock of supplies in a home and orders supplies over the Internet to assure that the appropriate level of supplies are always available.

A service to measure the walking rate, step size and posture of an elderly person to estimate health and predict the likelihood of a fall.

Some examples of advisor services include:

A service that provides shopping advice about where and when to shop.

A service that can propose possible menus based on the available food stuffs in the kitchen.

A service that observes the activities of humans and appliances within the home and can suggest ways to reduce the cost of heating, electricity or communications.

A service that observes lifestyle and can offer advice about improving health.

Some examples of media services include

A service that maintains a sense of informal non-disruptive presence with distant family members.

A robot device that communicates affection.

A device that renders the surface temperature of wall, floors and windows to show energy consumption and loss within a house.

Services that enable seamless tele-presence for communication with others

4.2. Ambient Services for Assisted Living

Participants: Rémi Barraquand, James Crowley, Frédéric Devernay, Amaury Nègre.

Ambient Assisted Living, Monitoring Services, Presence awareness.

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

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

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

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

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

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

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

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

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

4.3. 3-D video processing

Participants: Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

3-D display, Stereoscopy, View Interpolation, Auto-calibration

Stereoscopic cinema has seen a surge of activity in recent years, and for the first time all of the major Hollywood studios released 3-D movies in 2009. This is happening alongside the adoption of 3-D technology for sports broadcasting, and the arrival of 3-D TVs for the home. Two previous attempts to introduce 3-D cinema in the 1950s and the 1980s failed because the contemporary technology was immature and resulted in viewer discomfort. But current technologies such as accurately-adjustable 3-D camera rigs with onboard computers to automatically inform a camera operator of inappropriate stereoscopic shots, digital processing for post-shooting rectification of the 3-D imagery, digital projectors for accurate positioning of the two stereo projections on the cinema screen, and polarized silver screens to reduce cross-talk between the viewers left- and right-eyes mean that the viewer experience is at a much higher level of quality than in the past. Even so, creation of stereoscopic cinema is an open, active research area, and there are many challenges from acquisition to post-production to automatic adaptation for different-sized display [41], [43].

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

When the 3-D scene which has to be displayed is computer-generated, it is usually not a problem to generate a few dozen viewpoints. But when a real scene has to displayed, one would have to shoot it through the same number of synchronized cameras as there are viewpoints in order to display it properly. This makes 3-D shooting of real scenes for glasses-free 3-D displays mostly unpractical. For this reason, we are developing high-quality view-interpolation techniques, so that the many different viewpoints can be generated from only a few camera positions [42].

Our research focuses on algorithms derived from Computer Vision and Computer Graphics, applied to live-action stereoscopic 3-D content production or post-production, including [40]:

> Live monitoring of stereoscopic video: geometric image misalignment, depth budget (i.e. limits on horizontal disparity), left-right color balance, left-right depth-of-field consistency [59].
>
> Live correction of stereoscopic video: correct the above defects in real-time when it is possible, with the help of GPU-based architectures.
>
> Adaptation of the stereoscopic content to the display size and distance, to avoid divergence or geometric deformations [42].
>
> Novel camera setups and algorithms for unconstrained stereoscopic shooting (especially when using long focal length).
>
> Novel camera setups and algorithms for glasses-free 3D displays.
>
> Stereoscopic inpainting.
>
> Stereoscopic match-moving.
>
> Compositing stereoscopic video and matte painting without green screen.
>
> Relighting of stereoscopic video, especially when videos are composited.

4.4. User localization in large-scale Smart Spaces

Participants: Amaury Nègre, Lukas Rummelhard, Dominique Vaufreydaz [correspondant].

Multi-modal perception, Smart Spaces, Localisation.

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

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

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

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

Observing and tracking people in smart environments remains a challenging fundamental problem. Whether it is at the scale of a campus, of a building or more simply of a room, we can combine several additional localization levels (and several technologies) to allow a more accurate and reliable user perception system. Within the PRIMA project, we are currently experimenting with a multi-level localization system allowing variable granularity according to the available equipment and the precision required for the targeted service.

5. Software

5.1. OMiSCID Middleware for Distributed Multimodal Perception

Participants: Rémi Barraquand, Amaury Nègre, Patrick Reignier, Dominique Vaufreydaz [correspondant].

Middleware, Distributed perceptual systems

OMiSCID is new lightweight middleware for dynamic integration of perceptual services in interactive environments. This middleware abstracts network communications and provides service introspection and discovery using DNS-SD (*DNS-based Service Discovery* [31]). Services can declare simplex or duplex communication channels and variables. The middleware supports the low-latency, high-bandwidth communications required in interactive perceptual applications. It is designed to allow independently developed perceptual components to be integrated to construct user services. Thus our system has been designed to be cross-language, cross-platform, and easy to learn. It provides low latency communications suitable for audio and visual perception for interactive services.

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

The OMiSCID graphical user interface (GUI) is an extensible graphical application that facilitates analysis and debugging of service oriented applications. The core functionality of this GUI is to list running services, their communication channels and their variables. This GUI is highly extensible and many modules (i.e. plugins) have been created by different members of the team: figure 2 shows an example of some of these modules. OMiSCID GUI is based on the Netbeans platform and thus inherits from its dynamic installation and update of modules.

5.2. Detection and Tracking of Pedestrians in INRETS Intelligent Urban Spaces Platform

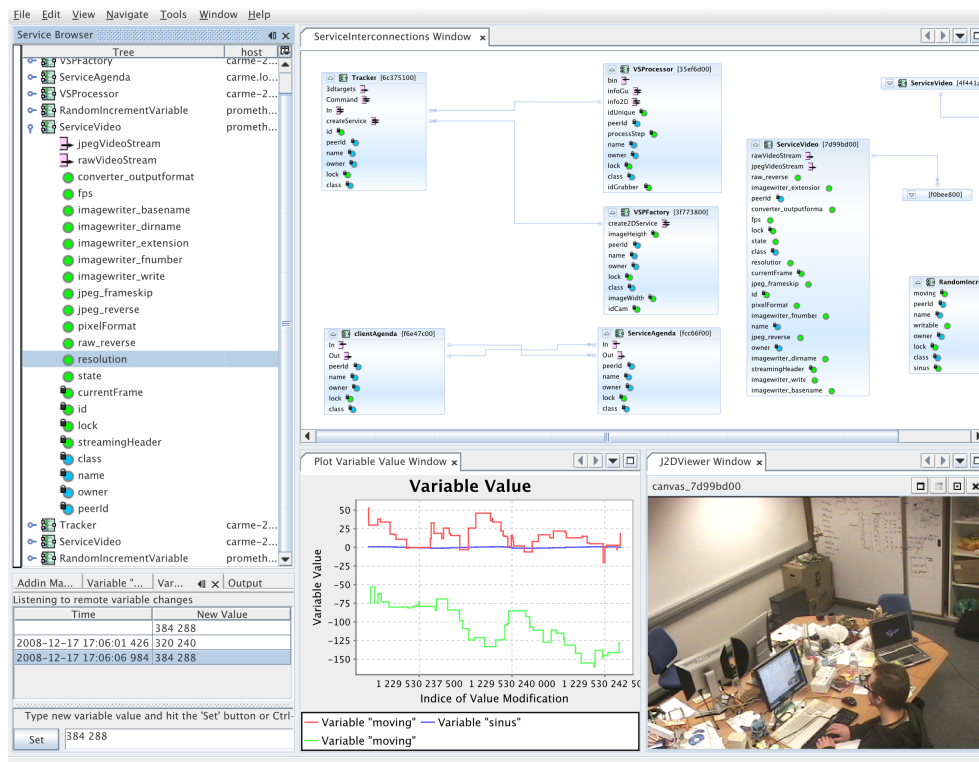**Participants:** Claudine Combe, James Crowley [correspondant], Lukas Rummelhard.

