Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online
CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-15 (2016) 175–180

Compact Data Association in Multiple Object Tracking: Pedestrian Tracking on Mobile Vehicle as Case Study

Songlin Piao, Tanittha Sutjaritvorakul, Karsten Berns *

* *Computer Science Department, University of Kaiserslautern, Kaiserslautern, Germany*

Abstract: This paper presents a compact and fast data association which can be used in tracking-by-detection based multiple pedestrian tracking approaches. The goal is to make the data association simple and robust so that it can be really useful in a compact computing environment. This is realized by replacing the computationally heavy Bayesian based filter with the compact Median Flow tracker. Two layers of data association are proposed for fulfilling different requirements of the tracking tasks. Afterwards, we assess the performance of the algorithm by evaluating it on a standard dataset. Experimental results show that it improves processing speed upon state-of-art methods tremendously with only trading off limited performance.

Keywords: pedestrian, multiple, detection, tracking, association, vehicle

1. INTRODUCTION

Pedestrian detection is one of the key points in vehicle safety technologies. Tracking can improve the detection rate, so that the number of pedestrian fatalities and injuries could be reduced. But in a real world scenario, pedestrian tracking from a vehicle is troublesome because of a potentially moving camera and noise, the unpredictable appearance of pedestrians, the cluttered background, the occlusions and abrupt scene changes (Maggio and Cavallaro, 2011). Moreover, it requires real-time computation on a compact CPU based board coupled with on-board FPGA processing in automotive applications.

Tracking-by-detection approaches become widely used to handle those tracking difficulties (Huang et al., 2008; Breitenstein et al., 2009). In general, a detection process is performed on each frame. The tracking algorithm later associates the detection results across the frames based on a matching score. *Data association*, which is the core to many multiple object tracking algorithms, links each detection result that belongs to the same person and assigns an ID to it over time. Thus, a final pedestrian trajectory is generated for each target with a unique ID.

Despite the significant developments on human detection (Dalal and Triggs, 2005; Felzenszwalb et al., 2010), the detection response remains inaccurate. As a consequence, the pedestrian trajectory is incorrectly generated, being subject to misdirected data association between detections and tracked results. Hence an accurate data association algorithm should generate a correct and long-term trajectory. The simplest data association approach is *Nearest Neighborhood (NN)*. Given a previous trajectory, the method chooses the closest measurement to the last state and assigns this measurement as a new state of the trajectory. In order to obtain cues for data association, the association elaborates the matching score definition. Instead of only considering a relative distance of the same target between two consecutive frames, spatio-temporal context can be taken into account.

Another approach is *target estimation*. It provides a trajectory assumption of the target location at the current frame based on the information of the last frame. The implementation of *estimators* is either based on image processing or filtering framework. First, a feature points-based tracking method is a common estimation technique via image processing. It estimates the targets' motion by searching optimized matching positions in the local surroundings. Second, Recursive Bayesian Filtering becomes popular in recent days; besides the usage of classic Kalman Filter (Welch and Bishop, 1995) in the linear system (Mittal et al., 2012), Particle Filter (Arulampalam et al., 2002) is also widely used in the tracking scenarios (Breitenstein et al., 2009; Stalder et al., 2010; Choi et al., 2013). In contrast to the feature-based tracking, the tracking uncertainty can be modeled by a probability distribution based on Markov theory. The concept is to approximate a true state of the system by checking each generated particle. Due to the random movement of human, it is difficult to model the movement using a linear filter. Unfortunately, the computation cost is very expensive for generating a large number of particles. As a result, these methods would be limited only for academic research or a system that contains powerful hardwares.

Considering the aforementioned problems, we propose a compact but robust data association method through integrating the feature-based tracking method in this work. We define three *track states* to clearly represent internal state transitions. Finally, we evaluate our framework on ETHZ Bahnhof dataset (Ess et al., 2008) by comparing detection and tracking performances with other state-of-

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.07.728

art methods. The remainder of this paper is organized as follows. Section 2 gives an overview of the related work; Section 3 explains the proposed framework with respect to system design; Section 4 introduces details of the proposed data association method. The experimental results will be discussed in Section 5, and Section 6 gives the conclusion and future work.

2. RELATED WORK

There are two types of regular requirements for the tracking. One is to improve the system performance by reducing the miss detection rate without losing the accuracy and increasing the whole processing time. Discontinuous problems of detection can be remedied by integrating tracking algorithm. This problem means an object cannot be detected continuously because of brightness changes, noises etc. It can happen even if the change is very small.

The other requirement would be to precisely generate a trajectory for each target. Human-Robot interaction and video analysis are the application scenario for such requirement. For the first type of requirement, ROC curve is used to evaluate the performance; CLEAR MOT Metrics (Leal-Taixé et al., 2015) is used for the second type, and details will be introduced in the experiment section.

