HR+: Towards an interactive autonomous robot

G. Bailly‡, L. Brèthes†, R. Chatila†, A. Clodic†, J. Crowley≀, P. Danès†,
F. Elisei‡, S. Fleury†, M. Herrb†, F. Lerasle†, P. Menezes†,
R. Alami†

† LAAS-CNRS
7, Av. du Colonel Roche
31077 Toulouse Cedex 4
France

‡ ICP
46, Av. Félix Viallet
38031 Grenoble
France

≀ GRAVIR
Av. de l'Europe
38330 Montbonnot
France

***Abstract*— The HR+ project investigates perception, decision and interface issues for human-robot interaction. We report here on the research activities that have been conducted by the partners of the project and that are relevant to HR+ context. We then present their partial integration in an autonomous interactive robot called Rackham. Rackham, has been deployed in a public area for relatively long periods in direct contact with non-professional persons, accumulating valuable data and information for future enhancements.**

I. Introduction

The development of personal robots is a new center of convergence and a motivating challenge in robotics research. One key aspect is "added" to the "standard challenge" of autonomous robots: the essential role of the "human in the loop". This has numerous consequences. Two of them are of particular interest for us:

1) the robot should be able to operate in an environment which has been essentially designed for humans, and
2) the robot will have to interact with human.

The human-centered theme is currently investigated in several areas. The spectrum of developments range from humanoids to wearable computing and sensing, human augmentation, telepresence, smart rooms or even intelligent objects.

The HR+ project investigates interaction paradigms considered on an incremental and interactive problem solving process based on:

- perception modalities (mainly based on vision) of human motion and gesture,
- decisional abilities taking into account explicit reasoning on the tasks, on the human environment and on the robot capacities to achieve them in a given context.
- robot abilities to convey information to the human through augmented reality and virtual talked heads

We report here on the research activities conducted by the partners that are relevant to HR+ context and on their partial integration in a autonomous interactive robot.

In section II, we briefly summarize the development of an architecture that is able to integrate in a principled manner vision-based processes for human activity observation.

Section III, we present a number of visual functions that have been developed to detect faces, track human limbs and interpret communication gestures.

Section IV reports on the development of flexible hybrid platform (hardware and software system) that allows to place users in a multi-modal face-to-face interaction loop with a talking agent and to record their activity for statistical analysis.

Then, section V describes Rackham, a new tour-guide robot that has been used as an integration platform. Rackham, has been deployed in a public area for long periods in direct contact with non-professional persons, accumulating valuable data and information for future enhancements.

II. Context-Aware Observation of Human Activity

PRIMA has developed an architecture in order to integrate in a principled manner vision-based processes for human activity observation [10]. The PRIMA robust tracker [14], shown in figure 1, will be used as an example to illustrate process architecture and components. Other forms of perceptual processes have been implemented in the process layer [11], [13].

Tracking is a cyclic process of recursive estimation. A well-known framework for such estimation is the Kalman filter. A complete description of the Kalman filter is beyond this paper. A general discussion of the use of the Kalman filter for sensor fusion is given in [7]. The use of the Kalman filter for tracking faces is described in [8].

Tracking provides a number of fundamentally important functions for a perception system. Tracking conserves information over time, thus provides object constancy. Object constancy assures that a label applied to a blob at time t_1 can be used at time t_2. Tracking enables the system focus attention, applying the appropriate detection processes only to the region of an image where a target is likely to be detected. Also the information about position and speed provided by tracking can be very important for describing situations.

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

To this set of phases, the PRIMA robust tracker adds a recognition phase, an auto-regulation phase, and a communi-

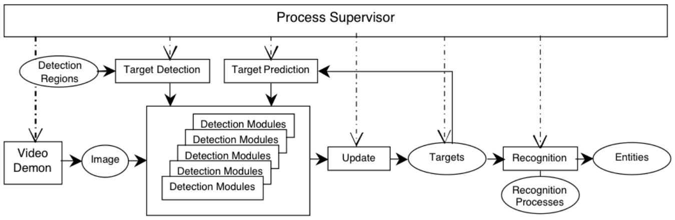


Fig. 1. The Prima Robust Tracker

cation phase. In the recognition phase, the tracker interprets recognition methods that have been downloaded to the process by a configuration tool. These methods are bits of code that may be expressed in scheme, CLIPS or C++. They are interpreted by a RAVI interpreter [15] and may result in the generation of events or the output to a stream. The auto-regulation phase determines the quality of service metric, such as total cycle time and adapts the list of targets as well as the target parameters to maintain a desired quality. During the communication phase, the supervisor responds to requests from other processes, the PFT (Process Federation Tool) or a federation supervisor. These requests may ask for descriptions of process state, or capabilities, or may provide specification of new recognition methods.

A. Supervisory Control

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

B. Process Scheduler

The process supervisor maintains a schedule of modules to be executed. The scheduler can interrupt processing after each phases to receive and react to events. Typically this schedule will be composed of phases, with the module calls within each phases determined by a list of data elements. We illustrate this with the PRIMA robust tracker. The robust tracker uses a schedule composed of the following six phases:

Module Execution Schedule for the Tracker:

1) GetNextImage();
2) For each current target: Predict target location and update target description;
3) For each detection region: If New Target Detected then add target to current target list;
4) For each recognition method in list: execute method();
5) Evaluate quality of result and adapt module schedule and parameters;
6) Interpret messages.

Each target and each detection region contains a specification for the module to be applied, the region over which to apply the module, and the step size to apply processing. Recognition methods are interpreted snippets of code that can generate events or write data to streams. These methods may be downloaded to a robust tracker as part of the configuration process by the process federation tool (PFT) or by a federation supervisor to give a tracker a specific functionality.

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

C. Homeostasis and Autonomic Control

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

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

Auto-description of processes is provided by the process supervisor’s response to requests over a communication channel. For example, in the robust tracker, auto-description requests include: $GetProcessMethods$ (returns the list of currently loaded recognition methods), $GetProcessDetectionModules$ (returns list of image processing modules available for detection), $GetProcessDetectionRegions$ (returns current list of detection regions).