Figure 2. OMiSCID GUI showing a list of running services and some modules for service interconnections, variable plotting, live video stream display and variable control
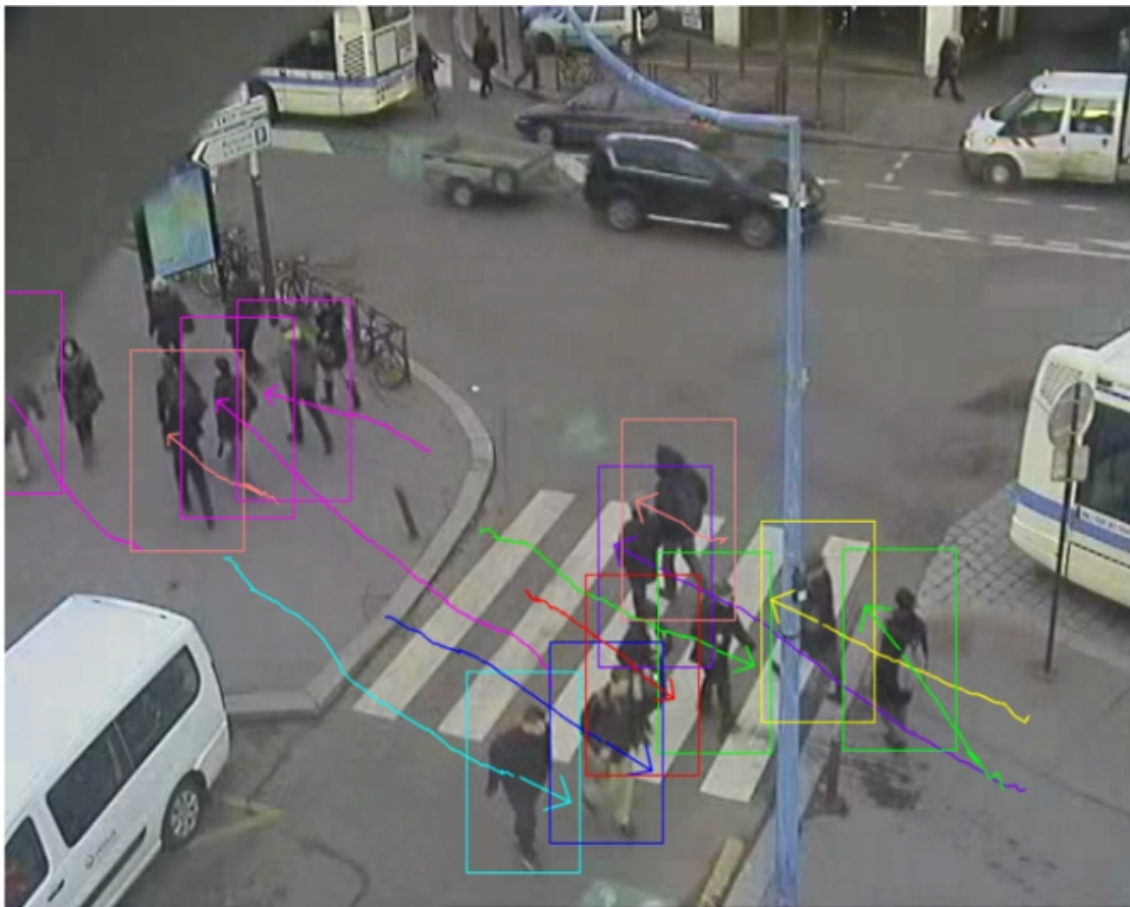
Figure 3. Cipebus: pedestrian tracking system.

Visual detection and tracking of pedestrians, Intelligent Urban Space

The project ANR-07-TSFA-009-01 CIPEBUS ("Carrefour Intelligent - Pole d'Echange - Bus) has been proposed by INRETS-IFSTTAR, in collaboration with Inria, Citilog, Fareco, and the city of Versaille. The Objective of the CIPEBUS project is to develop an experimental platform for observing activity in a network of urban streets in order to experiment with techniques for optimizing circulation by context aware control of traffic lights.

Within CipeBus, Inria jas developed a real time multi-camera computer vision system to detect and track people using a network of surveillance cameras. The CipeBus combines real time pedestrian detection with 2D and 3D Bayesian tracking to record the current position and trajectory of pedestrians in an urban environment under natural view conditions. The system extends the sliding window approach to use a half-octave Gaussian Pyramid to explore hypotheses of pedestrians at different positions and scales. A cascade classifier is used to determine the probability that a pedestrian can be found at a particular position and scale. Detected pedestrians are then tracked using a particle filter.

The resulting software system has been installed and tested at the INRETS CipeBus platform and is currently used for experiments in controlling the traffic lights to optimize the flow of pedertrians and public transportation while minimizing the delay imposed on private automobiles.

5.3. Multisensor observation of human activity for integrated energy and comfort management

Participants: Claudine Combe, James Crowley [correspondant], Lucas Nacsa, Amaury Nègre, Lukas Rummelhard.
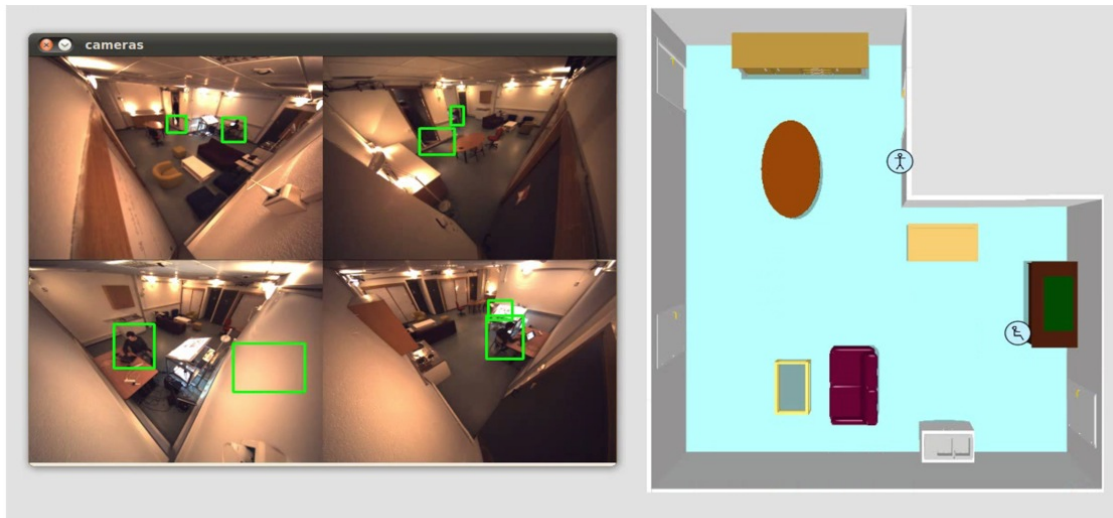
multimodal tracking of human activity

Figure 4. The 3D tracker integrates observations from multiple sensors

As part of Inria's contribution of ICTLabs Action TSES - Smart Energy Systems, we have constructed a system that integrates information from multiple environmental sensor to detect and track people in indoor environments. This system, constructed as part of activity 11831 Open SES Experience Labs for Prosumers and New Services, has been released to ICTLabs partners in June 2012. It has also been used for construction of a smart spaces testbed at Schneider Electric.

This software, named MultiSensor activity tracker, integrates information from multiple environmental sensors to keep track of the location and activity of people in a smart environment. This model is designed to be used by a home energy broker that would work in conjunction with a smart grid to manage the energy consumption of home appliances, balancing the needs of inhabitants with opportunities for savings offered by electricity rates. This database will also be used for by advisor services that will offer advice to inhabitants on the consequences to energy consumption and energy cost that could potentially result from changes to lifestyle or home energy use.

Work in this task draws from earlier result from a number of development projects at Inria. In the ANR Casper project Inria created Bayesian tracking system for human activity using a voxel based occupancy grid. Within the INRA ADT PAL project, Inria is creating methods for plug and play installation of visual and acoustic sensors for tracking human activity within indoor environments.

While a voxel based Bayesian tracker has served well for a number of applications, a number of limitations have been observed. For example, under certain circumstances, the sensor data can provide contradictory or ambiguous data about the location and activities of people. Resolving such cases required the Bayesian tracker to choose between a numbers of competing hypotheses, potentially resulting in errors. Several members of our group have argued that an alternative integration approach based on the use of a Particle filter would solve these problems and provide a more reliable tracking system. This task has been undertaken to evaluate this hypothesis. The system configured and optimized for detecting and tracking people within rooms using multiple calibrated cameras. The system currently uses corner mounted cartesian cameras, ceiling mounted cameras with wide angle lenses and panoramic cameras placed on tables. Cameras may be connected and disconnected while the component is running, but they must be pre-calibrated to a common room reference frame. We are currently experimenting with techniques for Bayesian estimation of camera parameters for auto-calibration. Cameras may be connected dynamically.

The original system 3DBT has been declared with the APP "Agence pour la Protection des Programmes" under the Interdeposit Digital number IDDN.FR.001.490023.000.S.P.2006.000.10000. A revised declaration for the latest version of the system is currently being prepared.

5.4. Stereo Viewfinder

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

Stereoscopy, Auto-calibration, Real-time video processing, Feature matching

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

5.5. Tracking Focus of Attention for Large Screen Interaction

Participants: Rémi Barraquand, Claudine Combe, James Crowley [correspondant], Varun Jain, Sergi Pujades-Rocamora, Lukas Rummelhard.

Embedded Detection and Tracking of Faces for AttentionEstimation.

Large multi-touch screens may potentially provide a revolution in the way people can interact with information in public spaces. Technologies now exist to allow inexpensive interactive displays to be installed in shopping areas, subways and urban areas. Thesis displays can provide location aware access to information including maps and navigation guidance, information about local businesses and and commercial activities. While location information is an important component of a users context, information about the age and gender of a user, as well as information about the number of users present can greatly enhance the value of such interaction for both the user and for local commerce and other activities.

The objective of this task is to leverage recent technological advances in real time face detection developed for cell phones and mobile computing to provide a low-cost real time visual sensor for observing users of large multi-touch interactive displays installed in public spaces.

People generally look at things that attract their attention. Thus it is possible to estimate the subject of attention by estimating where people look. The location of visual attention is manifested by a region of space known as the horopter where the optical axis of the two eyes intersect. However estimating the location of attention from human eyes is notoriously difficult, both because the eyes are small relative to the size of the face, and because eyes can rotate in their socket with very high accelerations. Fortunately, when a human attends to something, visual fixation tends to remain at or near that subject of attention, and the eyes are relaxed to a symmetric configuration by turning the face towards the subject of attention. Thus it is possible to estimate human attention by estimating the orientation of the human face.

We have constructed an embedded software system for detecting, tracking and estimating the orientation of human faces. This software has been designed to be embedded on mobile computing devices such as laptop computers, tablets and interactive display panels equipped with a camera that observes the user. Noting the face orientation with respect to the camera makes it possible to estimate the region of the display screen to which the user is attending.

