Pattern Recognition 43 (2010) 2390–2405

Contents lists available at ScienceDirect

Pattern Recognition

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

Particle filtering with multiple and heterogeneous cameras ☆

Rafael Muñoz-Salinas *, R. Medina-Carnicer, F.J. Madrid-Cuevas, A. Carmona-Poyato

Department of Computing and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain

ARTICLE INFO

Article history:
Received 3 December 2008
Received in revised form
11 December 2009
Accepted 21 January 2010

Keywords:
People tracking
Stereo vision
Particle filters
Multiple-views

ABSTRACT

This work proposes a novel particle filter for tracking multiple people using multiple and heterogeneous cameras, namely monocular and stereo cameras. Our approach is to define confidence models and observation models for each type of camera. Particles are evaluated independently in each camera, and then the data are fused in accordance with the confidence. Confidence models take into account several sources of information. On the one hand, they consider occlusion information from an occlusion map calculated using a depth-ordered particle evaluation. On the other hand, the relative precision of sensors is considered so that the contribution of a sensor in the final data fusion step is proportional to its precision. We have defined confidence and observation models for monocular and stereo cameras and have designed tests to validate our proposal. The experiments show that our method is able to operate with each type individually and in combination. Two other remarkable properties of our method are that it is highly parallelizable and that it does not impose restrictions on the cameras' positions or orientations.

1. Introduction

Multicamera people tracking has become an attractive research field in the last 10 years. The use of multiple cameras allows solutions to the occlusion problem and coverage of wider areas than can be covered by a single camera. Most of the work on multicamera people tracking has been developed using monocular cameras. Nevertheless, recent works have shown that stereo cameras provide better tracking results [26–28] than monocular cameras do. Therefore, people tracking using multiple stereo cameras seems to be a promising tracking alternative. However, the popularization of stereo tracking systems is mainly deterred by the fact that most of the current surveillance systems are already based on monocular cameras. Therefore, the technology transition should be smooth and allow for the reuse of existing hardware.

Particle filtering is now the most popular visual tracking approach [6,13,20,36]. It has gained popularity in the vision community because of its ability to deal with non-linear and non-Gaussian systems. The most common data fusion approach consists of obtaining independent likelihoods from each sensor and multiplying them. However, this approach has several limitations. First, in real scenarios one might need to use a set of heterogeneous sensors, each one with a different level of reliability. However, in such a case, fusing their observations by considering all the sensors as equally reliable is a suboptimal strategy, since it does not make the best use of each sensor. Second, in many cases, the reliability of a sensor might change dynamically according to the target state. For instance, a camera should be considered more reliable if it observes the target perfectly than if it observes the target partially occluded.

This work proposes a novel particle filter for fusing information from a set of heterogeneous sensors, assuming that they might produce measures with different levels of confidence. A sensor confidence model is defined for each type of sensor, indicating its reliability in observing a particle. All observations of a particle are then fused while taking confidence into account, so that sensors with low confidence levels are assigned a lower relevance in the final data fusion step.

The filter is employed to provide a novel solution to the multicamera people tracking problem, which is able to operate with either monocular cameras, stereo cameras or a combination of both types of cameras. For that purpose, we define observations and confidence models for each type of camera. The observation model of a monocular camera evaluates foreground and color information, whereas the stereo camera observation model also evaluates depth information. The confidence models of both cameras take occlusion into account in order to compute their reliability dynamically. Occlusion is computed using a depth-ordered particle evaluation scheme that permits a parallelized implementation of the algorithm.

The rest of the paper is organized as follows. First, Section 2 explains the most relevant related work. Section 3 provides a brief overview of particle filtering and stereo computation. Section 4

☆ This work has been carried out with the support of the Research Projects DPI2006-02608 and TIN2007-66367 funded by the Spanish Ministry of Science and Technology and FEDER.

* Corresponding author.
E-mail address: rmsalinas@uco.es (R. Muñoz-Salinas).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.01.015

presents the proposed particle filtering algorithm, whereas Section 5 explains the observation and confidence models proposed for monocular and stereo cameras. Finally, Section 6 shows the experiments carried out and Section 7 draws some conclusions.