Auto-description can also concern the state of the processes interpretation of the environment. For example, the robust tracker can respond to: $GetProcessEntities$ (return list of entities recognized by recognition methods), $GetProcessTargets$ (returns current list of targets), $GetProcessQoS$ (returns the current quality of service).

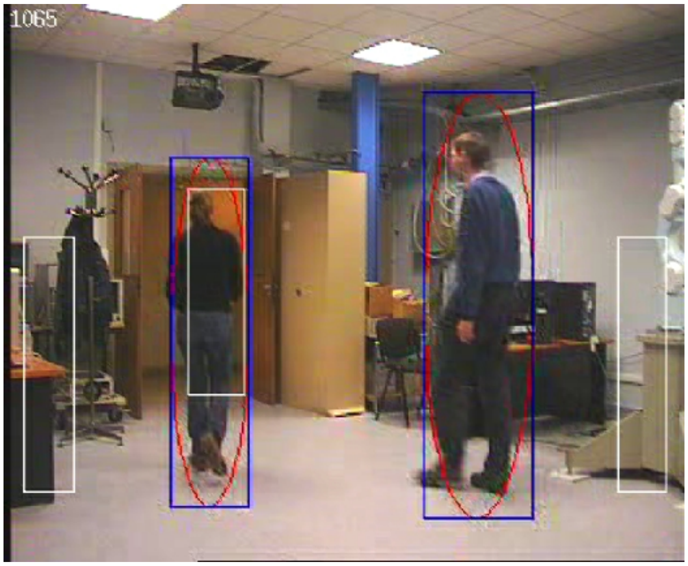


Fig. 2. An example of the Prima Tracker use

D. Communication between processes

Three classes of channels exist for communication between processes: events, streams and requests. Events are asynchronous symbolic messages that are communicated through a publish and subscribe mechanism provided by the Federation Supervisor. Streams provided serial high bandwidth data between two processes. Requests are asynchronous messages that ask for the current values of some process variables.

Figure 2 illustrates the use of the tracker in the HR+ context.

III. Vision-based Human-Robot interaction

LAAS contribution to vision-based human perception concerns the development of visual functions suitable to:

- detect and recognize faces, so as to identify possible persons in the robot's vicinity.
- track human limbs such as hands or faces in video streams.
- interpret communicative gestures, e.g. to symbolize referential actions to the robot, as well as pointing/manipulative gestures, e.g. to exchange objects with the robot.

For tracking purpose, the particle filtering formalism and alternative schemes have been investigated. Associated results have been compared in terms of performance and applicability to interaction modalities.

A first reason for focusing on particle filter as the tracking engine comes from its capability to cope with the non-Gaussian noise models required to represent cluttered environments. A second reason is that this framework allows the information from different measurements sources to be fused in a principled manner. Although this fact has been acknowledged before, it has not been fully exploited in this context. Combining or fusing a host of cues such as color, shape, motion, even —in the foreseeable future— sound can increase the reliability of trackers dedicated to human limbs.

Using shape cues requires that sufficiently precise shape models of the tracked limbs are learned beforehand. Tracking issues can be addressed considering either view-based (2D) shape models or 3D articulated models. The next section briefly outlines the aforementioned functions.

A forthcoming challenge consists in the integration of all these vision-based functions on an autonomous mobile robot — with on-board cameras — and in the evaluation of the robustness of the complete system in dynamic, cluttered and crowded environments with various lighting conditions. Further experiments will be inspired from scenarios which consider the robot as a museum guide.

A. Face detection, recognition and tracking

The face detector is based on a boosted cascade of Haar-like features to detect frontal faces while rejecting non-faces patterns (figure 3). The recognition module developed at ISR (Coimbra) takes as input the enclosing frontal face-like regions. The classification is based on eigen images and Mahanalobis-like distance [16].

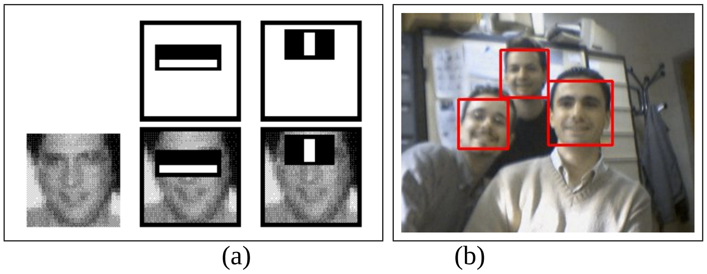


Fig. 3. (a) Haar-like features overlaying on a training face, (b) example of face detection

We have developed 2D trackers based on the combination or fusion of visual cues into various particle filtering schemes. In [4], we introduced mechanisms for data fusion within the original Condensation algorithm to develop face/hand trackers (at 20fps) fusing skin blobs and shape in a novel way. Faces and hands are here represented by their silhouette contours modeled by splines. The intermittent nature of skin regions and motion cues makes them candidate for the design of detection modules, from which efficient initialization strategies or importance functions can be defined for the particle filter —e.g. in the ICondensation framework— with the aim to avoid drift or target loss. Besides, color cues and shape cues tend to be remarkably persistent (figure 4) and are easy to fuse assuming they are independent. These considerations have been addressed and discussed in [5].

In a near future, we plan to adapt our tracking modules in order to fuse the above cues with other information — such as sound, motion or face regions (section III-A)— and to simultaneously track multiple targets.