The system uses a Bayesian Particle filter tracker operating on a Scale invariant Gaussian pyramid to provide integrated tracking and estimation of face orientation. The use of Bayesian tracking greatly improves both the reliability and the efficiency for face detection and orientation estimation. The scale invariant Gaussian pyramid provides automatic adaptation to image scale (as occurs with a change in camera optics) and makes it possible to detect and track faces over a large range of distances. Equally important the Gaussian Pyramid provides a very fast computation of a large number of image features that can be used by a variety of image analysis algorithms.

The software developed for this activity builds on face detections software that has recently been developed by Inria for the French OSEO project MinImage. MinImage was a five year, multi-million euro project to develop next generation technologies for integrated digital imaging devices to be used in cellphones, mobile and laptop computing devices, and digital cameras, that has begun in February of 2007. The project scope included research on new forms of retinas, integrated optics, image formation and embedded image processing. Inria was responsible for embedded algorithms for real time applications of computer vision.

Within MinImage, Inria developed embedded image analysis algorithms using image descriptors that are invariant to position, orientation and scale and robust to changes in viewing angle and illumination intensity. Inria proposed use of a simple hardware circuit to compute a scale invariant Gaussian pyramid as images acquired by the retina. Sums and differences of image samples from the pyramid provide invariant image descriptors that can be used for a wide variety of computer vision applications including detection, tracking and recognition of visual landmarks, physical objects, commercial logos, human bodies and human faces. Detection and tracking of human faces was selected as benchmark test case.

This work has been continued with support from EIT ICTlabs, to provide context information for interaction with large multi-touch interactive displays installed in public spaces.

Multitouch interactive displays are increasingly used in outdoor and public spaces. This objective of this task is to provide a visual observation system that can detect and count users of a multitouch display and to estimate information such as the gender, and age category of each user. us rendering the system sensitive to environmental context.

A revised software package has recently been released to our ICTlab partners for face detection, face tracking, gender and age estimation, and orientation estimation, as part of ICTlabs Smart Spaces action line, Activity 11547 : Pervasive Information interfaces and interaction. With Task 1207 of this activity we have constructed and released an "Attention Recognition Module". This software has been protected with an APP declaration.

An similar software was released in 2007 using face color rather than appearance. The system SuiviDeCiblesCouleur located individuals in a scene for video communications. FaceStabilsationSystem renormalised

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

5.6. Visual Emotion Recognition for Health and Well Being.

Participants: Rémi Barraquand, Claudine Combe, James Crowley [correspondant], Varun Jain, Sergi Pujades-Rocamora, Lukas Rummelhard.

Visual Emotion Recognition

People express and feel emotions with their face. Because the face is the both externally visible and the seat of emotional expression, facial expression of emotion plays a central role in social interaction between humans. Thus visual recognition of emotions from facial expressions is a core enabling technology for any effort to adapt ICT to improve Health and Wellbeing.

Constructing a technology for automatic visual recognition of emotions requires solutions to a number of hard challenges. Emotions are expressed by coordinated temporal activations of 21 different facial muscles assisted by a number of additional muscles. Activations of these muscles are visible through subtle deformations in the surface structure of the face. Unfortunately, this facial structure can be masked by facial markings, makeup, facial hair, glasses and other obstructions. The exact facial geometry, as well as the coordinated expression of muscles is unique to each individual. In additions, these deformations must be observed and measured under a large variety of illumination conditions as well as a variety of observation angles. Thus the visual recognition of emotions from facial expression remains a challenging open problem in computer vision.

Despite the difficulty of this challenge, important progress has been made in the area of automatic recognition of emotions from face expressions. The systematic cataloging of facial muscle groups as facial action units by Ekman [45] has let a number of research groups to develop libraries of techniques for recognizing the elements of the FACS coding system [33]. Unfortunately, experiments with that system have revealed that the system is very sensitive to both illumination and viewing conditions, as well as the difficulty in interpreting the resulting activation levels as emotions. In particular, this approach requires a high-resolution image with a high signal-to-noise ratio obtained under strong ambient illumination. Such restrictions are not compatible with the mobile imaging system used on tablet computers and mobile phones that are the target of this effort.

As an alternative to detecting activation of facial action units by tracking individual face muscles, we propose to measure physiological parameters that underlie emotions with a global approach. Most human emotions can be expressed as trajectories in a three dimensional space whose features are the physiological parameters of Pleasure-Displeasure, Arousal-Passivity and Dominance-Submission. These three physiological parameters can be measured in a variety of manners including on-body accelerometers, prosody, heart-rate, head movement and global face expression.

The PRIMA Group at Inria has developed robust fast algorithms for detection and recognition of human faces suitable for use in embedded visual systems for mobile devices and telephones. The objective of the work described in this report is to employ these techniques to construct a software system for measuring the physiological parameters commonly associated with emotions that can be embedded in mobile computing devices such as cell phones and tablets.

As part of Inria's contribution to ICT labs Action THWB Health and Wellbeing, Inria has participated in Activity 12100 "Affective Computing". In this activity we have provided a software system for detection, tracking of faces, and for visual measurement of Valence, Arousal and Dominance.

A software library, named PrimaCV has been designed, debugged and tested, and released to ICTLabs partners for real time image acquisition, robust invariant multi-scale image description, highly optimized face detection, and face tracking. This software has been substantially modified so as to run on an mobile computing device using the Tegra 3 GPU.

6. New Results

6.1. Social behaviors recognition

Participants: Wafa Benkaouar, Claudine Combe, Dominique Vaufreydaz [correspondant].

Figure 5. On the left image, one can see the telemeter range in red, the foot detection (blue spot) and the angle view from the Kinect (in green). the middle and right image represent RGB camera en depth view from the Kinect.

Recognition of social behaviors is an unconscious innate cognitive process vital to human communication. This skill enables anticipation and increases interactive exchanges quality between humans. Among social behaviors, engagement is the expression of intention for interaction. During engagement phase, many non-verbal signals are used to communicate this intention to the partner, e.g. posture, gaze, spatial information, gestures, vocal cues. Within the context of frail or elderly people at home, companion robots must also be able to detect the engagement of humans in order to adapt their responses during interaction with humans to increase their acceptability.

Classical approaches in the domain are dealing with spatial information. Our hypothesis was that relative spatial information of people and robot are not discriminative in a home-like environment [15]. Our approach integrates multimodal features gathered using a robot companion equipped with a Kinect from Microsoft (see figure 5). Confronted to a robot centered dataset for multimodal social signal processing recorded in a home-like environment, the evaluation highlights its robustness and validates use of such technique in real environment (50% of error reduction). Our experimentations also confirm results from cognitive science domain [61].

6.2. Live monitoring and correction of 3DTV broadcasts

Participants: Pierre Arquier, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora, Matthieu Volat.

6.2.1. 3D broadcast monitoring and correction:

One of the achievements of the 3DLive FUI project was the transfer of real-time 3D video monitoring and correction algorithms to the Binocle company, and their integration into the TaggerLive product, which was used during several 3DTV broadcasts between 2010 and 2012 for live monitoring and correction of stereoscopic video. The algorithms that were developed within the PRIMA team and transferred into the TaggerLive are:

Multiscale view-invariant feature detection and matching on the GPU.

Computation of a temporally smooth and robust correction (or rectification) to remove the vertical disparity in the stereoscopic video while keeping the image aspect.

Real-time monitoring of the "depth budget", or the histogram of the horizontal disparity;

Live alerts when stereoscopic production rules are broken, such as when the disparities are too large, or when there is a stereoscopic window violation.

Real-time implementation of a state-of-the-art dense stereo matching method on the GPU.

6.2.2. 3D content adaptation:

3D shape perception in a stereoscopic movie depends on several depth cues, including stereopsis. For a given content, the depth perceived from stereopsis highly depends on the camera setup as well as on the display size and distance. This can lead to disturbing depth distortions such as the cardboard effect or the puppet theater effect. As more and more stereoscopic 3D content is produced in 3D (feature movies, documentaries, sports broadcasts), a key point is to get the same 3D experience on any display. For this purpose, perceived depth distortions can be resolved by performing view synthesis. We have proposed [19] a real time implementation of a stereoscopic player based on the open-source software Bino, which is able to adapt a stereoscopic movie to any display, based on user-provided camera and display parameters.

6.2.3. Focus mismatch detection:

Live-action stereoscopic content production requires a stereo rig with two cameras precisely matched and aligned. While most deviations from this perfect setup can be corrected either live or in post-production, a difference in the focus distance or focus range between the two cameras will lead to unrecoverable degradations of the stereoscopic footage. We have developed a method [18] to detect focus mismatch between views of a stereoscopic pair in four steps. First, we compute a dense disparity map. Then, we use a measure to compare focus in both images. After this, we use robust statistics to find which images zones have a different focus. Finally, to give useful feedback, we show the results on the original images and give hints on how to solve the focus mismatch.

6.3. Simultaneous localization and mapping (SLAM)

Participants: James Crowley, Marion Decrouez, Frédéric Devernay.

Localisation, place recognition, object recognition. Live processing of a video sequence taken from a single camera enables to model an a priori unknown 3D scene. Metrical SLAM (Simultaneous Localization and Mapping) algorithms track the camera pose while reconstructing a sparse map of the visual features of the 3D environment. Such approaches provide the geometrical foundation for many augmented reality applications in which informations and virtual objects are superimposed on live images captured by a camera. Improving such systems will enable in the future precise industrial applications such as guided-maintenance or guided-assembly in wide installations.

