Implementing a Human-Aware Robot System

E. Akin Sisbot, Aurélie Clodic, Luis F. Marin U., Mathias Fontmarty, Ludovic Brèthes, Rachid Alami

LAAS/CNRS
7, Avenue du Colonel Roche 31077 Toulouse, France
Email: LastName@laas.fr

Abstract— **The presence of humans in the robot environment brings new challenges to the robotic research. From low level functions to high level planners, clearly the human has to be taken into account in all the layers of the robot control system. Indeed, the robot has to behave socially in order to interact friendly with its human partners.**

This paper describes the development of several components (human detection and tracking, planning and supervision) that take into account humans explicitly with some preliminary results of their integration.

I. INTRODUCTION

The introduction of robots in human environments raises a key issue that is "added" to the "standard challenge" of autonomous robots: the necessity to interact with humans and to behave socially. From the low level functions to the high level planners, clearly the presence of humans has to be taken into account in all the layers of the robot control system.

We investigate mobile robot abilities to navigate in the presence of humans and to approach persons in order to interact with them or to hand objects to them. To accomplish this the robot must be able to detect human presence in its vicinity and to interpret their trajectories, in order to secure their displacements —by avoiding them or getting out of their way— or to interact with them.

We report here on a contribution to the development of people detection and tracking modalities based on the use of a color camera and a laser range finder, as well as a Human Aware Navigation Planner and a supervision system that take explicitly into account humans.

The overall system has been implemmented on a robot called Rackham (figure 1), a B21r robot (iRobot). It is a 52 cm wide and 118 cm tall cylinder topped with a mast supporting a kind of helmet. It integrates 2 PCs (one mono-CPU and one bi-CPUs running P3 at 850 MHz). The standard equipment is extended with a pan-tilt Sony camera EVI-D70 attached under the helmet, a digital camera mounted on a Directed Perception pan-tilt unit, an ELO touch screen, a pair of loudspeakers, an optical fiber gyroscope and wireless Ethernet.
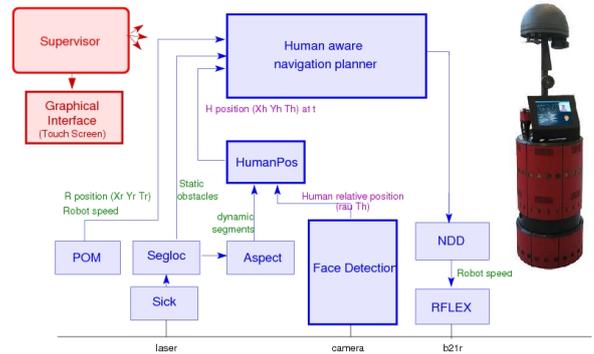
Fig. 1. Rackham and its Software architecture with all the modules involved

The paper is organized as follows: section II discusses an example of a human-robot interaction scenario, sections III, IV and V present three components which are visual human detection, laser human detection and navigation planner modules. Finally, section VI describes the supervision system.

II. SCENARIO

The environment model is assumed to have been acquired previously. Such a model allows the robot to navigate and to monitor its environment. Rackham is in a room, motionless, and faces entrance to detect and track any person entering. A person comes in. If the person approaches the robot, the robot will presents itself. If, after that, the person stays next to the robot, it will propose its services, otherwise the interaction is cancelled.

For now, the service proposed by the robot is to guide the person in the room, *i.e.* to conduct him to a given target point chosen between many destinations and to verbally present it. So, once the person has chosen the target, the robot plans a trajectory and navigates to the desired position. One originality comes from the fact that Rackham navigation has been specially designed to be human-friendly.

The human presence and his willingness to interact are checked during the task achievement and especially along the trajectory execution, e.g. if visual contact can't be established, the robot stops and asks for it. At any moment, the robot should be able to detect the presence and monitor the commitment of the person and to react accordingly. If the human shows no interest or if he has left, the robot abandons its current task considering that it is no more relevant.

There are four key components of the system: ICU – visual face detection and tracking module, HumPos –human detection and tracking module, NHP – Navigation in human presence module and finally a supervision system that controls the task and manages the overall system.

These components are developed using OpenRobots [8], a software environment that allows to build modules that can integrated in the LAAS architecture [7] (see figure 1).