1) Communicative gestures interpretation: Regarding communicative gestures, a mixed-state Condensation algorithm was proposed in [4] to recognize hand postures and automatic

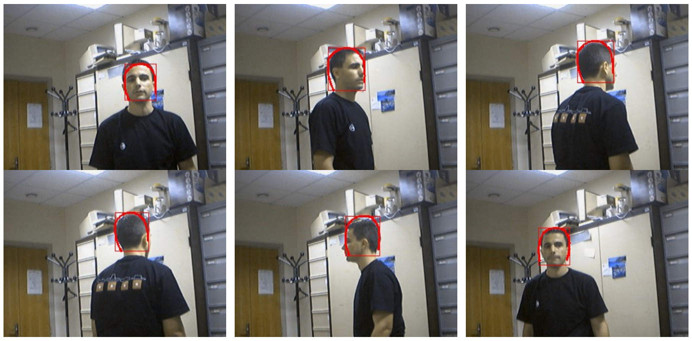


Fig. 4. Face tracker fusing color and shape cues : some snapshots from a tracking sequence

switch between multiple templates in the tracking loop. For a richer interaction, we recently extended this tracker[1] so as to handle multiple canonical motion models [5].

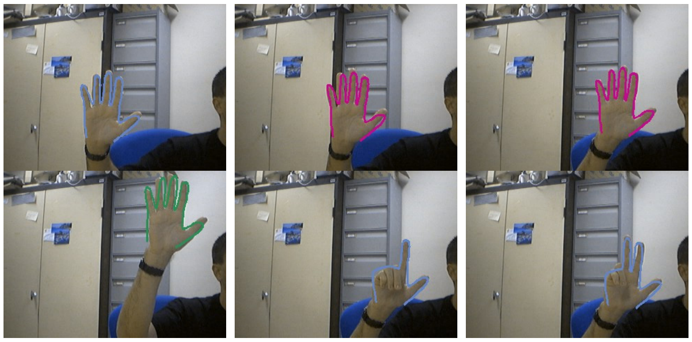


Fig. 5. Hand tracker with current posture and image motion (color) recognition: some snapshots from a tracking sequence

2) 3D gestures interpretation: 3D model-based approaches are also well suited to pointing gestures interpretation. We have focused on appearance-based tracking of high DoF 3D truncated quadrics (cones and cylinders) representing human hand (figure 6) or arm. A strategy is proposed to handle the projection and hidden parts removal efficiently. The non-Gaussian and non-linear character of the 3D model led also to particle filter based techniques. The criterion combines a measure based on the contours (from DT image) with a similarity measure of local color distribution.

The results presented in [17] show the feasibility of the approach when applied to the problem of tracking[2] a human arm (figure 7). The next step will be to extend our approach to deal with human posture tracking while occlusions will be handled using a multi-ocular system.

IV. Towards an Embodied Conversational Agent

ICP contributions are related to the way the robot will present itself, using a talking head displayed on its LCD panel, and the way it will communicate with human users. This setup allows users to gauge presence and engage in a mutual attention loop. In addition, it allows the robot to provide eye gazes towards widgets or other information displayed on the screen or towards objects in the real world. The long-term goal is to build an embodied conversational agent (ECA) able to maintain realistic face-to-face communication with a human interlocutor. This conversational agent is embodied by a video-realistic talking head. While most researchers focus on discourse interpretation and generation, the main challenge here is to provide the interlocutor with implicit and explicit signs of mutual interest and attention as well as with an awareness of environmental conditions in which interaction takes place.

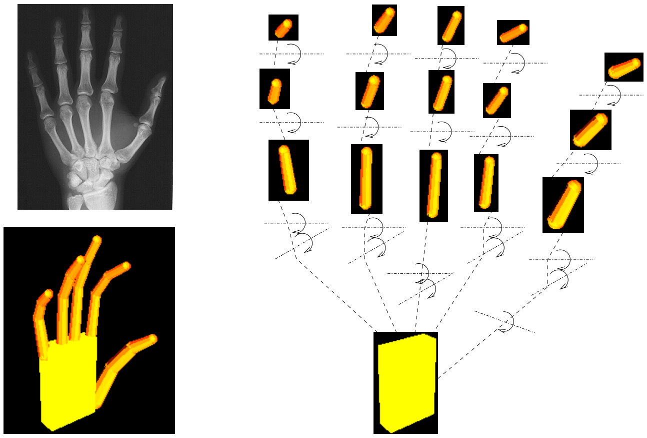


Fig. 6. A 3D hand structure with its DoF

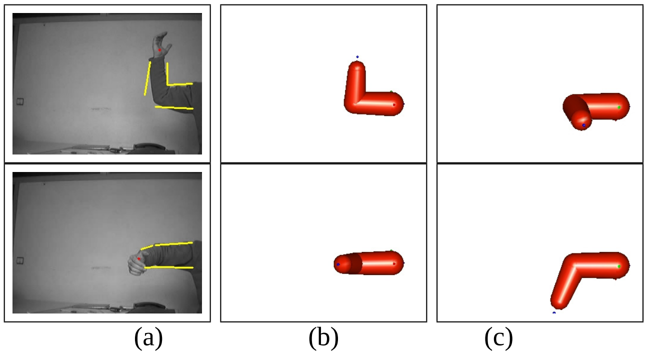


Fig. 7. Some snapshots from a tracking sequence: (a) projection of the 3D model, (b) resp. (c) associated frontal (resp. top) views of the model

A. *Talking head*

We have cloned the 3D appearance and articulation gestures of a real human [2], [20]. The eye gaze of the clone can be controlled independently to look at the user, to look at where the user is looking on the screen (giving signs of mutual attention) or to direct the user's attention to 2D objects on the screen (vergence of the eyes is handled and provides a crucial cue for inferring spatial cognition). The virtual neck is also articulated and can accompany the eye-gaze movements. The audiovisual messages can be either recorded by the original human speaker, or synthesized from text input. In practice the

[1] Videos of this last tracker results are available at www.laas.fr/∼lbrethes/icra05.

[2] Illustrative video sequences are available at www.laas.fr/∼pmenezes/t3final.mov.

synthetic signals are here generated off-line to avoid slight reaction delays.