A problem with current methods is the assumption that the environment is static. Indoor environments such as supermarket ailes and factory floors may contain numerous objects that are likely to be moved, disrupting a localization and mapping system. We explore methods for automatic detection and modeling of such objects. We define the scene as a static structure that may contain moving objects and objects are defined as a set of visual features that share a common motion compared to the static structure [39]. Using several explorations of a camera in the same scene, we detect and model moved objects while reconstructing the environment. Experiments highlignt the performance of the method in a real case of localization in an unknown indoor environment.

6.4. Post-production tools for 3-D Cinema

Participants: Laurent Boiron, Frédéric Devernay [correspondant], Sylvain Duchêne, Sergi Pujades-Rocamora.

Over the past 6 years, we have been developing 3D movie processing techniques which have been used for the production and post-production of 3D movies (mainly feature-length movies, documentaries and commercials). These include image alignment, view interpolation, depth map computation, etc. These algorithms were developed as C++ libraries, and can be executed using standalone tools. Since the movie post-production workflow relies mainly on standard tools for compositing, color grading, etc., and these tools can be extended by plugin mechanisms, we integrated our post-production algorithms into such a tool, namely Nuke by The Foundry.

We also developed a new method for stereoscopic video cut and paste. Video cut-and-paste consists in semi-interactively segmenting a video object from a video stream, and pasting the segmented video object in another video. The object segmentation is done using a small number of strokes made on a few frames of the video, and can be corrected interactively. Existing methods only worked on monoscopic videos, and extending it to stereoscopic videos required solving important challenges:

The video object must not only remain consistent over time, but also between the left and right views.

The video object may be partially occluded in one or both views.

The camera setup may be different between the first and the second video, causing depth distortion or different depth effects.

We solved the first two challenges by adding left-right stereo consistency based on dense stereo matching, as well as temporal consistency based on optical flow, in an optimization framework based on graph cuts. The user interface was also taken into consideration in the algorithm: any correction of the results (i.e. new strokes on an image) will only propagate forward in time.

6.5. Scene flow computation from RGBZ data

Participants: Frédéric Devernay [correspondant], Julian Quiroga.

The scene flow describes the motion of each 3D point between two times steps. With the arrival of new depth sensors, as the Microsoft Kinect, it is now possible to compute scene flow with a single camera, with promising repercussion in a wide range of computer vision scenarios. We proposed [22] a novel method to compute scene flow by tracking in a Lucas-Kanade framework. Scene flow is estimated using a pair of aligned intensity and depth images, but rather than computing a dense scene flow as in most previous methods, we get a set of 3D motion fields by tracking surface patches. Assuming a 3D local rigidity of the scene, we propose a rigid translation flow model that allows to solve directly for the scene flow by constraining the 3D motion field both in intensity and depth data. In our experimentation we achieve very encouraging results. Since this approach solves simultaneously for the 2D tracking and for the scene flow, it can be used for action recognition in existing 2D tracking based methods or to define scene flow descriptors.

6.6. Knit your Ideas Into Smart Spaces

Participants: Joelle Coutaz, Alexandre Demeure [correspondant], Emeric Fontaine.

We developed KISS (Knit your Ideas Into Smart Spaces), an end-user development system for the home. KISS enables users to program their home with sentences expressed in a pseudo-natural language. Programs can be tested either with the virtual home or in the real home. We led an evaluation that shows that users are able to program a real-life scenario. This work is described in the phd manuscript of Emeric Fontaine [46]. An experimental evaluation shows that KISS can be used to programm a real life scenario.

Particpants encountered some difficulties related to the restricted vocabulary used for the experiment. Some difficulties also occured relative to the understanding of "progressive verbs". To overcome these problems, we envision a system for co-constructing vocabulary with the system, which may lead to the definition of multiple language for communicating with the system.

6.7. Attention-Based Navigation

Participants: Thomas Fisher, Thierry Fraichard [correspondant], Patrick Reignier.

Assistant robots and robot companions are designed to share the human living space, to navigate among and interact with human beings. From the mobility point of view, roboticists have recently striven to develop navigation scheme geared towards achieving so-called "socially acceptable motions". To that end, various concepts borrowed from environmental psychology and anthropology have been used, the "personal space" concept from Proxemics being perhaps the most widely used.

The purpose of our work here is to further the research in this area by taking into account other factors such as human activities, interaction configurations and intentions. An attentional model derived from cognitive psychology is used to dynamically determine the "focus of attention" of the persons involved in a given task. Depending on the task at hand, the robot uses the attention information in order to decide its future course of action so as, for instance, to attract one person's attention or, on the contrary, to minimize the disturbance caused.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. CAPEE - plateforme Capteur d'Activité et de Présence pour l'Efficacité Enérgétique

Participants: James Crowley [correspondant], Lucas Nacsa, Amaury Nègre, Lukas Rummelhard.

In Cooperation with local PME Novazion, Inria has worked with and Schneider Electric to create a research plafform for activity recognition for Smart Energy Systems. This system integrates information from multiple environmental sensor to detect and track people in indoor environments. Copies of this system has been installed at Schneider Electric Homes research group in Grenoble and in the Smart Spaces lab at Inria Montbonnot.

An associated software, named MultiSensor activity tracker, integrates information from multiple environmental sensors to keep track of the location and activity of people in a smart environment. This model is designed to be used by a home energy broker that would work in conjunction with a smart grid to manage the energy consumption of home appliances, balancing the needs of inhabitants with opportunities for savings offered by electricity rates. This database will also be used for by advisor services that will offer advice to inhabitants on the consequences to energy consumption and energy cost that could potentially result from changes to lifestyle or home energy use.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. EquipEx AmiQual4Home - Ambient Intelligence for Quality of Life

Participants: Stan Borkowski, Sabine Coquillart, Joelle Coutaz, James Crowley [correspondant], Alexandre Demeure, Thierry Fraichard, Amaury Nègre, Patrick Reignier, Dominique Vaufreydaz.

Ambient Intelligence, Equipment d'Excellence, Investissement d'Avenir

The AmiQual Innovation Factory is an open research facility for innovation and experimentation with human-centered services based on the use of large-scale deployment of interconnected digital devices capable of perception, action, interaction and communication. The Innovation Factory is to be composed of a collection of workshops for rapid creation of prototypes, surrounded by a collection of living labs and supported by a industrial innovation and transfer service. Creation of the Innovation Factory has been made possible by a 2.140 Million Euro grant from French National programme "Investissement d'avenir", together with substantial contributions of resources by Grenoble INP, Univ Joseph Fourier, UPMF, CNRS, Schneider Electric and the Commune of Montbonnot. The objective is to provide the academic and industrial communities with an open platform to enable research on design, integration and evaluation of systems and services for smart habitats.

The core of the AmiQual Innovation Factory is a Creativity Lab composed of a collection of five workshops for the rapid prototyping of devices that integrate perception, action, interaction and communications into ordinary objects. The Creativity Lab is surrounded by a collection of six Living Labs for experimentation and evaluation in real world conditions. The combination of fabrication facilities and living labs will enable students, researchers, engineers, and entrepreneurs to experiment in co-creation and evaluation. The AmiQual Innovation Factory will also include an innovation and transfer service to enable students, researchers and local entrepreneurs to create and grow new commercial activities based on the confluence of digital technologies with ordinary objects. The AmiQual Innovation Factory will also provide an infrastructure for participation in education, innovation and research activities of the European Institute of Technology (EIT) KIC ICTLabs.

The AmiQual Innovation Factory is a unique combination of three different innovation instruments: (1) Workshops for rapid prototyping of devices that embed perception, action, interaction and communication in ordinary objects nased on the MIT FabLab model, (2) Facilities for real-world test and evaluation of devices and services organised as open Living labs, (3) Resources for assisting students, researchers, entrepreneurs and industrial partners in creating new economic activities. The proposed research facility will enable scientific research on these problems while also enabling design and evaluation of new forms of products and services with local industry.

The AmiQual Innovation Factory will enable a unique new form of coordinated ICT-SHS research that is not currently possible in France, by bringing together expertise from ICT and SHS to better understand human and social behaviour and to develop and evaluate novel systems and services for societal challenges. The confrontation of solutions from these different disciplines in a set of application domains (energy, comfort, cost of living, mobility, well-being) is expected to lead to the emergence of a common, generic foundation for Ambient Intelligence that can then be applied to other domains and locations. The initial multidisciplinary consortium will progressively develop interdisciplinary expertise with new concepts, theories, tools and methods for Ambient Intelligence.

The potential impact of such a technology, commonly referred to as "Ambient Intelligence", has been documented by the working groups of the French Ministry of Research (MESR) [35] as well as the SNRI (Stratégie Nationale de la Recherche et de l'Innovation).

8.1.2. *INRETS Intelligent Urban Spaces Platform*

Participants: Claudine Combe, James Crowley [correspondant], Lukas Rummelhard.

Visual detection and tracking of pedestrians, Intelligent Urban Space

The project ANR-07-TSFA-009-01 CIPEBUS ("Carrefour Intelligent - Pôle d'Echange - Bus) has been proposed by INRETS-IFSTTAR, in collaboration with Inria, Citilog, Fareco, and the city of Versaille. The Objective of the CIPEBUS project is to develop an experimental platform for observing activity in a network of urban streets in order to experiment with techniques for optimizing circulation by context aware control of traffic lights.

