Robotics and Autonomous Systems 58 (2010) 435–443

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Data association and occlusion handling for vision-based people tracking by mobile robots

Grzegorz Cielniak [a,*], Tom Duckett [a], Achim J. Lilienthal [b]

[a] School of Computer Science, University of Lincoln, LN6 7TS Lincoln, United Kingdom
[b] Centre for Applied Autonomous Sensor Systems, Örebro University, SE-701 82 Örebro, Sweden

ARTICLE INFO

Article history:
Received 1 May 2009
Received in revised form
4 February 2010
Accepted 11 February 2010
Available online 17 February 2010

Keywords:
AdaBoost
Occlusion detection
Thermal vision
Colour vision
Bayesian estimation

ABSTRACT

This paper presents an approach for tracking multiple persons on a mobile robot with a combination of colour and thermal vision sensors, using several new techniques. First, an adaptive colour model is incorporated into the measurement model of the tracker. Second, a new approach for detecting occlusions is introduced, using a machine learning classifier for pairwise comparison of persons (classifying which one is in front of the other). Third, explicit occlusion handling is incorporated into the tracker. The paper presents a comprehensive, quantitative evaluation of the whole system and its different components using several real world data sets.

1. Introduction

This paper addresses the problem of people detection and tracking by mobile robots in indoor environments. A system that can detect and recognise people is an essential part of any mobile robot that is designed to operate in populated environments. Information about the presence and location of persons in the robot's surroundings is necessary to enable interaction with the human operator, and also for ensuring the safety of people near the robot.

The presented people tracking system uses a combination of thermal and colour information to robustly track persons. The thermal camera simplifies the detection problem, which is especially difficult on a mobile platform. The system is based on a fast and efficient sample-based tracking method that enables tracking of people in real-time. The measurement model using gradient information from the thermal image is fast to calculate and allows detection and tracking of persons under different views. An explicit model of the human silhouette effectively distinguishes persons from other objects in the scene. Moreover the process of detection and localisation is performed simultaneously so that measurements are incorporated directly into the tracking framework without thresholding of observations. With this approach persons can be detected independently from current light conditions and in situations where other popular detection methods based on skin colour would fail.

A very challenging situation for a tracking system occurs when multiple persons are present on the scene. The tracking system has to estimate the number and position of all persons in the vicinity of the robot. Tracking of multiple persons in the presented system is realised by an efficient algorithm that mitigates the problems of combinatorial explosion common to other known algorithms. A sequential detector initialises an independent tracking filter for each new person appearing in the image, using thermal information. A single filter is automatically deleted when it stops tracking a person.

While thermal vision is good for detecting people, it can be very difficult to maintain the correct association between different observations and persons, especially where they occlude one another, due to the unpredictable appearance and social behaviour of humans. To address these problems the presented tracking system uses additional information from the colour camera, introducing several techniques for improving data association and occlusion handling.

First, an adaptive colour model is incorporated into the measurement model of the tracker to improve data association. For this purpose an efficient integral image based method is used to maintain the real-time performance of the tracker.

Second, to deal with occlusions the system uses an explicit method that first detects situations where people occlude each other. This is realised by a new approach, based on a machine

* Corresponding author. Tel.: +44 1522 837398; fax: +44 1522 886974.
E-mail addresses: gcielniak@lincoln.ac.uk (G. Cielniak), tduckett@lincoln.ac.uk (T. Duckett), achim.lilienthal@tech.oru.se (A.J. Lilienthal).

0921-8890/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2010.02.004

learning classifier for pairwise comparison of persons, that uses both thermal and colour features provided by the tracker. Our approach uses the AdaBoost algorithm [1] to build the classifier from the available thermal and colour features.

Third, the information from the occlusion detector is incorporated into the tracker for occlusion handling and to resolve situations where persons reappear in a scene.

Further to our previously published results [2], this paper presents a comprehensive, quantitative evaluation of the whole system and its different components using several real world data sets recorded in an office environment (see also [3] for further details). We analyse the relative influence of different visual features for occlusion handling, and further demonstrate the robustness and efficiency of the approach.

1.1. Related work

Many approaches for people tracking on mobile platforms are based on skin colour and face recognition (e.g., [4,5]). However these methods require persons to be close to, and facing, the robot so that their hands or faces are visible. Stereo vision provides extra range information that makes the segmentation of persons easier, allowing for detection and tracking of both standing and moving people regardless their orientation [6,7]. In both these systems, the coarse depth information provided by the stereo-camera has proven sufficient to resolve the majority of short-term occlusions.

