J. Vis. Commun. Image R. 55 (2018) 711–719

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Cross-camera multi-person tracking by leveraging fast graph mining algorithm ☆

Caiyou Zhang [a,*], Yuteng Huang [a], Zhiqiang Wang [a], Hongcheng Jiang [a], Dongfeng Yan [b]

[a] Information & Telecommunications Company of State Grid Zhejiang Electronic, China
[b] Xiamen Great Power Geo Information Technology Co., Ltd., China

ARTICLE INFO

Article history:
Received 23 June 2018
Revised 6 August 2018
Accepted 7 August 2018
Available online 11 August 2018

Keywords:
Multiple person
Tracking
Video surveillance
Matching

ABSTRACT

Multiple person tracking is a very useful task in intelligent video surveillance, which is hindered by many challenges such as the variations of illumination, the irregular changes of human shapes and the particle occlusions. To tackle these challenges, in this paper, we propose a new online learning tracking system to generate the complete trajectory for each tracking object. In detection stage, we build a classifier for each tracking object by online learning in order to provide more accurate detection results. Online learning could real-time update the classifier for an accurately tracking results in the future. In the tracklet generation stage, we apply the spatiotemporal constrain to generate a set of reliable tracklets. Finally, we propose a new Part-based matching method to get the correlation between different tracklets and apply linear programming and greedy algorithm to handle the data association problem to generate the complete trajectory for each tracking object. In particular, our approach is able to cast the multiple cameras tracking problem as a data association problem. The experiments on our proposed method demonstrate state-of-the-art performance in multiple person tracking.

1. Introduction

Object tracking, especially multiple person tracking, is an popular research point in recent years and it is widely used in many applications such as intelligent video surveillance, gesture and action recognition, and events recognition [23–26]. However, there are a few factors that make this problem challenging and difficult to tackle. For example, the irregular changes of human shapes often leads to variations of the features related to the person to be tracked, the partical occlusions leads to the pseudo-disappearance of tracking objects, and the variations of illumination of scene often leads big changes in features of objects. Multi-view camera will always lead to different shape and appearance of the same object from different angle. So it is still a challenge to detect the same object from different camera. Over the past two decades, a large number of different tracking algorithms [1,2] have been proposed to handle these problems. Traditional trackers such as mean-shift, kalman, particle filter [3–5] can be seen as a process of optimization. They need iterations to find the best results. However, they often fail in some special scenarios and there is no decision-making mechanism to judge whether the tracking objects disappear or are occluded by other objects. In the recent years, more and more tracking algorithms begin to add detected algorithms into tracking system. we call these algorithm as tracking-by-detection. The goal of this method is to improve the accuracy of tracking and cover the shortage of traditional trackers. Shu et al. applied an online learning method to track objects, where they trained a decision model by SVM for each tracking object in order to improve the accuracy of detected results [6]. In the processing, they need to update the decision model. This method of improving detected results by online learning classifier gives us a good deal of enlightenment. Andriluka et al. proposed a tracking-by-detection framework by combining the detector and the tracker [7]. This method relied on the performance of the detector and the stability of features to make data association between different detected results.

With the increasing application of the tracking technology, cross-camera person tracking gradually become a hotspot [8,9]. The main challenge of cross-cameras person tracking is how to judge two persons in the different cameras represent the same person or do not. H. Li combines mean-shift, detected methods, SIFT and new MoSIFT interest point trackers to handle the cross-camera person tracking in a nursing home because author think that different approaches are appropriate for different transitions

☆ This article is part of the Special Issue on TIUSM.

* Corresponding author.
E-mail address: 634423778@qq.com (C. Zhang).

https://doi.org/10.1016/j.jvcir.2018.08.006
1047-3203/

between cameras [10]. This method has a lot of limitations, because we must tell the tracking system about the location information of different cameras, and set the different matching methods for the different transitions between the different cameras. Z. Zhang's method is similar with H. Li's method [11]. The different is that the temporal information is added for matching between the different tracing objects in Z. Zhang's method. Now, many researchers apply traditional tracking algorithm to handle the tracking objects in a single camera, then use data fusion algorithm to implement the cross-camera tracking. In summary, finally, all of these tracking algorithms formulate the tracking problem to the detection problem and the data association problem.