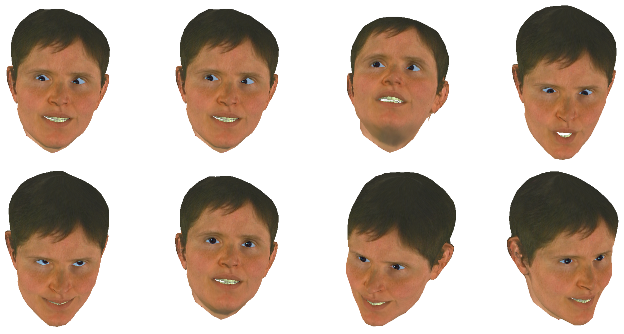


Fig. 8. The talking head: the neck and the eye movements of the 3D head have independent controls.

B. Dimensions of face-to-face interaction

Building an ECA that may engage into a face-to-face interaction/conversation with a human partner is quite challenging. Not only the ECA has to decode the user's needs and intentions through multimodal communication, but also must give direct and indirect signs that it actually knows about where the interaction is taking place, who is its interlocutor and what ambient/localized service it may provide to the user(s). Such a rich face-to-face interaction (see Figure 9) requires intensive collaboration between the scene analysis and the specification of the task to be performed in order to generate appropriate actions of the ECA.

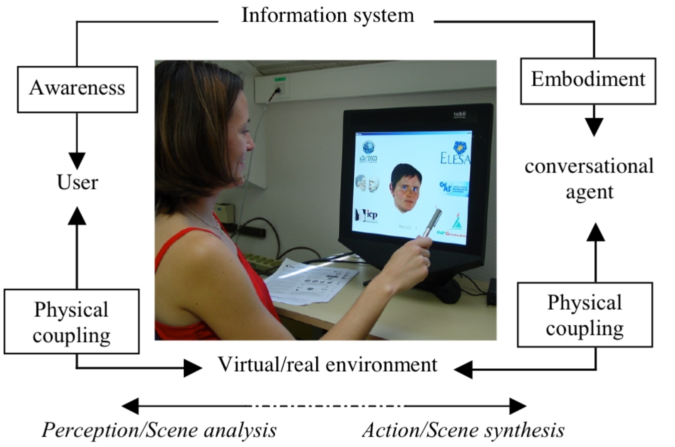


Fig. 9. Embodied conversational agents and ambient interaction. The control of agent actions should be aware of the user, the environmental conditions of the interaction and the competence of the information system to provide the user with relevant and reliable information. This involves a strong coupling between scene analysis and synthesis.

C. Research aims

Our perspective is to develop an embodied "Theory of Mind" (TOM) to link high-level cognitive skills to the low-level motor and perceptual abilities of a virtual conversational agent and to demonstrate that such a TOM will provide the information system with enhanced user satisfaction, efficient and robust interaction. The motor abilities is principally extended towards speech communication i.e. adapting content and speech style to pragmatic needs (e.g. confidentiality), speaker (notably age and possible communication handicaps) and environmental conditions (e.g. noise). If the use of a virtual talking head instead of a humanoid robot limits physical actions, it extends the domain of interaction to the virtual world: the user can also interact with other virtual objects (e.g. virtual icons) surrounding the virtual talking head (see the face-to-face system described below).

D. A dedicated face-to-face platform

The user sits in front of a standard-looking flat panel screen, where a 3D talking head faces him or her, as shown on Figure 2. Hardware and software specificities allow the user to interact with the system using eye gaze, a mouse and speech. The 3D clone can look at the user, talk to him, and react to where the user looks. These elements form the basis of a grounded virtual face-to-face situation.

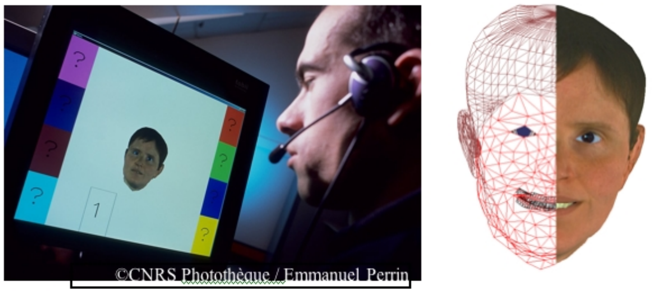


Fig. 10. Face-to-face interaction (speech, gaze and mutual attention) with a 3D clone.

E. Experiments

A first experiment [3] was conducted with a playing card scenario creating a "too much information at the same time" situation, where an agent was proposed to help retrieve the correct information. We expected that using eye gazes of the 3D clone as an extra modality might lead to faster performances or lower the cognitive load. Preliminary analysis showed that users willing to use this level of guidance could perform the task faster or easily: they could trust the clone and visit fewer cards. We demonstrated that cues of mutual attention may benefit the performance in information retrieval. We believe that the study and modeling of the components of human face-to-face interaction are crucial elements of intuitive, robust and reliable communication. We are currently investigating interactive real-time eye-gaze patterns of human speakers in face-to-face communication with a special focus on the speaking/listening state.These studies and results would of course benefit to personal robots such as Rackham.

V. Rackham

We have designed and implemented a new tour-guide robot. Besides robustness and efficiency in the robot basic navigation abilities in a dynamic environment, our focus was to develop and test a methodology to integrate human-robot interaction abilities in a systematic way.

To test and validate our developments, we have decided to bring regularly our robot to a museum in Toulouse. By regularly, we mean two weeks every three months. The robot, called Rackham, has already been used at the exhibition for hundreds of hours (July 2004, February 2005), accumulating valuable data and information for future enhancements. The project is conducted so as to incrementally enhance the robot functional and decisional capabilities based on the observation of the interaction between the public and the robot.

A. Mission Biospace and Rackham typical role

Mission BioSpace is an exhibition developed by the "Cité de l'Espace"[3] to illustrate what could be an inhabited spaceship. It presents about 14 interactive elements from "Lexigraph" to "Teleportation".