Within CipeBus, Inria jas developed a real time multi-camera computer vision system to detect and track people using a network of surveillance cameras. The CipeBus combines real time pedestrian detection with 2D and 3D Bayesian tracking to record the current position and trajectory of pedestrians in an urban environment

under natural view conditions. The system extends the sliding window approach to use a half-octave Gaussian Pyramid to explore hypotheses of pedestrians at different positions and scales. A cascade classifier is used to determine the probability that a pedestrian can be found at a particular position and scale. Detected pedestrians are then tracked using a particle filter.

The resulting software system has been installed and tested at the INRETS CipeBus platform and is currently used for experiments in controlling the traffic lights to optimize the flow of pedertrians and public transportation while minimizing the delay imposed on private automobiles.

8.1.3. FUI 3Dlive

Participants: Frédéric Devernay, Sylvain Duchêne, Matthieu Volat.

3Dlive (http://3dlive-project.com) is a collaborative project, supported by French Ministry of Industry, and involving 3 industry and research clusters: Images & Reseaux (Brittany and Pays-de-la-Loire regions), Imaginove (Rhône-Alpes region), Cap Digital (Paris region). The objectives of this project are to create expertise in France for the live filming and transmission of 3D stereo contents, and to help French industry and universities to be major global 3D actors. 3Dlive won the *Loading the Future* trophy from the Images & Reseaux cluster in 2011. The consortium consists of:

> R&D/industry:
>> Orange Labs (project leader), Technicolor (3D R&D), Thomson Video Networks (encoders) and Thales Angenieux (optics).
> Small companies:
>> AMP (TV shooting) and Binocle (specific 3D HW & SW manufacturer).
> University labs:
>> Inria/PRIMA and Institut Telecom.

The role of PRIMA within this project is to develop new algorithms for real-time processing of stereoscopic video streams. This includes:

> stereoscopic video rectification and geometric adjustments.
> view interpolation, and extraction of stereoscopic metadata for the adaptation of the stereoscopic content to the projection screen.

These algorithms rely on view- and scale- invariant feature extraction, feature matching, dense stereoscopic reconstruction, and computer graphics techniques (matting, and accelerated processing and rendering using the GPU).

8.1.4. FUI PRAMAD

Participants: Wafa Benkaouar, Claudine Combe, Dominique Vaufreydaz [correspondant].

Pramad is a collaborative project about *Plateforme Robotique dÁssistance et de Maintien à Domicile*. There are seven partners:

> R&D/industry:
>> Orange Labs (project leader) and Covéa Tech (insurance company),
> Small companies:
>> Wizarbox (game designer) and Robosoft (robot).
> Academic labs:
>> Inria/PRIMA, ISIR (Paris VI) and Hôpital Broca (Paris).

The objectives of this project are to design and evaluate robot companion technologies to maintain frail people at home. Working with its partners, PRIMA research topics are:

> social interaction,
> robotic assistance,
> serious game for frailty evaluation and cognitive stimulation.

8.1.5. *Large-scale initiative action PAL*

Participants: Rémi Barraquand, Thierry Fraichard, Patrick Reignier, Dominique Vaufreydaz.

The 12 Inria Project-Teams (IPT) participating in a Large-scale initiative action Personally Assisted Living (PAL http://pal.inria.fr) propose to work together to develop technologies and services to improve the autonomy and quality of life for elderly and fragile persons. The goal of this program is to unite these groups around an experimental infrastructure, designed to enable collaborative experimentations.

PAL is organized arround 12 IPT:

> Coprin, Demar, E-Motion, Flowers, Lagadic, Lagadic-Sophia, Maia, Phoenix, Prima, Pulsar, Reves and Trio.

The role of PRIMA within this project is to develop new algoritms mainly along two research axes:

> assessing frailty degree of the elderly,
> social interaction.

8.2. European Initiatives

8.2.1. *Collaborations in European Programs*

> Program: CATRENE - Communication and digital lifestyle
>
> Project acronym: AppsGate
>
> Project title: Applications Gateway
>
> Duration: September 2012 to March 2015
>
> Coordinator: ST Microelectronics
>
> Other partners: Pace, Technicolor, NXP, Myriad France SAS, 4MOD Technology, HI-IBERIA Ingenieria y Proyectos, ADD Semiconductor, Video Stream Network, SoftKinetic, Optrima, Fraunhofer, Vsonix, Evalan, University UJF/LIG, and Institut Telecom
>
> Abstract:
>
> AppsGate will develop an Open Platform to provide integrated home applications to the consumer mass market. The set-top box is the primary point of entry into the digital home for television services including cable TV, satellite TV, and IPTV. AppsGate will transform the set-box into a residential gateway, capable of delivering multiple services to the home, including video, voice and data. The AppsGate project is putting together chip suppliers, consumer electronics OEMs and service providers to demonstrate an advanced Set Top Box that provides a home gateway for applications in the areas of entertainment, home automation, energy management and healthcare. This project aims at developing an Open Platform to provide integrated home applications to the consumer mass market. The set-top box is the primary point of entry into the digital home for television services including cable TV, satellite TV, and IPTV. This device has evolved beyond its historical role as a simple black box sitting on top of a large TV set into a device that supports a variety of functions, notably interactive television applications. Another interesting development is the concept of residential gateway, which is a complex device capable of delivering multiple services to the home, including video, voice and data.
>
> Both the set-top box and the residential gateway can be combined into a unique platform to deliver the same rich experience to multiple users in different rooms. When various devices are connected to this platform and multiple applications are seamlessly integrated together, the concept of application gateway or AppsGate is born. This new platform, which offers the prospect of unprecedented business opportunities, is the focus of the project.

8.2.2. *Collaborations with EIT KIC ICTlabs*

ICTLabs is the KIC for ICT (http://eit.ictlabs.eu/ict-labs/) ICTlabs is set up as a network of 6 "co-location" centers in Helsinki, Stockholm, Berlin, Paris, Eindhoven and Trento. The Paris node is run by Inria with partners Alcatel Lucent, Orange, University Paris Sud Institut Telecom.

PRIMA actively participates in the thematic actions: Smart Spaces, Smart Energy Systems and Health and Well Being.

ICTLabs Action Line Smart Spaces (ASSP) Activity 11547 : PI3 - Pervasive Information interfaces and interaction.

With activity PI3 we have constructed and released an "Attention Recognition Module"

ICTLabs Action Line Smart Spaces (ASSP) Activity 12201 : TIK - The Interaction Toolkit

PRIMA coordinates the Activity TIK. This activity will deliver a standard library of tools for human computer interaction for smart Spaces.

ICTLabs Action Line TSES - Smart Energy Systems Activity 12201 : Activity 11831 Open SES Experience Labs

PRIMA has constructed a testbed that integrates information from multiple environmental sensor to detect and track people and recognize their activity.

ICTLabs Action Line THWB Health and Wellbeing, IActivity 12100 "Affective Computing".

PRIMA has constructed a embedded software system for mobile computing that can detect and track faces, and measure the physiological parameters of Valence, Arousal and Dominance in order to recognize and stimulate human emotion.

8.3. International Initiatives

8.3.1. *Inria International Partners*

Starting with the PERSPOS project (BQR Grenoble INP 2008-2009) PRIMA has a long standing collaboration the MICA center (UMI 2954 CNRS). Our current goal is to develop the concept of "large-scale" perceptive space that is an intelligent environment which will be deployed on a large surface containing several buildings (as a university campus for example). The user is assumed to wear one or many mobile intelligent wireless devices (telephone, Smartphone, PDA, notebook). Using these devices, one can use many different applications (read emails, browse the Internet, file exchange, etc.). By combining the concepts of large-scale perceptive environments and mobile computing, we can create intelligent spaces to propose services adapted to individuals and their activities. Our collaboration is focussing the user location within such a smart space.

Tracking people in smart environments remains a challenging fundamental problem. Whether it is at the scale of a campus, of a building or more simply of a room, we can dynamicaly combine several localization levels (and several technologies) to allow a more accurate and reliable user localization system. This collaboration was concrete with the Ph.D. thesis from Han Yue (started in November 2008). This thesis was co-supervided between Grenoble INP and Hanoi Polytechnical Institute.

8.4. International Research Visitors

8.4.1. *Internships*

Marco Polo Cruz Ramos (from Dec. 2011 until May 2012)

Subject: Design of Interaction Systems for Mobile Robots Collaboration.

Institution: Technológico de Monterrey (Mexico).

Thomas FISCHER (from Feb. 2012 until Dec 2012).

Subject: Design of a Robot Companion.

Institution: University of Buenos Aires (Argentina).

9. Dissemination

9.1. Scientific Animation

James Crowley has served as co-organiser for the Dagstuhl seminar on Human Activity Recognition in Smart Environments, Dagstuhl Castle, Wadern Germany, 3-7 Dec 2012

James Crowley has served as co-organiser for the Symposium Performance Evaluation for Tracking and Surveillance, PETS 2013, IEEE Winter Meeting, Clearwater Florida, January 2013.

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

ICRA 2012, IEEE International Conference on Robotics and Automation, Saint Paul, Minn, 14-18 May 2012.