III. ICU - DETECTION AND TRACKING SERVICES

The services provided by the ICU ("I See You") module are related to human detection and tracking using visual primitives. Our visual tracking strategy consists in fusing visual cues into particle filtering schemes. This fusion is performed in the definition of the particles likelihoods. In addition, outputs from detection primitives can be involved in the importance functions of the particle filters.

a) Implementation: The following detection services have been developed and integrated on the robot:

- a *face detector* computes the number of detected faces together with their positions and scales in a B/W or RGB image, using Haar-like features;
- a *face detector dedicated to tracking* computes a Gaussian mixture related to detected faces in an B/W or RGB image;
- a *motion detector dedicated to tracking* computes a Gaussian mixture related to the motion activity in a B/W or RGB image;
- a *skin-color blobs detector dedicated to tracking* computes a Gaussian mixture related to the detected color blobs in an RGB image.

These detectors have been used to set up three visual tracking modalities for H/R interaction, namely:

- a *head tracking modality* for interaction at short-range H/R distance (Figure 2 (a));
- a *human tracking modality* for interaction at mean-range H/R distance (Figure 2 (b));
- a *monitoring modality* for long-range H/R interaction (Figure 2 (c)).

Besides, the ICU module provides requests to control the embedded camera, and to manipulate the three previously described modalities. It can handle internally the switch between modalities, based on the *face detector* results (frequency and size of the detections). ICU can also approximate the location of a person at short or mean distance to the robot. All the tracking outputs from the various modalities are available for the other modules and more specifically for the supervisor.

b) Adopted Strategies: A preliminary evaluation enabled to select the most meaningful visual cues associations in terms of discriminative power, robustness to artefacts (clutter or illumination changes) and time consumption, be these cues involved in the particles likelihoods or in the filters' importance functions (See [20] for more details). As a result, dedicated visual cues were selected for each modality, see below.
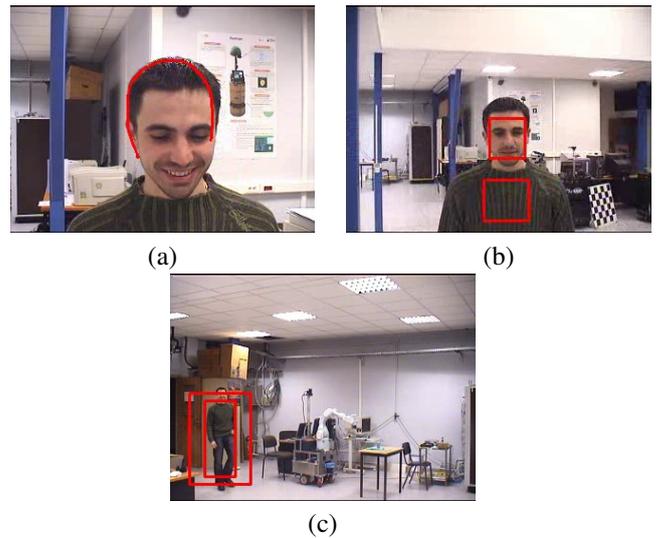

Fig. 2. Modalities provided by the ICU module

Several particle filtering strategies were evaluated in order to check which ones best fulfill the requirements of the considered H/R interaction modalities. The evaluated strategies are CONDENSATION [12], ICONDENSATION [13], the Auxiliary Particle Filter [14] and the Rao-Blackwellized Subspace History-Sampling Sampling Importance Resampling (RBSSHSSIR) algorithm [15]. For the sake of comparisons, importance functions rely on dynamics or measurements alone, or combine both. Further, each modality has been evaluated on a database of sequences acquired from the robot in a wide range of typical conditions: cluttered environments, appearance changes or sporadic disappearance of the targeted person, jumps in the dynamics… For each sequence, the mean estimation error with respect to "ground truth", together with the mean failure ratio (% of target loss), were computed from several filter runs.

Eventually, the *head tracking modality* combines motion and shape cues into a CONDENSATION algorithm, the *human tracking modality* merges shape and color information into a RBSSHSSIR algorithm, while the *monitoring modality* relies on color and motion distributions into an RBSSHSSIR algorithm.