2. Related work

The multicamera people tracking problem has been addressed from multiple perspectives, but most often using monocular cameras. A common approach used by many authors is to intersect the medial axes of the people blobs on the ground plane using the floor homography [2,16,17,19]. The main problem is that the majority of the people's silhouettes need to be visible, and in most cases also the people's feet. Therefore, cameras must be placed at elevated positions and relatively far from the people. Although this restriction may be feasible in outdoor scenarios, it might be difficult in indoor scenarios where the areas to be covered are small and the cameras must be placed closer to the people. A solution to the tracking problem in these scenarios is proposed in Ref. [9] by Fleuret et al. These authors present a tracking approach using multiple monocular cameras placed at eye level. The monitored area is discretized into cells to create a probabilistic occupancy map, and an iterative process is run at each frame in order to determine the locations of the people in the map. The authors claim that the computing time is improved by the use of integral images, but at the expense of imposing restrictions to the camera position and orientation, i.e., the cameras must be placed in such a manner as to prevent people from appearing inclined in the images. In [30], the authors describe a distributed self-configurable tracking system using multiple monocular cameras with and without overlap. In order to allow an efficient camera collaboration, the Kalman–Consensus filter is employed so that each camera comes to a consensus with its neighboring cameras about the actual state of the target. Wul et al. [34] proposed a method specially designed for tracking a large number of close objects from multiple cameras. Their proposal focus on the data association problem for which they propose the use of a greedy randomized adaptive search algorithm.

Although promising results have been obtained using multiple monocular cameras, some studies have shown that stereo vision provides better results for object tracking. The authors of this work have proposed several approaches for people detection and tracking using a single stereo camera. While Ref. [28] proposes a tracking approach combining color and stereo extracted directly from the camera image, Refs. [26,27] propose the use of plan-view maps to represent stereo information more efficiently. In all the tests performed, stereo information has improved the tracking results. However, using a single stereo camera still imposes strong limitations on the extent of the monitored area. In [25], Mittal and Davis present a probabilistic approach for tracking people in cluttered scenes using multiple monocular cameras. They employed a fine camera calibration in order to obtain depth information from camera pairs projected onto plan-view maps. In [21], Krumm et al. show a people tracking system for a smart room using a pair of stereo cameras with a short base line. People are detected by grouping 3D blobs extracted from the stereo information. The main drawback of their approach is the difficulty of reliably segmenting the 3D blobs. In [35], Zhao et al. propose a tracking system with multiple stereo short-baseline cameras. Each camera performs an independent tracking procedure and then reports its results to a centralized tracker.

Important deterrents to the popularization of stereo tracking systems are that most of the surveillance systems installed nowadays employ monocular cameras, and that tracking approaches have not been proposed that can employ the existing equipment in combination with stereo cameras. Our aim is to provide a framework for operating simultaneously with both types of sensors so that current tracking systems could be improved by simply adding stereo cameras.

Particle filters are probably one of the most important tracking frameworks in the vision community. However, the most frequently employed fusion approach consists of considering the observations from all sensors jointly, without considering their relative reliability. In other words, they generally treat all kind of sensors as equally reliable. However, some researchers have proposed alternative fusion approaches with particle filters. In [33] the authors propose an adaptive particle filter for tracking a single object using two monocular cameras. Each particle is assigned two likelihoods, one for each camera. Their fusion strategy consists of using a joint likelihood if the target is visible in both views, and the maximum likelihood in case it is only visible in one camera. The scope of their work is very restricted in the number of views and target that can be employed. In addition, their approach is only valid for tracking in the camera image plane, but cannot be employed to determine the 3D location of the target. In [8], Du and Piater propose a method for tracking people in multiple monocular cameras using a separate particle filter in each camera. At each time step, each filter resamples and evaluates particles using the ground plane homography and the people medial axis. Then, data fusion is performed using a Markov random field, and belief propagation is employed to reduce the computational effort. In Ref. [23] the authors proposed a multi-view-based cooperative tracking approach based on the homographic relation between different monocular views. Their approach is similar to [8] except for the explicit management of occlusion. They apply two hidden Markov processes (a tracking and an occlusion process) for each target in each view. Based on the occlusion process, the cooperative tracking process reallocates resources among the different trackers in each view. These three works are interesting contributions to the multisensor tracking problem using particle filters. Nevertheless, all three assume identical sensors (with the same precision and reliability), thus limiting their extension to more complex scenarios where heterogeneous sensors are available. A solution to that problem is given in [14]. Han et al. propose a kernel-based Bayesian filter in which more particles are assigned to the most reliable sensors. The sensors' reliability is dynamically determined based on the observed likelihood, i.e., sensors providing higher likelihoods are considered more reliable. Finally, independent posteriors are combined as a mixture of Gaussian kernels. One drawback of their approach is that, in practice, their approach does not really fuse information. Instead, the tracking is primarily biased towards the observations of the sensor with the highest likelihood, which attracts most of the particles. Besides, while the sensor with highest likelihood might be the one that observes the target best, it is not necessarily the one with the highest precision. Therefore, their approach does not consider the relative confidence levels of the different sensors employed.

2.1. Proposed contribution

As can be seen, most of the sensor fusion approaches used with particle filters fail to implement the idea of combining information from heterogeneous sensors so as to assign more relevance to the most reliable ones. This work aims to fill that gap by proposing a tracking framework that offers two main contributions. First, we propose a novel multisensor multitarget particle filter that takes into account the sensors' confidence levels. In our