HUMAN PRESENCE DETECTION AND TRACKING FOR A CONCIERGE ROBOT

Iñaki Rañó¹, Bogdan Raducanu², Sriram Subramanian³

¹ Parque Tecnologico de San Sebastian
Paseo Mikeletegi 53 (Ed. Central), 20009 San Sebastian, Spain

² Computer Vision Center, Edifici Q, Campus Universitat Autonoma de Barcelona
08193 Bellaterra (Barcelona), Spain

³ Department of Computer Science, University of Saskatchewan
1C129 Engineering Building, 57 Campus Drive, Saskatoon SK S7N5A9, Canada

Abstract: This paper reports on the development of a social-oriented mobile robot, that is able to reliably navigate in a semi-structured environment and to interact with people in some specific way. The navigation of the robot is achieved using a behaviour-based approach, where properly ordered sequences of actions, turns and transitions bring the robot from place to place. The robot can detect the presence and position of people through its vision system, that has been adapted to detect a special badge carried by the user. The result of this work is a “concierge robot” able to offer practical information to the visitors of a public building.

Keywords: autonomous mobile robot, robot navigation, robot dynamics, computer vision, pattern identification, human-machine interface.

1. INTRODUCTION

In the beginning, the robots started replacing people by performing repetitive jobs or dangerous tasks for a human operator. It was about robots that were mainly designed to work in industrial, completely structured, environments. For this reason, limited human-robot interaction was required, if any. In the recent years, however, we witnessed a dramatic shift in the robotics research efforts. Nowadays, mobile robots are increasingly becoming more integrated socially. Unlike industrial robots that operate in highly structured environments and respond to a fixed set of commands without sensorial feedback from the environment, mobile robots operating in human-populated environments must present more flexibility to changeable conditions and must be prepared for a wider range of tasks (Adams et al., 2000). One day, they will become our personal assistants, for entertainment purposes (AIBO; Papero; and Krose et al., 2003) or social ones, like taking care of elderly persons (Montemerlo et al., 2002). Or, they will populate public spaces like museums, art-galleries (Arras et al., 2003; Thrun et al., 1999) offering guided tours.

Unlike their industrial counterparts, which must explicitly be turned on or off in order to accomplish their mission, the social-oriented mobile robots must show some kind of awareness, that will allow them to be available for continuous interaction with people. Social robots are usually provided with the same perceiving capabilities than people: audio, vision, etc.

In this paper we introduce a mobile robot (called “MiReLa”) oriented towards interaction with people. The research involving MiReLa (which is a B21r mobile robot) takes place at the Technology Park of San Sebastian (TPSS), Spain, and its main task will consists in acting as a “concierge robot” for the main building of the park. It patrols continuously the
entrance of the building looking for visitors, tries to approach him/her once detected and offers help in finding a particular location or person in the building.

The paper is structured as follows: section 2 is dedicated to the description of navigation and obstacle avoidance strategy. Section 3 presents the vision algorithm used for people detection. In section 4, a possible scenario for a demo application is proposed, while section 5 contains our conclusions and guidelines for future work.

2. NAVIGATION AND OBSTACLE AVOIDANCE

Although humanoid robots could seem more appropriate to perform social interaction with humans, the degree of complexity involved in the control of movement poses a challenge itself. The agent used in the development of this work is a B21r robot from Real World Interfaces Inc. (now iRobot) (see figure 1).

![Fig.1 A B21r mobile robot from the iRobot company, similar to “MiReLa” used in the experiments](image)

Thanks to a large extent to its size, this kind of robot is suitable to build social interaction with humans. The set of sensors and driving mechanism are also appropriate to semi-structured environments, like the one in which the robot will perform its duties. The sensing capabilities of this B21r include a CCD camera, laser scanner, sonar and infrared sensors rings for proximity detection with different detection ranges, an electronic compass that provides a global orientation sense, and finally tactile sensors to detect contact with objects. The movement of the robot is performed by four wheels in a synchro-drive configuration hence it is holonomic and can turn “on-the-spot”. This kind of steering is highly desirable in this agent-environment system because any manoeuvre becomes much simpler. On the other hand synchro-drive steering configuration reduces the odometric errors.