ICPR 2012, 21st International Conference on Pattern Recognition, Tsukuba Science City, JAPAN, November 11-15, 2012

CVPR 2012, IEEE Conference on Computer VIsion and Pattern Recognition, Providence Rhode-Island, 16-21 June 2012.

Sabine Coquillart has served as a member of the program committee for the following conferences:

GRAPP'2012 - International Conference on Computer Graphics Theory and Applications - Rome Italy, Feb. 2012.

VisGra 2012 - International Workshop on Computer Vision and Computer Graphics - Reunion Island, Feb. - March 2012.

IEEE 3DUI 2012 - IEEE 3D User Interfaces - OrangeCounty, USA, March 2012.

WSCG 2012 - International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision - Prague, Czech Republic, 2012.

SVR 2012 - Symposium on Virtual and Augmented Reality - Brazil , May 2012.

CASA 2012 - International Conference on Computer Animation and Social Agents, Singapore, May 2012.

CGI 2012 - Computer Graphics International, June 2012.

ISVC 2012 - International Symposium on Visual Computing - Rethymnon, Crete, Greece, July 2012.

CGVCVIP 2012 - IADIS Computer Graphics, Visualization, Computer Vision and Image Processing, Lisbon, Portugal, July 2012.

VSMM 2012 - 18th International Conference on Virtual Systems and Multimedia, Milan Italy, Sept. 2012.

ICVRV 2012 - International Conference on Virtual Reality and Visualization 2012, Ginhuangdao, China, September 2012.

JVRC 2012 - Joint Virtual Reality Conference on Virtual Reality of EGVE - ICAT - EuroVR, Madrid, Spain, October 2012.

GRAPHICON'2012 - International Conference on Computer Graphics and Vision - Moscow, Russia, October 2012.

EUROMED 2012 - International Conference on Cultural Heritage, Lemesos, Cyprus, Oct. Nov. 2012.

ISMAR 2012 - IEEE International Symposium on Mixed and Augmented Reality - Atlanta, Georgia, USA, Nov. 2012 (reviewer). CGAG 2012 - Computer Graphics, Animation and Game, Kangwondo, Korea, December 2012

Sabine Coquillart has served as a member of the Conference Committee and International Program Committee co-chair of IEEE VR 2012 - IEEE Virtual Reality, Orange County, CA, USA, March 2012. She has served as co-chair for the 2012 Joint Virtual Conference of EGVE - ICAT - EuroVR, Madrid, Spain, October 2012.

Sabine Coquillart is:

Elected member of the EUROGRAPHICS Executive Committee.

Member of the EUROGRAPHICS Working Group and Workshop board.

Member of the 2012 Joint Virtual Reality Conference of EGVE -ICAT- EuroVR Best Papers Award Committee.

Member of the steering committee for the ICAT conference - International Conference on Artificial Reality and Telexistence.

Chairing the steering committee for the EGVE Working Group - EUROGRAPHICS Working group on Virtual Environments.

Member of the FET pool of experts for European Commission funding under the framework of FET Open.

Sabine Coquillart did evaluation for the Eurographics PhD Awards Committee, 2012. She participated to the following Journal Editorial Boards:

IEEE Transactions on Visualization and Computer Graphics: co-guest editor with S. Feiner and K. Kiyokawa of a special issue (vol. 18 n. 4, 2012).

Journal of Virtual Reality and Broadcasting.

Presence: Teleoperators & Virtual Environments Journal: co-Guest Editor for some papers of a special issue.

Scientific World Journal.

Thierry Fraichard has served as an Associate Editor [1] for three international conferences:

IEEE Int. Conf. on Robotics and Automation (ICRA), St Paul (US), May 2012.

IEEE-RSJ Int. Conf. on Intelligent Robots and Systems (IROS), Vilamoura (PT), Oct. 2012.

IFAC Symp. on Robot Control (SYROCO), Dubrovnik (HR), Sep. 2012.

He has reviewed an article for the Autonomous Robots journal and has served has an expert evaluator for the French Research Agency and the European Commission (FP7 framework). Along with James Kuffner from CMU, he has edited a special issue on guaranteed motion safety for the Autonomous Robots journal.

Patrick Reignier has served as a Program Committee member for the following conferences:

Reconnaissance des Formes et Intelligence Artificielle (RFIA 2012), Lyon, France.

6th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2012), Vitoria-Gasteiz, Spain.

First Workshop on the Rights and Duties of Autonomous Agents (RDA2).

Workshop Affect, Compagnon Artificiel, Interaction (WACAI 2012) Grenoble, France.

He has reviewed articles for:

ACM Transactions on Interactive Intelligent Systems.

Assistance and Service Robotics in a Human Environment workshop of IROS 2012.

He has been elected at the board of the Association Française pour l'Intelligence Artificielle (AFIA). He was a member of the selection committee for a Professor Position at the Joseph Fourier University and a member of the selection committee for a Professor Position at the Pierre et Marie Curie University.

Dominique Vaufreydaz has served as a Program Committee member for the following conferences:

UBICOMM 2012, The Sixth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, Barcelona, Spain, September 2012.

MTEL2012 at ISM2012, The Seventh IEEE International Workshop on Multimedia Technologies for E-Learning in conjunction with the IEEE International Symposium on Multimedia 2012, Irvine (CA), USA, December 2012.

---

[1] An Associate Editor has to handle the review process of a set of papers, i.e. assign reviewers, assess each paper and write the review reports.

He has reviewed articles for:

> International Journal On Advances in Internet Technology, November 2012.

> IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, May 2013.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

> Master: Sabine Coquillart, Virtual Reality and 3D User Interfaces, M2 MOSIG, Univ. of Grenoble, France.

> Master: Sabine Coquillart, 3D User Interfaces and Augmented Reality, one day, M2 Numerical Modeling and Virtual Reality, Univ. of Laval, France.

> Master: James Crowley is co-responsable for the Master of Science in Informatics at Grenoble (MoSIG).

> Master: James Crowley teaches the MoSIG M1 Cours - Intelligent Systems.

> Master: James Crowley teaches the MoSIG M2 Cours - Computer Vision.

> Master: Alexandre Demeure, Introduction â l'IHM, 45h eqTD, M1 MIAGE, UJF, France.

> Master: Alexandre Demeure, IHM avancées, 26h eqTD, M2 MIAGE, UJF, France.

> Master: Alexandre Demeure, Introduction â l'IHM, 36h eqTD, M1 RICM4, Polytech, France.

> Master: Alexandre Demeure, Architecture pour l'IHM, 26h eqTD, M2 RICM5, Polytech, France.

> Master: Alexandre Demeure, IHM avancées, 15h eqTD, M2 RICM5, UJF, France.

> Licence: Thierry Fraichard, Introduction à la robotique, 19h eqTD, L3 INFO, Univ. of Grenoble, France.

> Licence: Thierry Fraichard, Projet Java, 20h eqTD, L3 MIAGE, Univ. of Grenoble, France.

> Master: Thierry Fraichard, Introduction to Perception and Robotics, 23h eqTD, M1 MOSIG, Univ. of Grenoble, France.

> Master: Thierry Fraichard, Projet Génie Logiciel, 36h eqTD, M1, ENSIMAG/Grenoble INP, France.

> Master: Patrick Reignier, Projet Génie Logiciel, 55h eqTD, M1, ENSIMAG/Grenoble INP, France.

> Licence: Patrick Reignier, Projet C, 20h eqTD, L3, ENSIMAG/Grenoble INP, France.

> Master: Patrick Reignier, Programmation Internet, 18h eqTD, M1, ENSIMAG/Grenoble INP, France

> Master: Patrick Reignier, Programmation C, 65h eqTD, M2, Université Joseph Fourier, France.

> Master: Patrick Reignier, Produits Structuré (informatique pour la finance), 24h eqTD, M2, EN-SIMAG/Grenoble INP, France.

> Licence: Dominique Vaufreydaz, Traitement de texte et tableur pour l'économie, 48h eqTD ,L1, Grenoble II, France.

> Licence: Dominique Vaufreydaz, Informatique appliquée à l'économie et à la gestion, enseignement à distance, Licence, Grenoble II, France.

> Licence: Dominique Vaufreydaz, Pratique avancée du Tableur, 72 h eqTD, L3, Grenoble II, France.

> Licence Professionnelle: Dominique Vaufreydaz, Enquêtes et traitement d'enquêtes avec le logiciel Sphinx, 11h eqTD, Licence pro Métiers de l'Emploi et de la Formation, 11h eqTD, Grenoble II, France.

> Licence Professionnelle: Dominique Vaufreydaz, Administration Windows, 39.5h eqTD, Licence pro Métiers de l'Emploi et de la Formation, Grenoble II, France.

> Master: Dominique Vaufreydaz, Enquêtes et traitement d'enquêtes avec le logiciel Sphinx, 11h eqTD, M2 Strategies économiques du sport et du tourisme, Grenoble II, France.

Master: Dominique Vaufreydaz, Pratique avancée du Tableur, 22 h eqTD, M1 Économie interna-tionale et stratégies d'acteurs, Grenoble II, France.

Master: Dominique Vaufreydaz, Mise à niveau Informatique pour l'économie, 22h eqTD, M2 Diagnostic économique d'entreprise, Grenoble II, France.