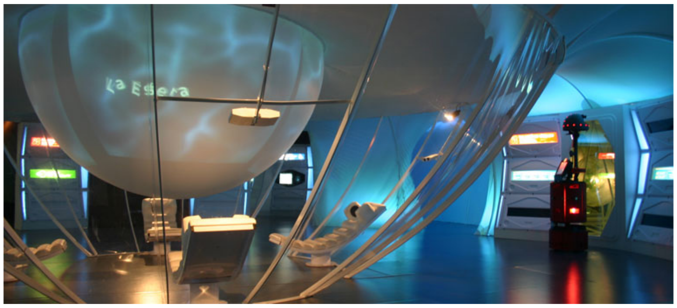


Fig. 11. The Tsiolkovski spaceship: A difficult environment context for navigation.

When Rackham is left alone with no mission, it looks forward to find out people to interact with. As soon as a person is detected, thanks to visual face detection, it presents itself "I'm Rackham and I can guide you in the spaceship" or alternatively explains how to use its services : "Select your destination using the touch-screen".

If the visitor finally selects a destination Rackham first confirms its new mission "OK, I will guide you to…", then plans and displays its trajectory and invites the visitor to follow it.

While navigating, the robot keeps on giving information about the progress of the on going travel : a congestion will require to temporarily stop or even to compute an alternative trajectory while a given level uncertainty on the position might call for a re-localization procedure; sporadic "disappearance" of the guided visitor are also detected and dealt with using sentences such as ”Where are you ?”,”Here you are again!”. The visitor may by himself stop and change the ongoing mission whenever he wants using various buttons displayed on the interface.

B. The robot

Rackham is a B21r robot made by iRobot. We have extended the standard equipment with one pan-tilt Sony camera EVI-D70, one digital camera mounted on a Directed Perception pan-tilt unit, one ELO touch-screen, a pair of loudspeakers, an optical fiber gyroscope and wireless Ethernet.

In order to integrate all these components in a robust and pleasant way the "Cité de l'Espace" has designed a "head" on a mast, the whole toped by an helmet which represents something between a one-eyed modern pirate and an African art statue (see picture 12). The eye is materialized by the EVI-D70 camera fixed upside-down above the helmet, the second camera is hidden in the helmet and one loudspeaker is placed within what represents the "mouth" (figure 12).

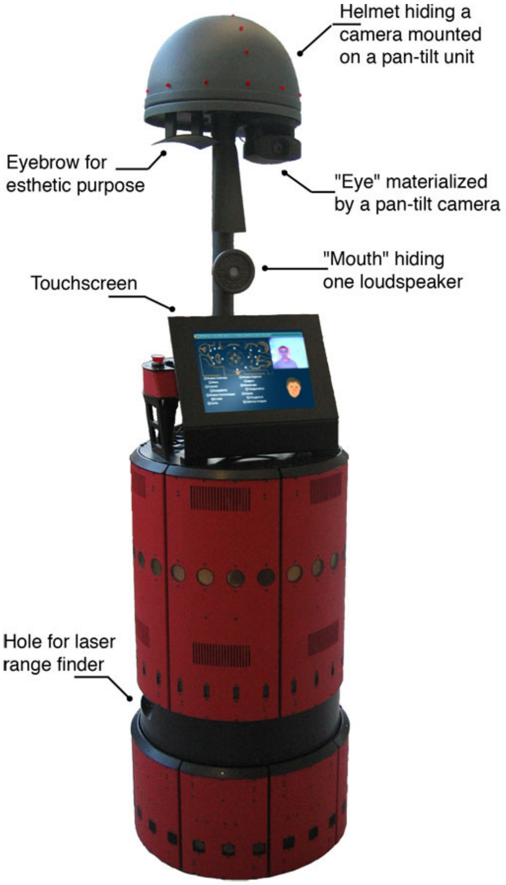

Helmet hiding a camera mounted on a pan-tilt unit
Eyebrow for esthetic purpose
"Eye" materialized by a pan-tilt camera
"Mouth" hiding one loudspeaker
Touchscreen
Hole for laser range finder

Fig. 12. Rackham and its equipment.

C. The software architecture

The software architecture is an instance of the LAAS[4] architecture ([1]). It is a hierarchical architecture including a supervisor written with openPRS[5] (a Procedural Reasoning System) that controls a distributed set of functional modules.

[3] http://www.cite-espace.com

[4] LAAS stands for: "LAAS Architecture for Autonomous Systems".

[5] The set of tools used to build an instance of this architecture (GenoM, openPRS, pocolibs, etc) are freely distributed at the following url: http://softs.laas.fr/openrobots.

A module is an independent software component that can integrate all the operational functions with various time constraints or algorithm complexity (control of sensors and actuators, servo-controls, monitorings, data processings, trajectory computations, etc.).

Each module is created using the generator of module GenoM and thus presents standard behavior and interfaces (see [12] and footnote 5). For Rackham, we have implemented 15 modules. We now present them according to their purpose in the system (see figure 13).

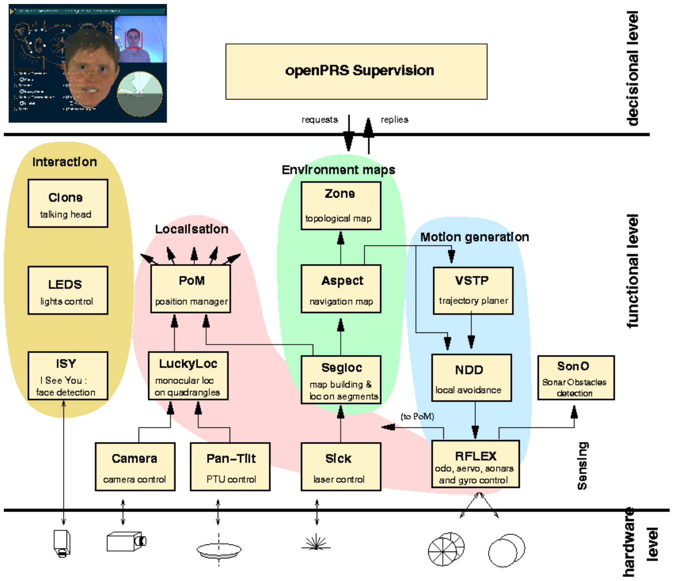


Fig. 13. The functional level of Rackham and its 15 modules.