Usually to avoid obstacles, some kind of local space representation is performed, mainly, because of the features of the sonar sensors. The representations’ main purpose is to maintain some kind of memory on the sensors readings, for example to overcome the “so called” sonar sensors’ weakness. On the other hand to make a proper sensor fusion the values must represent distances to the objects. While building the obstacle avoidance competence some mechanism uses to be anticipated for later generation of navigation. This generates the known and widespread hybrid architectures, with fast low-level reactive control loops and higher deliberative modules making extensive use of world models generated through sensor fusion.

The main alternative to this hybrid paradigm is known as the behaviour-based approach. Following this robotics trend, the complex behaviour displayed by the real robots emerges from the interaction with the environment, without the need of building world models. The control system of this robot has been developed with these ideas in mind in an incremental way. The first developed set of modules provides the robot with the obstacle avoidance behaviour. The second set is represented by the navigation behaviour. Each of these behaviours is constituted from modules. The output of each of these modules is further combined in a “Weight and Sum” module, whose output represents the linear ($V$) and angular ($\Omega$) velocities, respectively. The overall architecture is represented in figure 2.

![Fig. 2. The overall architecture of the behaviour-based control system, where the obstacle avoidance and navigation competences have been depicted. The output of this architecture, $V$ and $\Omega$, represents the linear and angular velocity, respectively.](image)

A more detailed explanation of these two behaviours (together with their constitutive modules) is offered in the following subsections.
2.1. Obstacle Avoidance Behaviour

The obstacle avoidance behaviour is achieved in a simple way, like the “aggression” behaviour proposed in (Braitenberg, 1986), but using as stimulus the detected free space. For each proximity sensorial system (sonar, infrared and laser) a module is implemented to perform obstacle avoidance. Indeed, three different sets of commands (linear speed and turn rate) are generated independently from these sensors to avoid obstacles. Since the ultrasonic waves suffer from specular reflections on the wide walls of the robot’s environment, the obtained measures are low-pass filtered. Moreover, as the speed of the robot is always non-negative only the front sensors from the sonar and infrared rings are used to compute the modules’ outputs.

Since this paradigm doesn’t build any world model there is no sensor fusion, instead there is the so-called “motor fusion”, with two main trends in the way this task is performed; cooperative or competitive motor fusion (Arkin, 1998). The selected approach has been the cooperative one because the resulting behaviour is smooth, while using a competitive combination the movement of the robot is too variable. Therefore the resulting motion command is a weighted combination of the speeds provided by each obstacle avoidance module, the sonar, infrared and laser ones.

2.2. Navigation behaviour

Although it is well known that odometry sensors show increasing errors for distances higher than ten meters or so, for short distances they are quite precise. It is also known that the main problem on integration performed by these sensors comes from the absolute world heading which accumulates unbounded errors. Even though odometry is extensively used by animals to navigate, the heading usually comes from an external reference, the turning movements being not integrated. In this case, the navigation behaviour in the environment is a combination of the obstacle avoidance behaviour and two other modules that drive the robot a given distance in a magnetic heading, provided by its compass. As in the case of animal navigation, the used magnetic orientation is external, without error accumulation.

The first additional module for navigation keeps the robot moving in the direction of a target place, with a configurable magnetic heading as reference. The linear speed has a non-negative value while the turn rate provided by this module is proportional to the difference between the current magnetic heading and the target orientation parameter. When this module controls the robot in isolation with a zero linear speed the resulting behaviour is a turn towards some assigned orientation. Jointly with the avoidance modules and using a non-null speed, the robot wanders with a preferred orientation.