IV. HUMPOS - HUMAN DETECTION MODULE

The HumPos module provides human detection and tracking services based on laser data. It feeds the motion planner with a list of humans in the environment. This list contains positions, orientations of the detected humans associated with a confidence index. The module is composed of three stages. The first two stages make use only of laser data to detect legs while the last one integrates information delivered by the ICU module.

At the first phase of the human detection, HumPos receives small dynamic segments (i.e. that are not considered as part of the environment map) created from laser scans, it selects those segments that could potentially form legs by their

relative configuration and shape, then the resulting legs (fig. 3-A) are coupled and form a list of items.

In the second phase, HumPos receives raw laser data and creates legs from a group of points, called blobs, obtained in a neighborhood and then these blobs are coupled by their proximity. Similar work is explained in [9] to detect cylinder and lines to find legs by analyzing their geometric characteristics. The resulting items obtained in this phase (fig. 3-B) are compared with those of the first phase. Matching items are sent to the last stage of the detection process.

The third and final phase involves the matching with the position estimation obtained from ICU (fig. 3-C). This phase increases dramatically the confidence of the matched items to be humans. In [10] a similar work is conducted by combining camera and laser to track people. The main difference comes essentially by the fact that our vision system allow to estimate the positions of the humans relatively to the robot.
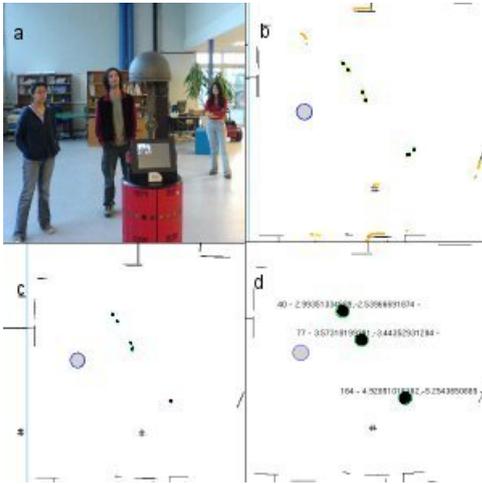

Fig. 3. Results of each phase in the HumPos module, a) a photo of the situation, b) legs obtained form laser raw data indicated by the colored circles, c) legs obtained from dynamic segments, d) humans detected. The black circles indicate detected legs (small) or humans (big). The circle in center left position of a), b) and c) images and in lightgray color marks the robot's position in the global environment map.

For the human tracking, a simple method is developed to estimate humans' next position and orientation (supposing that people naturally face to their walking direction). In [11] or [16] a particle filter for tracking moving objects with a laser scan is applied with good results, but we needed a lighter and simpler way to track items while leaving enough processor resources to the other modules.

Finally, a list of humans containing its Id's, positions and orientations in environment are made available to NHP.

V. NHP - NAVIGATION IN HUMAN PRESENCE

The navigation module is an implementation of the Human-Aware Navigation Planner[1]. This planner computes acceptable and human-friendly paths by reasoning on humans' state, their vision field and their accessibility. User studies[2][3][4] show a number of properties like keeping a certain distance, maintaining eye contact, etc. that must be taken into account in human-robot interaction context. From these properties we extracted two different criteria, namely "Safety" and "Visibility" criteria, that must be taken into account for a safe and friendly human-robot interaction. And these two criteria serve as the core of the planner and separate it from other path planners. Each criterion is represented by a set of numerical value stored in a 2D grid combining various costs depending robot's position in the environment. One can consider these grids as a set of cells containing various costs derived from the relative position to the human.

Safety Criterion: The first criterion, called "safety criterion", focuses on ensuring the safety by controlling the distance between robot and humans. By this property the robot is encouraged to keep a certain distance to the people in the environment. However in some cases, as the necessity of their interaction, the robot has to approach to a person whom it wants to interact with. Hence, this distance between the robot and the human is neither uniform nor fixed and depends on the properties of their interaction.

These properties are presented in the current system by "the safety grid". Figure 4 illustrates this grid containing a human centered gaussian form of cost distribution. The height of vertical lines represents the value of the cost associated with each cell.