The object tracking can be separated further into single and multiple object tracking. Single object tracking refers to track an object based on the predefined object's information such as the appearance template. Besides the short-term tracking algorithms such as Optical Flow (Horn and Schunck, 1981), Meanshift (Comaniciu and Meer, 2002), Continuously Adaptive Meanshift (Bradski, 2000), a lot of long-term model-free tracking algorithms have been published. Babenko et al. (2011) proposed online multiple instance learning (MIL) for robust object tracking. They used one positive bag consisting of several image patches to update a MIL classifier instead of using several positive patches so the drift problem in the traditional tracking-by-detection algorithms was solved. Online random forest and on-line multi-class LPBoost (Saffari et al., 2009, 2010) were proposed to overcome the multi-classification problem in the on-line boosting (Grabner and Bischof, 2006) which could only solve the binary classification tasks. Hare et al. (2011) proposed an on-line learned kernelized structured output support vector machine for adaptive tracking. Kalal et al. (2012) explicitly decomposed a long-term tracking task into tracking, learning and detection parts. Tracking failures could be detected automatically through estimating forward-backward errors (Kalal et al., 2010). Oron et al. (2014) extended classic Lucas-Kanade method (Bouguet, 2000) by using pixel-based object/background likelihoods in the optimization. Although these state-of-art algorithms could show relatively high accuracy in some situations such as static background and indoor environments, the algorithms cannot be directly applied to multiple object tracking in outdoor environments because of occlusions, similarities between the objects and clutter in the background and so on.

Huang et al. (2008) proposed a detection-based three-level hierarchical association for tracking multiple objects in crowded environments robustly. But this algorithm is suitable only for video analysis because future frames are used to estimate the global optimized trajectory. Breitenstein et al. (2009, 2011) proposed a Particle Filter based on-line tracking-by-detection algorithm, in which the confidence of detector was considered in the likelihood model, and object's appearance model was updated regularly using on-line boosting (Grabner and Bischof, 2006). In addition, Choi et al. (2013) proposed a general way of determining tracking trajectories robustly in a 3D coordinate system by estimating both the camera's ego-motion and the people's paths within a single coherent framework. Although these algorithms showed a good performance on the test dataset, it is not realistic to use them in a real time scenario, especially for the ARM-based compact CPU.

3. PROPOSED FRAMEWORK

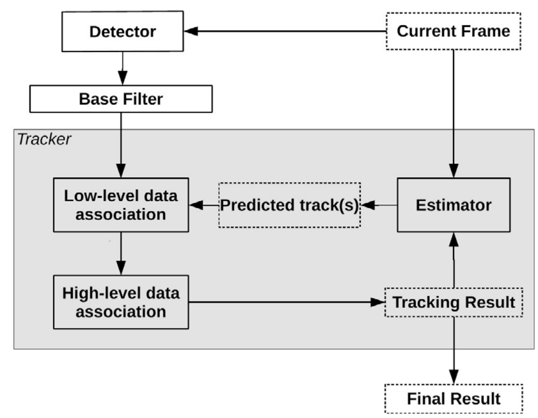

Fig. 1. Structure Overview of the Proposed Framework.

Fig. 1 shows the structure of the proposed framework. The gray rectangle represents the tracking part, in which two levels of data association and an estimator part are integrated. For the estimator, we propose to use the Median Flow tracker, however, we also tested with pure Kalman Filter, mixture of Kalman Filter and Median Flow tracker in the experiments to compare the system performance horizontally. The data association is solved by a linear assignment algorithm (Munkres, 1957). In order to manage the target consistency in a good manner, the framework internally defines three target states: **unstable**, **stable_active** and **stable_inactive**. The **unstable** and **stable_active** trackers will firstly get associated with the incoming detector outputs in the **Low Level**. Later, the **stable_inactive** trackers get associated in the **High Level**. More details about how to calculate a similarity score between a tracker and a detection output can be found in Section 4.2.

The outputs of detector are first fed into a false positive filter. We denote this filter as *Base Filter*. In the experiments, we used a discriminative part model (DPM) to get raw detection outputs (Felzenszwalb et al., 2010). Our *Base Filter* first discards raw detections which are below a certain confidence θ_{det}. Then the *Base Filter* removes any target, of which feet position is above the configured vanishing line θ_v, e.g. red bounding boxes in Fig. 2 will be filtered out. Given the raw detection set $\{z_t^1, z_t^2, ..., z_t^M\}$ at frame I_t, the valid detection set $\{d_t^1, d_t^2, ..., d_t^N\}$ can be defined as

$$d_t^n = z_t^m, \text{ if } (conf(z_t^m) > \theta_{det}) \, and \, (p(z_t^m) < \theta_v) \qquad (1)$$