People Detection and Tracking in a Receptionist Robot[1]

Ana Cristina Aleixo, Manuel Malhado, Rodrigo Ventura, and Pedro Lima
Institute for Systems and Robotics
Instituto Superior Técnico
Av. Rovisco Pais, 1
1049-001 Lisboa, Portugal
{cristinaleixo, manuel.malhado}@gmail.com, {yoda, pal}@isr.ist.utl.pt

Abstract

This paper presents a receptionist robot that is currently being developed at the Institute for Systems and Robotics (ISR) - Lisbon, focusing the presentation on the people detection and tracking algorithm. The followed methodology focuses on the integration of several modules, such as navigation, human-robot interaction, and people detection, in order to achieve an autonomous system. In order to save time and effort, as well as obtaining a robust solution, "off-the-shelf" software packages are being used whenever possible. However, for the people detection and tracking algorithms, it was necessary to develop a solution from scratch, using the robot's omnidirectional vision system.

1. Introduction

The goal of the receptionist robot is to approach visitors that arrive at ISR's 7th floor, interact with them, and to be able to guide them to any requested room on that floor. The robot will be stationed at a pre-defined place (*e.g.*, at ISR's lobby, near a power outlet) where it will wait for a person to approach it. At this point, it will attempt to establish oral communication in order to find out whether the person wishes to be lead to a specific room. If the person accepts the invitation, the robot guides him/her to the chosen destination, while making sure it is being followed. Upon reaching that room, the robot indicates the room, and returns to its starting position.

2. The robot platform

The receptionist robot is being built on top of a modified Nomadic SuperScout II, a commercial unicycle robot. The robot has two CPUs: a FlyBook V33i laptop (with a touch screen, microphone, and speakers) running Windows XP, and a single board computer with a Pentium 3 CPU running Linux.

This robot is equipped with a range of sensors particularly useful in this context: a Hokuyo-URG laser range finder, odometry, a sonar ring, a webcam (this camera will be facing backwards, like the touch screen), and an omnidirectional catadioptric system (whose camera provides a "below the waist-line" 360º image of the robot's surrounding area) [1].

3. Description of the subsystems

For the robot to perform the functionalities described in Section 1, the following subsystems are being developed:

- **Coordination.** The top-level coordination is implemented by a finite state machine, specifying what the robot should do next, based on the sequence of events. The communication among modules is provided by YARP (Yet Another Robot Platform) [2], an open-source multi-platform IPC toolkit.
- **Navigation and Localization.** For this subsystem, the Carnagie Mellon navigation toolkit (CARMEN) [3] is being used. This software package performs autonomous navigation based on laser and odometry, using a previously generated map.
- **Speech recognition.** For this module, the Microsoft's well known Speech Application Programming Interface (SAPI) is being used, along with a context dependent lexicon to improve recognition rate (*e.g.*, depending on whether the robot is expecting a yes/no answer, a room number, and so on).
- **On-screen interface.** The user can interact with the robot using a touch screen displaying a human-like face, runtime information about the robot (*e.g.*, a map), and GUI elements providing an alternative interface to speech. This module is being implemented using the Xface toolkit[2], which also incorporates SAPI for speech synthesis, and wxWidgets[3] for GUI components.

[1] This work was supported by FCT (ISR/IST plurianual) through the POS_Conhecimento Program that includes FEDER funds.

[2] URL: http://xface.itc.it/

[3] URL: http://www.wxwidgets.org/

- **Face detection.** This module is used to establish eye contact with the person, thus achieving a more personal human-robot interaction. Moreover, establishing eye contact provides an alternative way of detecting people. For this module, the OpenCV's[4] face detection algorithm is being used.
- **People detection and tracking.** Since we have not found readily available people detection or people tracking software packages, using omni-directional vision, we were forced to develop one from scratch. The next section describes the research being conducted to develop this module.

4. People detection and tracking

4.1. Detection

It is imperative that the receptionist detects the nearby presence of people. Since the robot performs this task while standing still, a motion detection based approach is being adopted, using background subtraction. A threshold is then applied to the resulting image, from which a binary image is obtained. This image is subsequently subject to erosion and dilation morphological operations, in order to eliminate most of the noise while approximately retaining the moving object's shape.

At this point, the binary image is sectioned, isolating the shape closest to the robot (this is accomplished by means of a labelling algorithm), thus focusing on a single person, if more than one is nearby. By working on the obtained section, when compared to the processing of the whole image, the processing time is significantly reduced. By calculating the first and second order moments [4], the shape's dimensions, position and orientation are then estimated.

When analysing the omnidirectional camera's images, we realised that the region corresponding to a standing person resembles a triangle. Based on this observation, and using the previously obtained moments, a triangular mask is created, which is then compared to the binary image, thus deciding if the captured shape corresponds to a person.

4.2. Tracking

In order to make sure that the robot is being followed while guiding a visitor to the requested room, a people tracking module is required. An algorithm similar to the people detection cannot be used while the robot is in motion, since the background is no longer static. Therefore two different approaches to this problem are being considered at this point.

The first approach is to sample the colour of the pixels located inside the triangle previously obtained by the people detection algorithm. The mean and variance of the samples is then used for tracking the person while the robot is in motion.

The second approach consists in tracking the boundary between one of the person's legs and the background. Admitting that this boundary is a straight line, several points are sampled along a side of the triangle initially obtained in people detection. By tracking these points across successive images, the linear conformal transformation (translation, rotation and/or scaling) is then estimated. This transformation explains the points' transition from the previous image to the current one. The obtained transform provides an estimate of the line's position and orientation in the current image, from which a new set of points can be considered. In order to find the points in an image, the pixels' value will be sampled along small straight lines, orthogonal to the previous boundary, at the points already considered. The point's new position is determined by maximization of the correlation between the values along each line and a step function template.

5. Current status

The implementation of the people detection algorithm is on its final stage. We are presently tuning its parameters, as well as testing its robustness and efficiency. We are also currently developing one of the people tracking approach outlined in section 4.2.

6. References

[1] P. Lima, A. Bonarini, C. Machado, F. Marchese, C. Marques, F. Ribeiro, and D. Sorrenti, "Omni-Directional Catadioptric Vision for Soccer Robots", *Special issues on the EuroRoboCup of the Journal of Robotics and Autonomous Systems*, Elsevier, Amsterdam, Vol 36, no 2, 2001.

[2] G. Metta, P. Fitzpatrick, and L. Natale, "YARP: Yet Another Robot Platform", *Journal of Advanced Robotics Systems, special issue on Software Development and Integration in Robotics, Volume 3(1)*, 2006.

[3] M. Montemerlo, N. Roy, and S. Thrun, "Perspectives on Standardization in Mobile Robot Programming: The Carnegie Mellon Navigation (CARMEN) Toolkit", *Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3*, October 2003, pp. 2436-2441.

[4] R.O. Duda and E.Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, USA, 1973.

[4] URL: http://www.intel.com/technology/computing/opencv/