Visibility Criterion: Another important property of a human-robot interaction is maintaining a certain level of visibility. This criterion, called "visibility criterion", aims to improve humans' comfort by encouraging the robot to prefer to stay in the humans field of view. This criterion allows to take into account the humans' vision field and allows us to produce more "acceptable" robot motions. The visibility grid is constructed (figure 4) according to costs reflecting the effort required by the human to get the robot in his field of view. Evidently passing in front of a human is preferable for the robot than passing behind of him.

Both of these criteria loses their importance inverse proportionnaly to the distance to the humans, so visibility and safety are not taken into account when the robot is far from humans.

An Extension: Hidden Zones: In the grids illustrated above, the costs are calculated without taking into account the obstacles in the environment. However, obstacles in the close vicinity of the human can have various effects on the safety and comfort. If the robot is behind an obstacle, the human would feel much comfortable because the obstacle would block the direct way between human and the robot. So the safety criterion must be cancelled in the zones located behind the obstacles.

On the other hand, when the robot becomes hidden by an obstacle the visibility costs lose their importance. To handle this issue, we introduce an extension to visibility and safety, called "hidden zones" criterion. This criterion helps to determine better costs for positions hidden by the obstacles.

An important effect of obstacles to the comfort of the human is the surprise factor. When the robot is hidden by an obstacle and loom in the human field of view, it can

cause surprise and fear especially if it is close to the human. To avoid this effect, we must discourage the robot to pass behind an obstacle too closely, and must allow it to get into the humans field of view when sufficiently far away. This is done by putting costs to the zones hidden from the view because of the obstacles.
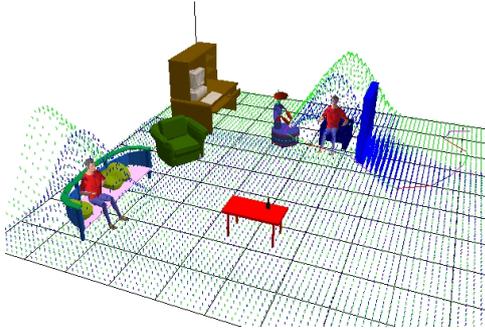

Fig. 4. The grids of the two criteria. A Safety (in blue) grid is built around every human in the environment. Visibility grid (in green) is computed by taking into account humans' field of view. The zone hidden by the wall has additional costs.

Each human in the environment has his own Visibility and Safety grid. Not only the shape of each grids differs one from another but also each human in the environment can have different parameters, so different shapes, for the same type of grid. These grids are combined to form a final grid in which we find a minimum-cost trajectory for the robot by using A* search algorithm. The synthesized path is not only collision free and feasible but also human friendly with satisfying both criteria. One must note that none of the grid, is constructed explicitly. The cost of a cell in the grids, including the final grid, is calculated in case of a request from A*. This approach allows us to the planning process and limit memory space usage.

As in real world scenarios there can be multiple objects and humans in motion, if the environment changes during execution, the calculated path can loose its validity. This path can be replaced very rapidly by replanning. As the module's planning method is fast, replanning allows us to be sufficiently reactive to the environment changes and allows us to execute paths smoothly by making the replanning totally transparent to the humans.

The "Human-Aware Navigation Planner" is implemented within the Move3D [5], a 3D geometric planner, and grids built the core of the planner module. This module works with a static internal 3D map along with each humans model, his grid construction parameters and the robot model. The humans' positions are updated by the HumPos module and the robot's current position is updated by the Position Manager. With all these inputs, the navigation planner module calculates a path and sends it to the path execution service. In figure 5, we see the architecture of NHP as well as its
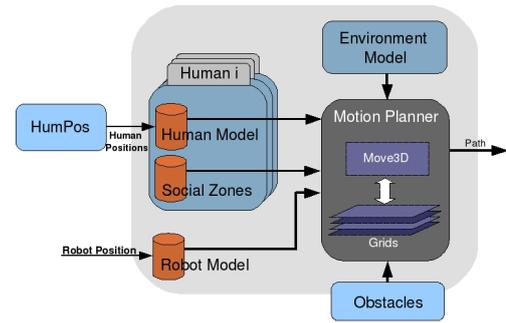

Fig. 5. The architecture of the Human-Aware Navigation Planner module showing the inner working and inputs/outputs of the planner

inputs and outputs.