The other module needed to generate the navigation behaviour, using magnetic orientation and small odometric displacements, integrates both, the relative distance travelled from a pre-defined starting point in the target direction (the same as for the compass module) and the perpendicular distance, the deviation from the nominal path. Two parameters control the displayed behaviour of this module, the target heading and the target distance. The provided output speed is constant while the target distance has not been travelled and zero once the target reached. On the other hand, the turning velocity is proportional to the lateral integration, the shift from the nominal path. Although this module doesn’t control the robot heading (it can boundless travel in the opposite target direction), it can be shown that a weighted combination of the compass module and this one produces an emergent behaviour that brings the robot to the target. Though the compass heading is not accurate due to the magnetic interferences of the environment, it has been experimentally verified that for long distances it is appropriated, since the errors, or deviations due to the magnetic field tend to compensate for long travels. In a homing travel of sixty metres the error can be smaller than one metre.

Finally, to have a real navigation behaviour, a special module has been implemented to sequentially configure all the parameters (magnetic orientation and relative displacements) on the previous modules. A graph containing magnetic heading and relative distance information is defined and used by this special module to generate a proper navigation behaviour through the environment, depending on the final target. The navigation is performed by serialization of turns and fixed size displacements on given magnetic directions from the starting position. The robot can return to its home after any sequence of movements by finding the appropriated connections in this graph of transitions. To solve the small shifts in long travels, two multi-layer perceptron artificial neural networks have been trained using a standard back-propagation algorithm to fix the position of the robot at the starting place. The network inputs were the compass heading and the laser scanner sweep, and the trained output was some approximated desired linear and turn speeds near the home place. To train the neural networks, readings from the sensors were collected and stored near and at the robot starting position, but due to the fact that the robot’s home was near some glass surfaces not properly detected with the laser scanner, the network did not performed exactly, but the generated behaviour was appropriate.

3. PEOPLE DETECTION

A crucial task for a social-oriented mobile robot is to detect human presence. In (Koku et al., 2000), human detection and tracking based on sound source localization is performed, while in (Schlegel et al., 1998), computer vision techniques have been used for the same purpose.
When addressing the problem of person detection from the computer vision perspective, the most natural way to look for people presence is to try to perform identification of face-like patterns. Despite of the fact that face detection is an old problem in the field of computer vision, which concentrated a lot of the research effort in the last years, it cannot be said that a universal solution has been found yet. Among the factors that make face detection a very difficult problem are: changes in illumination conditions, changes in face appearance (due to the presence of natural or artificial make-ups), changes in pose, etc.

As an alternative, it is also possible to perform person detection by the localization of a special “tag” (badge) attached to the user. By measuring whether the distance and orientation of the badge to the camera is below a critical threshold we wish to assess if the robot becomes the user’s focus of attention, in order to initiate the interaction. This approach was successfully exploited in (de Ipiña et al., 2002), but a main drawback (insufficiently explained in the paper) was the lack of robustness against changes in illumination conditions.

The badge is a rectangle made of rigid paper and has attached patches of reflector tape (in shape of discs), on each of its four corners. Both the modified camera and the badge are illustrated in figure 3. A very important property of this material is that it reflects the infrared light back, on the same direction it came from the source (infrared LEDs). The size of the badge is 10x15 centimetres and the discs have a diameter of 2 centimetres.

Usually, in order to estimate the 3D coordinates of a point in space, a stereovision system is needed. Thus, the 3D coordinates are estimated based on the pixel disparity, i.e. the difference in object pixels’ location in the pair of images captured by the two cameras. But, in our case we have been restricted by the fact that the robot has a monocular vision system.

In this situation, we explore a technique that allows us to estimate the 3D coordinates based on the information about the points and lines whose perspective projection we observe. Such relations with the perspective geometry constraints can often provide enough information to uniquely determine the 3D coordinates of the object. Such knowledge can come about when we have a model of the object being viewed in the perspective projection. The technique to inference the 3D point coordinates when knowing its 2D coordinates on the image plane of the camera is called “inverse perspective projection”. In our case, the method we propose is adapted from (Haralick and Shapiro, 1993) and allows the 3D reconstruction based on the observed perspective projection of two parallel line segments (the lines connecting the centres of the two patches situated along the vertical edge of the badge, for instance).

