Robotics and Autonomous Systems 60 (2012) 1021–1036

Contents lists available at SciVerse ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot[☆]

V. Alvarez-Santos[a], X.M. Pardo[a], R. Iglesias[a,*], A. Canedo-Rodriguez[a], C.V. Regueiro[b]

[a] Department of Electronics and Computer Science, University of Santiago de Compostela, E-15782, Spain

[b] Department of Electronic and Systems, University of A Coruña, E-15071, Spain

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form
30 March 2012
Accepted 11 May 2012
Available online 28 May 2012

Keywords:
Image features
Feature selection
Human recognition
Person following
Robot vision

ABSTRACT

One of the most important abilities that personal robots need when interacting with humans is the ability to discriminate amongst them. In this paper, we carry out an in-depth study of the possibilities of a colour camera placed on top of a robot to discriminate between humans, and thus get a reliable person-following behaviour on the robot. In particular we have reviewed and analysed the possibility of using the most popular colour and texture features used in object and texture recognition, to identify and model the target (person being followed). Nevertheless, the real-time restrictions make necessary the selection of a reduced subset of these features to reduce the computational burden. This subset of features was selected after carrying out a redundancy analysis, and considering how these features perform when discriminating amongst similar human torsos. Finally, we also describe several scoring functions able to dynamically adjust the relevance of each feature considering the particular conditions of the environment where the robot moves, together with the characteristics of the clothes worn by the persons that are in the scene. The results of this in-depth study have been implemented in a novel and adaptive system (described in this paper), which is able to discriminate between humans to get reliable person-following behaviours in a mobile robot. The performance of our proposal is clearly shown through a set of experimental results obtained with a real robot working in real and difficult scenarios.

1. Introduction

In the next few years, personal service robots are expected to become part of our everyday life, playing an important role as our appliances, servants and assistants; they will be our helpers and elder-care companions. These robots will need to be able to acquire a sufficient understanding of the environment, be aware of different situations, or detect and track people with minimum instruction and with high quality and precision. This will allow the achievement of natural human–robot interaction, and it will also allow the robot to focus its attention on one individual. Dautenhahn [1] summarised the importance of recognising a human with the following expression: 'Humans are individuals, and they want to be treated as such'. Thus, we believe that many robots could enhance their abilities by including human recognition and discrimination in the tasks that they are already able to carry out. Robots should be able to discriminate between humans or in general amongst similar objects even when they move through cluttered areas, there are objects overlapping in the visual field, shadows, illumination changes, objects being introduced or removed from the scene, etc. Hence, robots need robust human recognition and tracking that can account for such a wide range of effects.

As an example of service robot applications that require such person recognition, we can mention the entertainment robot SDR-4X [2], the museum or exhibition guide robot TOURBOT [3], a shopping mall guide robot by Kanda et al. [4], and also robots designed to care for the elderly such as the well-known Care-O-Bot [5]. In all the aforementioned examples the robot needs a robust and flexible human discrimination ability to distinguish the person which it is interacting with (target), avoiding mistaking this target for the rest of the people present in the same scene (distractors).

Our goal is to create a robust and adaptive system for human recognition able to discriminate our target from the rest of the people. Our system must be flexible enough to handle important variations in illumination, scene clutter, multiple moving objects, and other arbitrary changes to the observed scene. On the other hand, the person being recognised and followed will not need to

☆ This work was funded by the research grants TIN2009-07737 and INCITE08PXIB262202PR.

* Corresponding author.
E-mail addresses: victor.alvarez@usc.es (V. Alvarez-Santos), xose.pardo@usc.es (X.M. Pardo), roberto.iglesias@usc.es, roberto.iglesias.rodriguez@usc.es (R. Iglesias), adrian.canedo@usc.es (A. Canedo-Rodriguez), cvazquez@udc.es (C.V. Regueiro).

0921-8890/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2012.05.014

wear special clothes or location gadgets, thus achieving a more natural human–robot interaction. We intend to use the system described in this paper in a general-purpose guide robot which can be deployed at different museums or events where it will have to show routes of interest to the visitors. The robot will learn the routes that later it will have to show to the visitors, by following a person who will demonstrate them to the robot. We will refer to this person as a demonstrator or instructor (we will use both words indistinctly). This person will probably be a staff member of the museum or the event, without any kind of expertise in robotics. Due to this, we will also need a person-following behaviour on the robot, so that the robot is able to recognise a target that must be followed or a visitor who wants to follow the robot across the event. If our robot mistakes a distractor for the target during the route learning process, it will probably need to start over in learning the route.