As the planner currently does not take into account the dynamic obstacles and does not update its map, we compensated this gap by using a complementary module, namely NDD[6], that executes and makes local modifications of the path produced by the planner in order to avoid collisions.

Fig. 6. A comparison between a classic motion planner and the human-aware navigation planner. Clearly the last one produces a more acceptable path by taking into account the safety and visibility of each human in the environment

In figure 6, we can see clearly the safety and comfort difference between a classic path planner and our Human-Aware Navigation Planner. In this scenario, there are two persons in the robot's environment talking. One of them has his back turned to the robot and thus does not see the robot. The robot's goal is to go to the other corner of the room. A classic motion planner results (fig. 6-A) with a straight line path from one corner to the other and the human collision avoidance is obtained by an obstacle avoidance at the execution. In figure 6-B, the robots motion planning system is replaced by the NHP. The path is produced by taking into account humans' positions and their orientations.

The figure 7 contains paths calculated during the motion of the robot. The robot begins his motion with a straigth path (fig.7-1), when it detects a human, it immediately changes his path according to his position and his field of view (fig.7-2). The path is more deviated when the second human gets in the vision of the robot (fig.7-3). The robot continues this motion by constantly cheking environmental changes (fig.7-4,5,6).
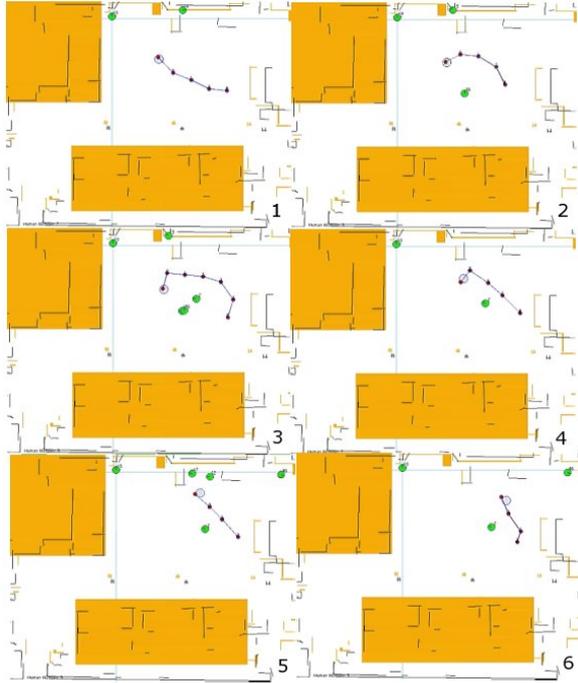
Fig. 7. Paths produced by our "Human Aware" Navigation Planner. During the motion of the robot, once the humans are detected the planner changes the current path to a more friendly one that takes into account the field of view and position of humans.

VI. SUPERVISION

We have developed and implemented an agent-based supervision system that integrates cooperative task achievement inspired by joint intention theory [21]. Every human entering the close vicinity of the robot is considered as an agent with whom the robot might collaborate or at least interact. To be homogeneous, the robot itself is also considered as an agent. Each agent is characterized by its abilities (the tasks that it can perform) and its potential commitment to the task at hand. Once involved, sensor-based monitors help to measure each agent involvement and belief about a task state (unachieved, achieved, impossible, irrelevant). This information is then used by the robot to achieve its task and to adapt it to human action, reaction or lack of reaction. In our system, tasks may be defined as individual tasks (only the robot is involved), joint tasks (the robot and another agent are involved) and activities that correspond to low level functionalities that are no further decomposed. Tasks are refined hierarchically a succession of sub-tasks and/or activities.

Consequently, the system can be controlled at different levels at the same time. If something happens at a given level, the system is able to take it into account at that level by applying adapted solutions and propagating, when necessary, events towards the higher or the lower levels.

This supervision system is a key component of a more general framework defined in [18]. More detailed information about the supervision system can be found in [19].

In the current implementation, the supervisor is written in Open-PRS[8]. The task plans are hand-coded (a set of predefined task library) and only the robot is able to propose a task. However, the supervisor is built, taking into account future extensions involving on-line task planning.

This supervision scheme has been tested on Rackham (Figure 8) according to the scenario with the following plan: **Initialization**, **Wait For People**, **Want to interact ?**, **Guide**. The system reacts in the following way.