9.2.2. Supervision

PhD : Mathieu Guillame-Bert, Learning Temporal Association Rules on Symbolic Time Sequences, Thesis defended December 2012, Thesis Director James Crowley

PhD : Antoine Meler, BetaSAC et OABSAC, Deux Nouveaux Echantillonnages Conditionnels pour RANSAC, Thesis defended January 2013, Thesis Director James Crowley

PhD in progress : Marion Decrouez, Modelisation et Localisation Visuelle dans les Environnement Dynamiques, Thesis Directed by Frédéric Devernay and James Crowley, Soutenance PrŽvu mars 2013,

PhD in progress : Evanthia Mavridou, Visual Invariants for Detection and Recognition, Thesis Director James Crowley

PhD in progress : Varun Jain, Perception of Human Emotions. Thesis Director James Crowley

PhD in progress : Julian Quiroga, Visual Perception of Gestures, Thesis Directed by Frédéric Devernay and James Crowley

PhD : Emeric Fontaine, Programmation d'espace intelligent par l'utilisateur final, UJF, July 2012, Joelle Coutaz et Alexandre Demeure.

PhD in progress : Dimitri Masson, Modèles et outils pour favoriser la créativité dans les premières phases de conception d'IHM, Gaelle Calvary et Alexandre Demeure

PhD: Alessandro Renzaglia, Distributed Control for Autonomous Helicopters, Univ. of Grenoble, France, April 2012, Thierry Fraichard and Agostino Martinelli.

PhD: Marco Polo Cruz Ramos, Design of Interaction Systems for Mobile Robots Collaboration; a Marsupial Robot Team for Search and Rescue Operations Case Study, Technológico de Monterrey, Mexico, December 2012, Jose-Luis Gordillo and Thierry Fraichard.

PhD: Rémi Barraquand, Designing Sociable Technologies, February 2nd 2012, Univ. of Grenoble, James Crowley and Patrick Reignier.

9.2.3. Juries

Thierry Fraichard (reviewer), PhD, Jim Mainprice, Univ. of Toulouse (FR), December 2012.

Patrick Reignier (examiner), PhD, Charbel El Kaed, Univ. of Grenoble, January 13th. 2012

Patrick Reignier (reviewer), HDR, Cédric Buche, National Engineering School of Brest, France, February 10th 2012.

Patrick Reignier (examiner), PhD, Michelle Leonhardt Camargo, Universidade Federal do Rio Grande do Sul, Brasil, March 13th 2012.

Patrick Reignier (examiner), PhD, Issac Noé Garcia Garza, Grenoble University, June 18th 2012.

Patrick Reignier (examiner), PhD, Cyrille Martin, Univ. of Grenoble, October 4th 2012.

Patrick Reignier (examiner), PhD, Shirley Hoet, Pierre and Marie Curie University, Paris, France, December 17th 2012.

9.3. Invited Presentations

Sabine Coquillart: First-person Visuo-haptic Environment - From Research to Applications, VISI-GRAPP‰¢¥ú2012, Rome, Italy, February 2012.

Sabine Coquillart: Multimodal Virtual Environment: from Research to Applications, ISVC‰¢¥ú2012, Rethymnon, Crete, Greece, July 2012.

Sabine Coquillart: Augmented Reality: State of the Art and Perspectives, Sharesight, Paris, December 2012.

Thierry Fraichard, Will the driver seat ever be empty?, LAAS Lab., Toulouse (FR), Dec. 2012.

Thierry Fraichard, The Difficulty of Safely Navigating Dynamic Environments, Ben Gurion Univ., Be'er Sheva (IL), Dec. 2012.

James Crowley, Perception for Social Interaction, Seminar de SFR Pole Cognition, Grenoble, 31 May 2012

James Crowley, A Cognitive Architecture for Situation Awareness, German-French Workshop on Perspectives on Cognitive Interaction and Technology, Beilefeld, Germany 4 - 6 June 2012

James Crowley, Intelligence Ambiante et l'Innovation Factory d'AmiQual4Home, Lundis de l'Innovation, SFR InnoVacs, 5 Nov 2012.

James Crowley, Intelligence Ambiante : Invasion Pprogressive de l'Informatique, InnoFond Seminar, Paris 29 November 2012.

James Crowley, Visual Recognition of Human Activity, in Dagstuhl seminar on Human Activity Recognition in Smart Environments, 3-7 Dec 2012

10. Bibliography

Major publications by the team in recent years

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

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

[3] O. BRDICZKA, J. CROWLEY, P. REIGNIER. *Learning Situation Models in a Smart Home*, in "IEEE Transactions on System, Man and Cybernetics, part B", 2009, vol. 39, n° 1, http://hal.inria.fr/hal-00665319/en.

[4] O. BRDICZKA, J. MAISONNASSE, P. REIGNIER, J. CROWLEY. *Detecting Small Group Activities from Multimodal Observations*, in "International Journal of Applied Intelligence", 2009, vol. 30, n° 1, p. 47–57, http://www-prima.imag.fr/jlc/papers/articleAI_brdiczka12Mars07.pdf.

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

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

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

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

[9] D. HALL. *Automatic parameter regulation of perceptual systems*, in "Image and Vision Computing", August 2006, vol. 24, n° 8, p. 870–881, http://dx.doi.org/10.1016/j.imavis.2006.02.011.

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

Publications of the year

Articles in International Peer-Reviewed Journals

[11] R. BARRAQUAND, D. VAUFREYDAZ, R. EMONET, A. NÈGRE, P. REIGNIER. *The OMiSCID 2.0 Middleware: Usage and Experiments in Smart Environments*, in "International Journal On Advances in Software", March 2012, vol. 4, n° 3&4, p. 231-243, http://hal.inria.fr/hal-00693395.

[12] S. BOURAINE, T. FRAICHARD, H. SALHI. *Provably Safe Navigation for Mobile Robots with Limited Field-of-Views in Dynamic Environments*, in "Autonomous Robots", April 2012, vol. 32, n° 3, p. 267-283 [*DOI :* 10.1007/S10514-011-9258-8], http://hal.inria.fr/hal-00733913.

[13] T. FRAICHARD, J. KUFFNER. *Guaranteeing Motion Safety for Robots*, in "Autonomous Robots", April 2012, vol. 32, n° 3, p. 173-175 [*DOI :* 10.1007/S10514-012-9278-Z], http://hal.inria.fr/hal-00733712.

Articles in National Peer-Reviewed Journals

[14] E. FONTAINE, A. DEMEURE, J. COUTAZ, N. MANDRAN. *Retour d'expérience sur KISS, un outil de développement d'habitat intelligent par l'utilisateur final*, in "IHM 2012", October 2012, http://hal.inria.fr/hal-00779660.

International Conferences with Proceedings

[15] W. BENKAOUAR, D. VAUFREYDAZ. *Multi-Sensors Engagement Detection with a Robot Companion in a Home Environment*, in "Workshop on Assistance and Service robotics in a human environment at IEEE International Conference on Intelligent Robots and Systems (IROS2012)", Vilamoura, Algarve, Portugal, Anne Spalanzani, David Daney, Olivier Simonin, Jean-Pierre Merlet, October 2012, p. 45-52, Workshop FW1 "Assistance and Service Robotics in a Human Environment" - Session3: Behavioral modeling and Human/Robot Interaction - http://www.iros2012.org, http://hal.inria.fr/hal-00735150.

[16] S. BOURAINE, T. FRAICHARD, H. SALHI. *Provably Safe Navigation for Mobile Robots with Limited Field-of-Views in Unknown Dynamic Environments*, in "ICRA 2012 - IEEE International Conference on Robotics and Automation", Saint Paul, MN, United States, IEEE, May 2012, p. 174-179, Technical session [*DOI :* 10.1109/ICRA.2012.6224932], http://hal.inria.fr/hal-00768527.

[17] M. DECROUEZ, R. DUPONT, F. GASPARD, J. CROWLEY. *Extracting planar structures efficiently with revisited BetaSAC*, in "21st Int. Conf. on Pattern Recognition (ICPR)", Tsukuba Science City (JP), November 2012.

[18] F. DEVERNAY, S. PUJADES, C. VIJAY. *Focus Mismatch Detection in Stereoscopic Content*, in "SD&A 2012 - Stereoscopic Displays and Applications XXIII", San Francisco, United States, A. J. WOODS, N. S. HOLLIMAN, G. E. FAVALORA (editors), SPIE, January 2012, vol. 8288, p. 82880E - 82880E:13, Session 4: 3D Image Processing and Content Adaptation - paper [8288-12] - Conf. website: http://stereoscopic.org/2012 [*DOI :* 10.1117/12.906209], http://hal.inria.fr/hal-00654115.

[19] S. DUCHÊNE, M. LAMBERS, F. DEVERNAY. *A Stereoscopic Movie Player with Real-Time Content Adaptation to the Display Geometry*, in "SD&A 2012 - Stereoscopic Displays and Applications XXIII", San Francisco, United States, A. J. WOODS, N. S. HOLLIMAN, G. E. FAVALORA (editors), SPIE, January 2012, vol. 8288, p. 82882Q - 82882Q:8, Poster Session: Stereoscopic Applications - paper [8288-15] - Conf. website: http://stereoscopic.org/2012 [*DOI :* 10.1117/12.908741], http://hal.inria.fr/hal-00657928.

[20] M. GUILLAME-BERT, J. CROWLEY. *Learning Temporal Associative Rules on Symbolic Time Sequences*, in "Asian Conf. on Machine Learning (ACML)", Singapore, November 2012.

[21] M. GUILLAME-BERT, J. CROWLEY. *Planning with Inaccurate Temporal Rules*, in "Int. Conf. on Tools with Artificial Intelligence (ICTAI)", Athens (GR), November 2012.