1) Localization: Several modules are involved in the localization of the robot.

First the rflex module, which interfaces low level software provided by the manufacturer.

To localize itself within its environment the robot uses a SICK laser, controlled by the module sick, that exports at the required rate the laser echoes, and segments deduced from aligned echoes. Another module, segloc, is able to match these segments with segments previously recorded in a map thanks to a classical SLAM procedure. However the map is effectively updated only during closing time. The resulting map is composed of 232 segments (see figure 14).

The localization being a very critical ability, a third localization procedure, based on vision, has been designed. It consists on the identification of the furniture of the spaceship with one color camera. The camera is controlled by the module camera that produces images to be processed by the module luckyloc that extracts, identifies and localizes planar quadrangles. However, if luckyloc is already able to identify the various pieces of furniture, the localization procedure is not yet totally functional.

Finally, the various uncertain positions exported by the modules rflex, segloc and luckyloc are merged by pom, the position manager module. This module is able to integrate positions computed at various frequencies and even to propagate "old" position data (because of the time taken to acquire and process the data). The supervisor can be informed in case of localization problems with one of the modules, fusion difficulties or significant uncertainties on the position. Depending on the problem, various strategies are applied.

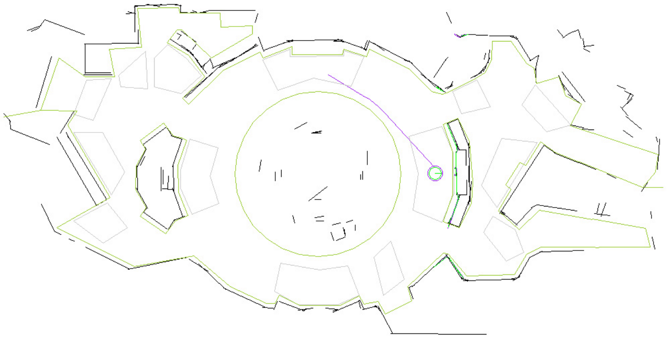


Fig. 14. The map of the environment built by the Rackham contains 232 segments (black) and has been augmented with forbidden zones (green or dark grey) and target zones (light gray).

Several areas corresponding to places of interest ("TARGETS"), forbidden zones, or other special areas ("SPECIAL") has been defined in the environment. The zone module continuously monitors the entrance and the exit of the robot from these zones and informs the supervisor.

2) Obstacles and people detection: Obstacle detection is a critical function both for security reasons and for interaction purposes. The most efficient sensor is once again the laser. However our laser can only look forward (over 180 degrees) in an horizontal plan. To partially overcome these limitations, the laser data are integrated in a local map by the aspect module and filtered using knowledge about the global map, its segments and the virtual obstacles. Every 40 milliseconds, aspect exports a local map all around the robot which represents the free space and which distinguishes static (i.e, that belong to the environment or the virtual obstacles) and dynamic obstacles (probably visitors). This local map is permanently displayed on the bottom right of the interface (see figure 15).

Using this representation, aspect is able to inform the supervisor when the robot is surrounded by unpredicted obstacles. The red LED's on the helmet flicker at a frequency proportional to the obstruction density by dynamic obstacles.

To reinforce the assumption of presence near the robot, the supervisor can use the services of the sono module that detects motion all around the robot using ultrasonic sensors. Unfortunately our ultrasonic sensors produce some audible(!) noise which seems to disturb visitors interacting with the robot.

A much more robust people detector is offered by a module called isy (or, "I See You") which is able to detect a face in real time from one color camera image. The detector uses a cascaded classifier and a head tracker based on a particle filter (see [4]). Isy controls the camera orientation in order to follow the detected face as long as possible. It informs the supervisor when it catches or looses a face. From the direction and the size of the face it is able to estimate the 3D position of the detected person with a sufficient precision (about 10cm for the height and 20cm for the range).

3) Trajectory and motion: Rackham being a guide, it must be able to take visitors to places of interest in the exhibition. These places are displayed on the interactive map. For the robot they correspond to a polygonal `target zones` (see §V-C.1) and to the position of the element of interest (which can be itself out of the polygon) that the robot will have to comment.

The robot motion implies mainly three modules :

- `rflex` that manages the lower servo-control loop, transmitting the reference speeds at the micro-controller.
- `ndd` integrates a local avoidance procedure based on an algebraic instance of Nearness Diagrams (see [18]). The input obstacles are provided the aspect map (see §V-C.2).
- `vstp` is a Very Simple [but very efficient] Trajectory Planner based on an algebraic visibility graph optimized with hash tables[6]. A main visibility graph is precomputed for the static segments of the map. Dynamic obstacles can be added and removed in real-time upon supervisor requests.

The strategy used to coordinate the implied modules is dynamically established by the supervisor. The objective is of course to reach the target zone while avoiding obstacles. The planned trajectory is an Ariadne's clew for `ndd`: the vertices of the broken line are sub-goals. Usually the supervisor has to intervene only if `ndd` does not progress anymore along this path. In such a case, various strategies can be applied: computing of a new trajectory taking into account the encountered obstacles, waiting for a while, starting an interaction with people around, etc.

The maximum speed that the robot can achieve in this mode is about 0.6 meters per second.

4) Interactions: For now the interactions are mainly established through the following components:

- the dynamic "obstacles" detectors (`aspect` and `sono`),
- the `isy` face detector,
- an animated face with speech synthesis,
- displays and inputs from the touch-screen,
- control of the robot lights.