To handle these problems, we propose an online learning tracking system to generate a reliable trajectory for each tracking object. First, we detect the persons in the first video frame by Part-based model. We save the position **(xi, yi)** and feature **fi** (such as color and HoG) for each detected object **i** and initialize the training set in order to learn classifier for each object, which will detail in the section *. Second, in the subsequent frames, we not only apply Part-based model to detect the tracking objects, but also use a set of classifiers to detect each tracking object in motion region got by Optical Flow. Then, the tracking system will observes the detection results by Part-based model and the Classifier to estimates the right detection result by the feature similarity. At the same time, we add these new detection results into the training set and update these classifiers. Third, we assumes that the frame to-frame motion is limited and the object is visible and apply the spatiotemporal constrain and feature similarity between different detection results to generate the reliable tracklets. Fourth, we use the states of initial and termination of tracklet by a fully new Part-based matching method to compute this similarity between two different tracklets. This part will detail in the section *. Finally, in the matching stage, we formulate the matching problem as a Maximum A Posteriors (MAP) problem and use linear programming to handle the occlusion problem and apply greedy algorithm to generate the final trajectory for each tracking object.

The main contributions are third-fold:

1) We propose a new online learning detection method. This method builds individual classifier for each tracking object, corrects the false detection results in current frame and updates the classifier to reduce the errors in the future processing.
2) We propose a fully new matching method. We apply Part-based matching method instead of the whole matching in order to improve the matching accurate and effectively handle the occlusion matching problem.
3) we formulate the matching problem as a Maximum A Posteriors (MAP) problem and use linear programming to handle the occlusion problem. Finally, the greedy algorithm is used to generate the complete trajectory for each tracking object.

The remainder of this paper is organized as follows. Section 2 will introduce the state of the art in visual tracking related to this paper. Section 3 will introduce the tracking system framework. Section 4 elaborates detector + classifier method in detection stage. Data association included Part-matching method will be illustrated in Section 5.

2. Related work

A vast amount of work has been published on multiple person tracking [12–14]. Trackingby-detection is an popular class tracking algorithm in the last decades. It associates the detection candidates with the spatiotemporal constraints to generate a set of reliable tracklets, then applies data association algorithms to get the final trajectory.

In the detection stage, different researchers applied many different methods in order to get more accurately detection results in each video frame. Obviously, Occlusion is a hard challenge in this stage. Background Substraction is used to detect object for a long time because the low calculated amount [15,16]. However, a good calculated amount must sacrifice the performance of the whole tracking system. Background Substraction often does not provide good detection results in detection stage. Shu et al. proposed a new training model to handle the occlusion problem in order to get the higher and better precision and recall in detected results, then applied greedy algorithm to handle the data association problem [6]. This method bringed an obvious improvement in detected results especially in the occlusion condition. However, on the other hand, this method does not provide a stable detected results in the low resolution video because this training model needs the part pictures of body as the training set to learn a classifier for each tracking object, while the low resolution video obviously does not provide a clear part pictures of body. This condition will affect the tracking results. At the same time, the false samples of the training set will bring catastrophic effects in the future process because there is no fault-tolerant mechanism in this method. For handling this problem, we applied detector + individual classifier method to detect person in each video frame aiming to improve the detected accuracy. Yang et al's approach was similar to ours but they used a classifier to predict the potential positions of the tracking person, while we first use Optical Flow to predict motion regions, then apply spatial and motioning information to get the final predicted region. So our approach is more efficient in terms of the computational complexity. We also added decision mechanism to guarantee the correct rate of training set in order to improve the performance of the classifier.

In the tracklet generation stage, several approaches apply spatiotemporal constraints and the similarity of feature to generate a set of reliable such like [2] (Fig. 1). However, these methods often produces so many tracklets and adds the calculated amount in the data association stage. At the same time, the length of tracklet produced by the method often is so short and adds the difficulty of matching because the bigger change of feature or shape of pedestrian will appear in a long time moving. Zdenek et al. proposed a TLD tracking algorithm [1] for a long time tracking, which used an online learning algorithm to learn an classifier for the tracking object, then applied this classifier to track objects. While, in the tracking process, Zdenek et al. does not propose an effective method to handle the partial occlusion problem and the multiple target tracking problem. However, inspired by this work, we propose an individual classifier and a new Part-matching method to handle the partial occlusion problem between different objects in

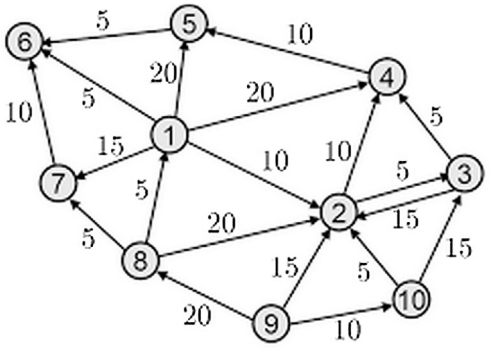

Fig. 1. An illustration of graphs in different cases.