Wait For People Rackham stands still and monitors the room entrance (fig.8-a) . If someone enters, Rackham detects him using a monitoring modality of ICU. Ludovic enters in the room. Once detected, the robot has to check if he wants to interact with it. **Want to interact ?** checks if Ludovic approaches (*human tracking*, *head tracking* and finally *face detection* of ICU module). If Ludovic comes closer (fig.8-b), Rackham presents itself (otherwise it will do nothing to not disturb the person) : "hello, i'm Rackham and i can guide you in the lab"(fig.8-c).

Ludovic shows interest, the robot proposes to engage with him: "tell me or click on the screen to choose your destination", there begins the joint task **Guide**.

The robot is the first committed to the task. Following Rackham proposition, Ludovic is informed that the robot is committed to the task, we write:

- (commitment **Guide** Rackham Rackham yes)
- (commitment **Guide** Ludovic Rackham yes)

Now, if Ludovic agrees the proposition and chooses a destination, we obtain: (commitment **Guide** Ludovic Ludovic yes). If Ludovic goes away, that will be: (commitment **Guide** Ludovic Ludovic no) and the joint task will be abandonned by the robot.

So, we have seen that our model includes three commitments that are complementary: the robot commitment to the task, the robot's belief concerning Ludovic commitment to the task and the robot's belief concerning Ludovic's knowledge about the robot commitment to the task.

Now, commitment to the task is established by both parts. Then the robot confirms "ok, i will guide you to that destination. Please, follow me.". The task achievement begins (fig.8-d). Rackham will bring Ludovic to the destination and Ludovic is supposed to follow it and both are supposed to consider the task as unachieved. We consider 3 state beliefs:

- (state **Guide** Rackham Rackham unachieved)
- (state **Guide** Ludovic Ludovic unachieved)
- (state **Guide** Ludovic Rackham unachieved)

If during the task, Rackham's navigation stops for any reason while the destination is not reached, there will be a change: (state **Guide** Ludovic Rackham achieved) i.e. we consider that the robot stop could mislead Ludovic thinking that the task is over, that represents, the robot belief concerning the agent belief concerning the robot. In the same way, if during the task, Ludovic goes away, the change will occur concerning his state: (state **Guide** Ludovic Ludovic achieved).

Each case leads to a difference in state beliefs, i.e. no more mutual belief. According to the case, the robot will

Fig. 8. Rackham looking at the room entrance and then interacting and guiding someone.

react differently, e.g. in the case Ludovic goes away, it will stop, ask and display "where are you ?" and the state will change to unachieved if Ludovic comes back (detected by the help of the *face detection* or if he touches the screen), if he does not the state will change to impossible and the task will be abandonned.

It is possible to define as many state monitors as needed and they could be tagged to allow better reaction: (monitor **Guide** Ludovic Rackham achieved LOC_PB)

At the end of the task, Rackham has also to ensure that there is a mutual belief concerning the achievement of the task.

Comparatively to the previous version of our system ([17]), it is easier to add new capabilities to the robot and to program new tasks involving interaction. In addition, monitoring is facilitated by the help of layered levels that are all reactive.

VII. CONCLUSION

We have presented several key components of our Human-Aware Robot architecture. Individual results of each component are good and the working of the system as a whole is promising.

More examples and videos showing the results of the individual components as well as the whole system can be found at http://www.laas.fr/~easisbot/roman06 .

This is a first implementation of a system and a scenario that will be extended toward manipulation scenario where robot and humans accomplish manipulation tasks cooperatively. This goal requires designing a human aware manipulation planner. Strengthening the link between HumPos and ICU modules will result in a better detection of humans in the environment.

In the near future, the supervision system will be improved with the integration of an on-line planning ability that explicitly considers Human-Robot Interaction and Collaborative Problem Solving. Concerning the supervisor itself, we envisage to extend it in order to be able to take into account more than two agents (the robot + a human) by task and to manage several tasks simultaneously at the robot level.

REFERENCES

[1] E. A. Sisbot, R. Alami, T. Simeon, K. Dautenhahn, M. L. Walters, S. N. Woods, K. L. Koay, C. L. Nehaniv. "Navigation in the presence of humans", Proc. IEEE-RAS International Conference on Humanoid Robots, December 5-7, 2005 Tsukuba, Japan. pp. 181-188.