[22] J. QUIROGA, F. DEVERNAY, J. CROWLEY. *Scene Flow by Tracking in Intensity and Depth Data*, in "CVPRW 2012 - Computer Vision and Pattern Recognition Workshops", Providence, RI, United States, W. LI, Z. ZHANG (editors), IEEE, June 2012, p. 50-57, Workshop "Human Activity Understanding from 3D Data" (HAU3D) [*DOI :* 10.1109/CVPRW.2012.6239237], http://hal.inria.fr/hal-00762557.

[23] J. RUIZ-HERNANDEZ, J. CROWLEY, C. COMBE, A. LUX, M. PIETIKAINE. *Robust and Computationally Efficient Face Detection using Gaussian Derivative Features of Higher Orders*, in "European Conf. on Computer Vision (ECCV), Workshops", 2012.

National Conferences with Proceeding

[24] M. DECROUEZ, R. DUPONT, F. GASPARD, F. DEVERNAY, J. CROWLEY. *Détection et apprentissage automatique d'objets pour la modélisation de milieux intérieurs*, in "RFIA 2012 - Colloque sur la Reconnaissance des Formes et l'Intelligence Artificielle", Lyon, France, January 2012, vol. ISBN : 978-2-9539515-2-3, 8, Session "Posters", http://hal.inria.fr/hal-00656550.

Scientific Books (or Scientific Book chapters)

[25] J. COUTAZ, G. CALVARY, A. DEMEURE, L. BALME. *Systèmes interactifs et adaptation centrée utilisateur : la plasticité des Interfaces Homme-Machine*, in "Informatique et intelligence ambiante : des capteurs aux applications", G. CALVARY, T. DELOT, F. SÈDES, J.-Y. TIGLI (editors), IC2 - Traité Informatique et Systèmes d'Information, L'Harmattan, July 2012, chap. 9, 57, http://hal.inria.fr/hal-00765504.

[26] J. COUTAZ, J. CROWLEY. *Intelligence ambiante : effet de mode ou d'avenir ?*, in "Informatique et intelligence ambiante : des capteurs aux applications", G. CALVARY, T. DELOT, F. SÈDES, J.-Y. TIGLI (editors), IC2 - Traité Informatique et Systèmes d'Information, L'Harmattan, July 2012, chap. 1, 13, http://hal.inria.fr/hal-00765452.

[27] T. FRAICHARD, T. HOWARD. *Iterative Motion Planning and Safety Issue*, in "Handbook of Intelligent Vehicles", A. ESKANDARIAN (editor), Springer, 2012, p. 1433-1458 [*DOI :* 10.1007/978-0-85729-085-4_55], http://hal.inria.fr/hal-00768956.

Books or Proceedings Editing

[28] S. COQUILLART, S. FEINER, K. KIYOKAWA (editors). *Proceedings of 2012 IEEE Virtual Reality, VR 2012*, IEEE, IEEE Computer Society, IEEE VGTC, March 2012, 183.

[29] S. COQUILLART, S. FEINER, K. KIYOKAWA (editors). *Special Issue of IEEE Transactions on Visualization and Computer Graphics*, IEEE, IEEE Computer Society, IEEE VGTC, March 2012, vol. 18, n° 4, 130.

[30] T. FRAICHARD, J. KUFFNER (editors). *Special Issue on Guaranteeing Motion Safety for Robots*, Autonomous Robots, 2012, vol. 32, n° 3.

References in notes

[31] *DNS-SD : DNS Service Discovery*, 2000, http://www.dns-sd.org.

[32] *Robosoft company*, http://www.robosoft.com/.

[33] M. BARTLETT, G. LITTLEWORT, M. FRANK, C. LAINSCSEK, I. FASEL, J. MOVELLAN. *Fully Automatic Facial Action Recognition in Spontaneous Behavior*, in "Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on", april 2006, p. 223 -230, http://dx.doi.org/10.1109/FGR.2006.55.

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

[35] J. COUTAZ, J. CROWLEY. *Plan « intelligence ambiante »: défis et opportunités*, Groupe de Travail « Intelligence Ambiante » du Groupe de Concertation Sectoriel (GCS3)., October 2008.

[36] J. COUTAZ, J. CROWLEY, S. DOBSON, D. GARLAN. *Context is Key*, in "Communications of the ACM", March 2005, vol. 48, n° 3, Special issue on the Disappearing Computer.

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

[38] S. DAVIDOFF. *Routine as resource for the design of learning systems*, in "Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Adjunct", New York, NY, USA, Ubicomp '10 Adjunct, ACM, 2010, p. 457–460, http://doi.acm.org/10.1145/1864431.1864486.

[39] M. DECROUEZ, R. DUPONT, F. GASPARD, F. DEVERNAY, J. CROWLEY. *Modélisation explicite des objets et de l'environnement en combinant les approches topologique et métrique pour la localisation*, in "ORASIS 2011 - 13e Congrès des jeunes chercheurs en vision par ordinateur", Praz-sur-Arly, France, Inria Grenoble Rhône-Alpes, June 2011, Session: SLAM et localisation, http://hal.inria.fr/hal-00651823.

[40] F. DEVERNAY, P. BEARDSLEY. *Stereoscopic Cinema*, in "Image and Geometry Processing for 3-D Cinematography", R. RONFARD, G. TAUBIN (editors), Geometry and Computing, Springer Berlin Heidelberg, 2010, vol. 5, p. 11-51 [*DOI :* 10.1007/978-3-642-12392-4], http://hal.inria.fr/inria-00509217/en.

[41] F. DEVERNAY, S. DUCHÊNE. *New View Synthesis for Stereo Cinema by Hybrid Disparity Remapping*, in "International Conference on Image Processing (ICIP)", Hong-Kong Hong Kong, Sep 2010, http://hal.inria.fr/inria-00531369/en.

[42] F. DEVERNAY, S. DUCHÊNE, A. RAMOS-PEON. *Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis*, in "SD&A 2011 - Stereoscopic Displays and Applications XXII", San Francisco, California, États-Unis, A. J. WOODS, N. S. HOLLIMAN, N. A.

DODGSON (editors), SPIE and IS&T, March 2011, vol. 7863, p. 786302-786302:12, Proceedings of SPIE-IS&T Electronic Imaging [*DOI :* 10.1117/12.872883], http://hal.inria.fr/inria-00565131.

[43] F. DEVERNAY, A. RAMOS-PEON. *Novel View Synthesis for Stereoscopic Cinema: Detecting and Removing Artifacts*, in "3DVP 2010: ACM Workshop on 3D Video Processing", Italie Firenze, Oct 2010, http://hal.inria.fr/inria-00531364/en.

[44] A. K. DEY, T. SOHN, S. STRENG, J. KODAMA. *iCAP: interactive prototyping of context-aware applications*, in "Proceedings of the 4th international conference on Pervasive Computing", Berlin, Heidelberg, PERVA-SIVE'06, Springer-Verlag, 2006, p. 254–271, http://dx.doi.org/10.1007/11748625_16.

[45] P. EKMAN, W. V. FRIESEN. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, 1977.

[46] E. FONTAINE. *Programmation d'espace intelligent par l'utilisateur final*, Université de Grenoble, France, 2012, 204, Thèse soutenue le 12 juillet 2012.

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

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

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

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

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

[52] P. JOHNSON-LAIRD. *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*, Harvard University Press, 1983.

[53] P. JOHNSON-LAIRD. *Computer and the Mind: An Introduction to Cognitive Science*, Harvard University Press, 1998.

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

[55] S. KOO, D.-S. KWON, FILTERING, EXTENDED KALMAN. *Recognizing Human Intentional Actions from the Relative Movements between Human and Robot*, in "Nonlinear Dynamics", 2009, p. 939–944.

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

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

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

[59] S. PUJADES, F. DEVERNAY. *Stereoscopic Focus Mismatch Monitoring*, in "CVMP - European Conference on Visual Media Production", London, Royaume-Uni, The Foundry, 2011, 1, Short Paper selected for Poster Presentation, http://hal.inria.fr/hal-00651843.

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

[61] E. A. SCHEGLOFF. *Body Torque*, in "Social Research", 1998, vol. 65, n$^{\text{o}}$ 3, p. 535–596 [*DOI :* 10.1017/S0140525X04400053], http://www.questia.com/googleScholar.qst?docId=95791590.

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

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

[64] K. N. TRUONG, E. M. HUANG, G. D. ABOWD. *CAMP: A Magnetic Poetry Interface for End-User Programming of Capture Applications for the Home*, in "Proceedings of UBICOMP 2004: The 6th International Conference on Ubiquitous Computing", Nottingham, England, September 2004, p. 143–160.

[65] A. VINCIARELLI, M. PANTIC, H. BOURLARD. *Social signal processing: Survey of an emerging domain*, in "Image and Vision Computing", 2009, vol. 27, n$^{\text{o}}$ 12, p. 1743 - 1759, <ce:title>Visual and multimodal analysis of human spontaneous behaviour:</ce:title> [*DOI :* 10.1016/J.IMAVIS.2008.11.007], http://www.sciencedirect.com/science/article/pii/S0262885608002485.

[66] D. VOGEL, R. BALAKRISHNAN. *Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users*, in "Proceedings of the 17th annual ACM symposium on User interface software and technology", ACM, 2004, p. 137–146.