The experiments performed were intended to assess the accuracy of the distance measured by the proposed algorithm. In the case of infrared technology, the only factor that can affect the performances of the algorithm described above is the amount of infrared radiation presented in the environment. Empirical experiments demonstrated the robustness of our system in case of diffused natural light, considered during different moments of the day, and also artificial light. In our view, these are the most likely scenarios in an office or home environment. We found that only direct sunlight, presented in the scene covered by the camera, can affect system’s performance.

Since the maximum sensitivity of the CCD cameras is in the infrared part of the light spectra, in our approach the camera is enhanced with additional infrared devices. This consists of an array of infrared LEDs placed around the objective in form of a ring and an infrared filter. The role of the filter is to let pass only those frequencies of the light that are close to the infrared rays spectrum. By using this kind of filter, the camera will perceive mostly the light that is reflected from these reflector patches. In consequence, background information is discarded from the beginning, making the image analysis process much simpler highly suitable for this kind of applications.

![Infrared Enhanced Video Camera and Badge](image.png)
We estimated the accuracy of the distance measured when the target was positioned at orientations of 0, 30 and 45 degrees with respect to the camera. The distance range was set between 40 centimetres and 2 meters. The error in distance estimation, at 2 meters (which is the maximum distance perceived by the system in its current configuration), was about 4%. The results are presented in figure 4.

![Graph](image1)

![Graph](image2)

![Graph](image3)

Fig. 4. The distance accuracy estimated over 350 frames. Graphics (a), (b) and (c) correspond when the pattern is successively positioned at 0, 30 and 45 degrees with respect to the camera.

Besides the absolute error (in cm) calculated in reference to the measured distance, we also estimated the standard deviation of the data collected for each distance. Due to sensor inaccuracy or other errors, the initial results present a large variation. A Kalman filter (Welch and Bishop, 2002) was applied in order to smoothen the variable calculated (distance) and to reduce its variance. Figure 5 shows the standard deviation of our data before and after the application of the filter. The improvements introduced by the application of Kalman filter are obvious.

![Graph](image4)

Fig. 5. Standard deviation for the distance measured, before and after the filtering.

Following the behaviour-based approach described in section 2, the proposed algorithm for person detection will be soon implemented as another competence, to be combined with already existing two behaviours.

4. A SIMPLE DEMO APPLICATION

The proposed scenario for the demo application is the following. In the “stand-by” mode, the robot is patrolling the entrance of the building near its starting position. When it detects a visitor, it tries to approach him/her by following the direction and distance obtained from the people detection module. The interaction between robot and the user would take place through a visual interface displayed on the screen. Besides the greeting from arriving at the TPSS, the robot could also offer a set of possible places to go, for example through a web browser on the screen. Once a place is selected the robot starts the corresponding sequence of behaviours, turns and relative displacements while avoiding obstacles, to reach that place. The user must follow the robot. Once the person has been driven into the selected place, the robot roughly approaches its home following the proper sequence of behaviours and uses the neural networks to control the final movement until it reaches the right home point from the surroundings. The whole process is repeated when a new visitor is detected.

5. CONCLUSIONS AND FUTURE WORK

In the last few years, we have noticed an increasing research effort towards social-oriented robots. Having as departure point this reality, we presented in this paper an application of a “concierge robot” able to interact with humans and to offer assistance and help in order to achieve their objectives.

The results obtained are promising, although a fault tolerant system is needed to have a really reliable
“concierge robot” that can perform such tasks without any supervision. The commercially available robotic platforms are still far from supporting real applications.

The robot makes use of a vision system in order to notice the presence of people who visit the building. The presence is acknowledged by the detection of a special badge only carried by visitors. Due to the infrared technology used to enhance the video-camera functionality, we achieved a high robustness against changes in illumination conditions, a phenomenon inherent in normal, semi-structured environments.

In this moment, the interaction is simple, highly effective but not too natural for humans, the user has to communicate explicitly with the robot through the display. In the future the communication mechanisms could be raised to support near-human communication, for example by integrating voice and gesture recognition capabilities.

ACKNOWLEDGEMENTS

The authors would like to thank to San Sebastian Technological Park and Computer Vision Center Barcelona for partially supporting this work.

REFERENCES