While the two firsts allow to detect the presence or the departure of people, the last ones permit the robot to "express" itself and thus establish exchanges.

The 3D head embedded in the screen interface can talk, thanks to the audiovisual speech synthesis system: audio was generated along with the synchronized movements of the face (for the lips, the jaw, the cheeks...). Meaningful messages have been prepared, corresponding to the various situations encountered by the robot or to the places that will need to be described during the visit.

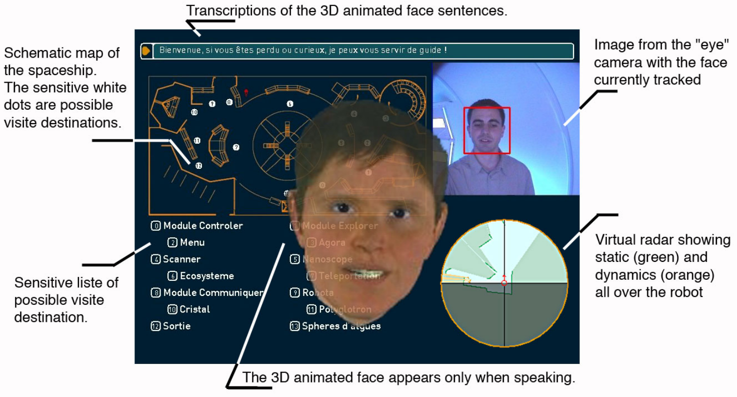
The neck and the eye movements of the 3D head have independent controls, that can be driven by the facial tracking module. That way, the user knows which messages are specifically addressed to him, and that the system is still aware of his or her presence, or has become aware of the user, even in situations where the screen is not yet facing the user perfectly: These synthesized movements take place on screen faster than the physical movements or orientation changes of the robot in real world. The orientation are computed thanks to the position of the interlocutor face detected by `isy` and to the location of the robot in the map maintained by the system.

To help interaction with both the graphical interface and the real world, the talking head can also synchronize some eye gazes and eye movements with the uttering of some keywords. For example, when talking about the "map", the clone face and eye gazes can direct the user to where on screen the map is displayed. As the robot knows its location and orientation in the world, he can also indicates a nearby object of interest, still using face and eye gaze when the corresponding keyword is being pronounced.

The robot interface, written with Java, is made of independent components or micro-GUI directly controlled by the supervisor through a dedicated communication channel.

The available micro-guis are (figure 15):

- a map of the environment including the current robot position and trajectory
- the local "aspect" map displayed as a radar
- the image of the "eye" camera with the faces currently detected by isy
- the clone or talking head
- pop-up warning messages
- top messages
- localization window (init).



Bienvenue, si vous êtes perdu ou curieux, je peux vous servir de guide !

Transcriptions of the 3D animated face sentences.
Schematic map of the spaceship. The sensitive white dots are possible visite destinations.
Image from the "eye" camera with the face currently tracked
0 Module Controler
2 Menu
4 Scanner
6 Ecosysteme
8 Module Communiquer
10 Cristal
12 Sortie
1 Module Explorer
3 Agora
5 Nanoscope
7 Teleportation
9 Robota
11 Polyglotron
13 Spheres d'algues
Sensitive liste of possible visite destination.
Virtual radar showing static (green) and dynamics (orange) all over the robot
The 3D animated face appears only when speaking.

Fig. 15. A view of the interface of the touch-screen.

D. Supervision

Rackham is used in a context where there is no need for a high level planner i.e. a system that synthesizes a partially ordered set of tasks to be performed to reach a given goal. Consequently, the highest level of decision is to select what

[6] VSTP is freely distributed: http://softs.laas.fr/openrobots/.

task to achieve. Indeed, the robot is able to perform a number of tasks in a variety of contexts and depending on various conditions (availability of visitors, energy level. . .).

Hence, the role of the robot supervisor involves several aspects:

- task selection,
- context-based task refinement,
- adaptive task execution control.

In its current configuration, Rackham, as a tour-guide in the exhibition, has basically to deal with two different tasks: *the search for interaction* (where the robot, left alone in the exhibition, tries to attract a visitor in order to interact with him), *the mission* (where the robot, according to the visitor's choice, brings him to a selected place).

Our choice was to use relatively low level observation and action primitives in order to leave as much flexibility as possible at the supervision level. Indeed, as we will see in the sequel, the performance of tasks in the vicinity and/or in interaction with humans is not compatible with a "black-box" strategy.

Another interesting aspect on which we focus is how the task execution process is influenced by the need for human-robot interaction.

When a task is given, our robot not only needs to execute it, but it also needs to be able to explain it (by exhibiting a legible behavior or by displaying relevant information) and it should allow humans to act on the course of its actions during their execution.

For instance, during the *mission* task, Rackham should not only be moving toward its goal and avoiding obstacles, it also has to maintain the interaction with the humans (waiting for possible inputs like abort or change the mission and displaying any relevant information that may be needed).

There are a number of speech-based or visualization-based functions that allow to give feedback to the user mainly in terms of messages. Other information such as trajectory, robot position, etc, are displayed directly by the interface as soon as there are available.

VI. Results and future work

Between march 2004 and February 2005, Rackham has spent ten weeks at the ”Cité de l'Espace” in five venues[7].

During the last two stays, the robot was sufficiently robust to be operated by the personnel of the Cité de l'Espace without our intervention.

We collected various data for analysis purposes: all the requests to the modules and their reply, the covered distance, the visitors interactions, etc. The results presented below are a synthesis of the data collected during the last two stays. Rackham has executed 1575 missions requested by the visitors of the exhibition and traversed nearly 16.5 km.

Fig. 16. Head of Rackham emerging from a sea of kids.