In this paper we describe our proposal of a person-following behaviour which enhances the role of a camera as a full-body human discriminator. We start by giving a general description of the design of our multimodal person-following behaviour. We highlight the major enhancements performed on the camera, and we give details on how our robot merges the information from the different sensors. After this introduction to the behaviour, we get into details on how our discrimination works. First, we review the most important colour and texture features (extracted from the human's torso region) which can be used for human recognition in a robot. In this part of the work, our goal is to get a subset of these features which achieve a good discrimination rate, but with a low computational cost. Then, we describe how we compare these features, and investigate mechanisms which provide our proposal with important adaptability, necessary to work on crowded and changing environments. Basically we pursue a strategy able to dynamically select those features that allow the best recognition of the person being followed at each instant. Finally, in the experimental results we show the performance of our proposal when it is applied on a real robot working on real and challenging environments.

2. Related work

In the past few years, there have been many works which have dealt with the problem of object tracking [6,7]. However, in most cases the system is restricted to rigid objects, such as cars, or semi-rigid such as faces or humans with limited variations in their pose. These works often describe strategies that have been developed for conventional cameras located in a fixed position and far from the people or objects being tracked. In this paper we want to present a new alternative to achieve a system that is really able to follow a target in crowded scenarios under strict real-time conditions, and strong changes of illumination. Moreover our system will have to track humans, whose movement is completely unfettered, from a camera located on a non-stationary robot. This results in a highly evolving environment. According to these needs, we can broadly classify the related work into two main categories. The first category is *human tracking from mobile robots*. The second category is *human recognition* (also called human identification in some works).

Most of the human trackers in robotics use particle filters, multiple hypothesis tracking (MHT) and Kalman filters (in any of its versions) to estimate the current position of people, starting from their last known positions and the current sensor information (Treptow et al. [8] use thermal images to locate and track the position of people, Mucientes and Burgard [9] and Kai et al. [10] use the information provided by a laser scanner to detect people's legs, [11,12] combine leg and face detection to obtain the position and track a person, [13] emphasises the necessity of multi-sensor fusion and probabilistic aggregation instead of single sensor tracking systems, etc.). Most of these human trackers will probably miss the person being followed if they do not see this person for a short period of time. This represents an important handicap for us, since in our case it is often that the robot loses sight of the person being followed for short periods of time due to the specific characteristics of the environment (when the robot is going round a corner, for example), or due to important occlusions in crowded areas.

Regarding the second category of the aforementioned related research, *human recognition*, there are few mobile robots that are able to actually recognise a person who does not wear either special clothes or specific gadgets. Most works carry out face recognition [14,15], or speaker direction detection [16,17] to recognise a human. Nevertheless this requires the person to stand close to the robot, and to look at it. However, other works [18–20] avoid this restriction by recognising humans using their clothes.

In this paper we want to exploit the visual features to solve the human recognition problem. Because of this, and as we will see in section Section 4, we will analyse the possibility of using several state-of-the-art feature-based recognition approaches [18–28]. Contrary to what happens in other papers, we will not detect the human's face to detect the presence of a person and extract its features, but instead we will apply a strategy similar to the one developed by Dalal [26] to detect humans in the images. Finally, most of the works that address the problem of human recognition do not consider enhancing the discrimination between them. In our case we will try to find a general adaptive strategy of a combination of visual features that allow the discrimination of people in different environments and under changing illumination conditions. To do so, we will take into account the strategies already developed by Kuo et al. [28] and Collins et al. [29], and which deal with the problem of enhancing the differences amongst two objects either by means of building models that are later recognised using AdaBoost, or by means of selecting the most discriminative visual features according to their distribution in the objects that are to be distinguished.

3. The person-following behaviour

An overview of the system we have developed to solve the person-following task can be seen in Fig. 1. The robot will use information provided by a camera and a laser scanner (Fig. 2) to obtain the position of a target. This information will be sent to a motion controller which will have to determine the motor commands that the robot must carry out so that it follows the target and avoids colliding with the environment.

Therefore, in our system there are two clearly separated modules which work in parallel (Fig. 1): *the camera module* and the *laser module*. *The camera module* will detect the presence of people around the robot using Dalal's human detector [26], and it will recognise the target from its torso, using both colour and texture/edge features. This camera module will provide the angle at which the target is located with respect to the forward direction of the robot. *The laser module* will obtain the position of the humans near the robot through a leg detection process, and will merge this information with the information provided by the camera module, to track the target.

A novel aspect in the system we propose is the sensor fusion process, which is carried out at the *tracker* (Fig. 1). This sensor fusion merges the outcome of the human recognition carried out at the camera module, with the human positions obtained from the leg detection carried out at the laser module.