[2] E, Pacchierotti, H. Christensen and P. Jensfelt, "Human-Robot Embodied Interaction in Hallway Settings: a Pilot User Study", IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man), Nashville, USA, 2005

[3] K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, C. Nehaniv, E. Sisbot, R. Alami, T. Simeon, "How may I serve you?, A Robot Companion Approaching a Seated Person in a Helping Context", Proc. 1st Annual Conference on Human-Robot Interaction (HRI06), Salt Lake City, Utah, USA.

[4] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv,D. Lee, I. Werry , "The Influence of Subjects Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment", Proc. IEEE Ro-man 2005, Nashville, August 2005.

[5] T. Siméon, JP. Laumond, F. Lamiraux, "Move3D: a generic platform for motion planning", 4th International Symposium on Assembly and Task Planning, Japan,2001.

[6] J. Minguez, J. Osuna, L. Montano, "A Divide and Conquer Strategy based on Situations to Achieve Reactive Collision Avoidance in Troublesome Scenarios", Proc IEEE International Conference on Robotics and Automation (ICRA), 2004. New Orleans, USA.

[7] R. Alami, R. Chatila, S. Fleury, M. Ghallab and F. Ingrand, "An architecture for autonomy", International Journal of Robotic Research, Vol.17, N°4, pp.315-337, April 1998.

[8] LAAS OpenRobots, http://softs.laas.fr/openrobots/index.php

[9] Joo Xavier, Marco Pacheco, Daniel Castro and Antonio Ruano, "Last line, arc/circle and leg detection from laser scan data in a Player driver" In IEEE International Conference on Robotics and Automation - ICRA 2005 , Barcelona, April 18-22, 2005.

[10] M. Kleinehagenbrock, S. Lang, J. Fritsch, F. Lomker, G. A. Fink and G. Sagerer, "Person Tracking with a Mobile Robot based on Multi-Modal Anchoring" In Proc. IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN), Germany, September 2002.

[11] D. Shulz, W. Burgard, D. Fox, and A.B. Cremers, "Tracking multiple moving objects with a mobile robot", In. Proc. of the IEEE Computer Society Conference on computer vision and pattern recognition (CVPR),Kauai,HW, December 2001.

[12] A. Blake and M. Isard, "The CONDENSATION Algorithm — Conditional Density Propagation and Applications to Visual Tracking", Advances in Neural Information Processing Systems, vol. 9, 1997.

[13] Michael Isard and Andrew Blake, "ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework", Lecture Notes in Computer Science, vol. 1406, p. 893-908, 1998.

[14] Michael K. Pitt and Neil Shephard, "Auxiliary Variable Based Particle Filters", in Sequential Monte Carlo Methods in Practice, A. Doucet, N. de Freitas, and N. Gordon, Eds., chapter 13, Springer, 2001.

[15] Peter Torma and Csaba Szepesvari, "Sequential Importance Sampling for Visual Tracking Reconsidered", Proc. AI and Statistics, 2003.

[16] Abedallatif Baba and Raja Chatila, "Simultaneous environment mapping and mobile target tracking", 9th International Conference on Intelligent Autonomous Systems (IAS-9), Tokyo, 2006.

[17] Clodic, Aurélie and Fleury, Sara and Alami, Rachid and Herrb, Matthieu and Chatila, Raja, "Supervision and Interaction: Analysis from an Autonomous Tour-guide Robot Deployment", In the 12th International Conference on Advanced Robotics, 2005.

[18] Clodic, Aurélie and Montreuil, Vincent and Alami, Rachid and Chatila, Raja, "A Decisional Framework for Autonomous Robots Interacting with Humans", IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN), 2005.

[19] Clodic, Aurélie and Montreuil, Vincent and Alami, Rachid and Chatila, Raja, "Models and Algorithms for robot supervision in HRI context", Laas Progress Report, 2006.

[20] L. Brèthes , F. Lerasle , P. Dans, "Data fusion for visual tracking dedicated to human-robot interaction", IEEE International Conference on Robotics and Automation (ICRA 05)",2005.

[21] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213–361, 1990.