From October 5, 2004 to October 15, 2004					
day	number of missions	distance in meters	duration hh:mn (motion)	number of requests	average speed (km/h)
1	17	71	0:34	379	0.44
2	63	543	2:39	2100	0.57
3	46	495	1:27	2210	0.61
4	9	100	0:11	318	0.63
5	76	815	2:15	2377	0.63
6	97	802	2:20	2967	0.54
7	54	542	2:12	2081	0.52
8	89	904	3:41	2810	0.59
9	54	607	2:19	1751	0.60
10	58	681	1:57	2019	0.58
11	170	1611	5:37	5084	0.57
	733	**7171 m**	**32:12**	24096	**0.57**

From February 7, 2005 to February 20, 2005					
day	missions	distance	duration	requests	speed
1	40	395	1:25	2801	0.51
2	49	555	1:32	2719	0.56
3	44	487	1:18	2557	0.62
4	82	851	3:32	4338	0.44
5	82	881	2:28	4209	0.58
6	70	739	1:49	3609	0.56
7	85	884	2:14	4338	0.50
8	71	815	2:24	3984	0.53
9	55	663	1:31	3154	0.60
10	78	912	2:29	4742	0.49
11	71	872	2:08	4214	0.54
12	91	994	2:49	4632	0.53
13	14	161	0:27	733	
	832	**9209 m**	**26:06**	46030	**0.54**

[7] see http://www.laas.fr/ sara/laasko.

Although the HR+ project has reached its end, further experiments and more detailed analysis of the collected data will be conducted in the future. HR+ has also opened for us several research issues such as the need to elaborate a framework for sharing decisions and actions between the robot and the human and more generally for collaborative problem solving.

Acknowledgments

Numerous persons have contributed to the project.

The authors want to thank: A. Arnal, Abdelatif Baba, Alban Caporossi, Maxime Cottret, Xavier Dollat, Véronique Hallard, Guillaume Infantes, Stephan Lauriol, Coline Lelong-Pantel, Florent Lanterna, H. Loevenbruck, Jérôme Manhes, Jean-Noël Plachez, Stéphane Plais, Maxime Randisi, C. Savariaux, Aymeric Yvernes.

References

[1] R. Alami, R. Chatila, S. Fleury, M. Ghallab, F. Ingrand, “An architecture for autonomy”, *International Journal of Robotic Research*, Vol.17, N°4, pp.315-337, Avril 1998.

[2] G. Bailly, M. Bérar, F. Elisei, and M. Odisio, “Audiovisual speech synthesis”, *International Journal of Speech Technology*, 6:331-346, 2003.

[3] G. Bailly, F. Elisei, and S. Raidt “Multimodal Face-to-Face Interaction with a Talking Face: Eye Gaze, Mutual Attention and Deixis” submitted to HCI-2005

[4] L. Brèthes, P. Menezes, F. Lerasle, and J. Hayet, “Face tracking and hand gesture recognition for human robot interaction,” *International Conference on Robotics and Automation, New Orleans, pp. 1901-1906, May 27 - June 1 2004*.

[5] L. Brèthes, F. Lerasle and P. Danès, ”Data Fusion for visual Tracking dedicated to Human/Robot Interaction,” Int. Conf. on Robotics and Automation, April 2005.

[6] A. Clodic, S. Fleury, R. Alami, M. Herrb, R. Chatila “Supervision and Interaction. Analysis from an Autonomous Tour-guide Robot Deployment”, submitted to ICAR 2005.

[7] J. L. Crowley and H. I Christensen, “Vision as Process”, Springer Verlag, Heidelberg, 1993.

[8] J. L. Crowley and F. Berard, “Multi-Modal Tracking of Faces for Video Communications”, IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’97, St. Juan, Puerto Rico, June 1997.

[9] J. L. Crowley, J. Coutaz and F. Berard, “Things that See: Machine Perception for Human Computer Interaction”, Communications of the A.C.M., Vol 43, No. 3, pp 54-64, March 2000.

[10] J. L. Crowley, J. Coutaz, G. Rey and P. Reignier, “Perceptual Components for Context Aware Computing”, UBICOMP 2002, International Conference on Ubiquitous Computing, Goteborg, Sweden, September 2002.

[11] J. L. Crowley, “Context Driven Observation of Human Activity”, European Symposium on Ambient Intelligence, Amsterdam, 3-5 November 2003.

[12] S. Fleury, M. Herrb, R. Chatila, “GenoM: a Tool for the Specification and the Implementation of Operating Modules in a Distributed Robot Architecture”, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Grenoble, France, 1997.

[13] D. Hall, James L. Crowley, Détection du visage par caractéristiques génériques calculées à partir des images de luminance, *Reconnaissance des Formes et Intelligence Artificielle* (RFIA), 2004

[14] D. Hall, A. Caparossi, Agent tracking and identification in video sequences, submitted, 2004

[15] A. Lux, ”The Imalab Method for Vision Systems”, International Conference on Vision Systems, ICVS-03, Graz, april 2003.

[16] P. Menezes, J.C. Barreto and J. Dias, ”Face tracking based on Haar-like Features and Eigenfaces,” Int. Conf. on Advanced Vision, July 2004.

[17] P.Menezes, F.Lerasle, J.Dias and R.Chatila, ”Monocular Visual Tracking of 3D Articulated Structures using Particle Filters,” Int. Conf. on Intelligent Robots and Systems, Submission, May 2005.

[18] J. Minguez, L. Montano. “Nearness Diagram Navigation (ND): Collision Avoidance in Troublesome Scenarios”,*IEEE Transactions on Robotics and Automation*, p 154, 2004.

[19] J. Piater and J. Crowley, ”Event-based Activity Analysis in Live Video using a Generic Object Tracker”, Performance Evaluation for Tracking and Surveillance, PETS-2002, Copenhagen, June 2002.

[20] L. Reveret. G. Bailly and P. Badin, “MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation”. International Conference on Speech and Language Processing, 755-758, Beijing - China, 2000.

[21] K. Schwerdt and J. L. Crowley, ”Robust Face Tracking using Color”, 4th IEEE International Conference on Automatic Face and Gesture Recognition”, Grenoble, France, March 2000.