Our system makes use of thermal vision, taking advantage of the fact that humans have a distinctive thermal profile compared to nonliving objects. Moreover, thermal information is not influenced by changing lighting conditions and allows detection of people even in darkness. Infrared sensors have been applied to detect pedestrians in driving assistance systems (e.g. [8,9]), but their use in robotic applications is limited, probably due to the high price of the sensors. So far, thermal cameras have been deployed mostly on mobile platforms designed for search and rescue missions [10,11]. The recent work of [12] describes the use of a thermal sensor for detection and classification of non-heat generating objects used for mobile robot navigation.

Other people tracking systems are based on range-finder sensors such as laser scanners and sonar, which are very popular sensors in mobile robotics for navigation and localisation tasks. The system in [13] uses a laser scanner sensor to track multiple persons. It is based on a particle filter and JPDAF data association, uses a global representation of the environment, requires thresholded sensor data and deals with occlusions of non-interacting persons only. In contrast, our system uses sensor coordinates, incorporates unthresholded data and can reason about occlusions of interacting persons.

Classical tracking algorithms usually handle the detection and tracking tasks separately in order to simplify the whole problem [14,15]. However, such an architecture can cause loss of information between these steps, in addition to the computational cost of detection by an exhaustive search of all possible object states [16]. The alternative approach considers these two problems simultaneously (track-before-detect, also called unified tracking [17]). The presented system is designed in this latter spirit, using a track-before-detect technique.

To deal with problems of occlusions several authors proposed solutions that use special sensors or their special arrangement. One example system uses a camera placed above the observed scene [18]. Persons observed from such a view-point cannot occlude each other. Another example is a multi-camera system [19] where ambiguities caused by occlusion are resolved by combining information from different cameras placed in different places. All these solutions can be used only in a few, controlled scenarios, and their use in mobile applications would be especially troublesome, if not impossible.

In the majority of people tracking systems the problem of occlusion is solved within the tracking framework. Possible approaches handle occlusions either implicitly without reasoning, or model them explicitly. Implicit solutions use kinematic information as well as dedicated measurement models [20–23]. However the behaviour of people tends to be highly unpredictable in general, and they may or may not interact. Therefore implicit approaches can deal only with specific cases, i.e., short-term occlusions. The proposed system uses an explicit approach to deal with occlusions. This reasoning requires domain specific knowledge, i.e., detection of situations when persons appear to merge and split, and making decisions about their behaviour during occlusion (see for example [24–27]). We use colour as additional information that helps to detect occluded persons and resolve occlusions when occluded persons appear again on the scene.

In the next section we introduce the experimental platform. Section 3 presents the basic tracker using gradient information from the thermal camera. The next sections describe the techniques developed to maintain the correct associations between observations and persons, by exploiting a combination of thermal and colour vision: incorporation of colour information into the measurement model (Section 4), an occlusion detector based on the machine learning algorithm AdaBoost (Section 5) and the occlusion handling procedure (Section 6). Experimental results are presented in Section 7, followed by conclusions and suggestions for future work.

2. Experimental set-up

We used an ActivMedia PeopleBot robot (Fig. 1) equipped with different sensors, including a colour pan-tilt-zoom camera (VC-C4R, Canon) and, a thermal camera (Thermal Tracer TS7302, NEC) and an Intel Pentium III processor (850 MHz). The colour and thermal camera are mounted close to each other, which simplifies the calibration procedure between the two cameras (see Section 4.1).

The robot was operated in an indoor environment (a corridor and laboratory room). Persons taking part in the experiments were asked to walk in front of the robot while it performed a corridor following behaviour, or while the robot was stationary. At the same time, image data were collected with a frequency of 15 Hz. The resolution of both thermal and colour images was 320×240 pixels. In our set-up the visible range on the grey-scale thermal image was equivalent to the temperature range from 24 °C to 36 °C.

3. Basic tracker using thermal vision

3.1. Particle-based tracking of a single person

To reliably estimate the location and movement of persons it is necessary to apply a tracking procedure. Our system uses a particle filter to provide an efficient solution to this problem despite the high dimensionality of the state space. The particle filter performs both detection and tracking simultaneously without exhaustive search of the state space. Moreover the measurements are incorporated directly into the tracking framework without any preprocessing, such as thresholding, that could cause loss of information.

The posterior probability $p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t})$ of the system being in state $\boldsymbol{x}_t$ given a history of measurements $\boldsymbol{z}_{1:t}$ is approximated by a set of N weighted samples such that

$$p(\boldsymbol{x}_t|\boldsymbol{z}_{1:t}) \approx \sum_{i=1}^{N} w_i \delta(\boldsymbol{x}_t - \boldsymbol{x}_t^i). \quad (1)$$