Neurocomputing (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Part-aware trajectories association across non-overlapping uncalibrated cameras

De Cheng, Yihong Gong*, Jinjun Wang, Qiqi Hou, Nanning Zheng

The Institute of Artificial Intelligence and Robotic, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian Ning West Road No. 28, Shaanxi 710049, PR China

ARTICLE INFO

Communicated by Shaoting Zhang

Keywords:
Part-aware
Association
MRF
Group activity

ABSTRACT

This paper focuses on the problem of multi-person tracking across non-overlapping uncalibrated cameras using data association method. The problem is extremely difficult as we have very limited cues to associate persons between cameras. To tackle the problem, our system consists of firstly building multiple trajectories from each camera independently, and then finding associations of trajectories between every two cameras of interest, where the later is the most challenging process. Our contributions are mainly two folds: First, we introduce a method to explore the human part configurations on every trajectory to describe the inter-camera spatial-temporal constraints for trajectories association. Second, we formulate trajectories association across non-overlapping cameras as a multi-class classification problem via the Markov Random Field (MRF) to effectively utilize domain priors such as group activity between persons. With the proposed part-aware correspondences and pair-wise group activity constraints of trajectories, we can achieve robust multi-person tracking. Experimental results on a benchmark dataset validates the effectiveness of our proposed approach.

1. Introduction

Surveillance systems have become almost ubiquitous in large public spaces, even in private palaces. A common topic generating much interests in surveillance is the multi-target tracking. Especially, tracking multiple persons in the camera network have much more essential applications in the real world surveillance systems. Multi-target tracking in a single camera is a classic problem in computer vision and has been extensively investigated [1–4], where data association based approaches are most popular. Tracking multiple targets across camera network is a more common scenario. And the task has great application potentials in areas such as video surveillance, traffic scene analysis [47], and social security, etc. However, it is also very challenging. Nowadays, the cross camera tracking problem is issued by the following ways: 1)transforming the multi-task tracking problem cross camera network into person re-identification problem, which focuses on matching the same individuals by the appearance model [5–8]; 2) tracking multi-target by applying the prior knowledge about the camera placement, geometrical configuration or viewpoint overlapping [9], which needs the calibration of the cameras; 3)data association method extended from multi-target tracking in single camera, which is relatively rarely studied but very useful. In this paper, we focus on the third category and build associations of pedestrians trajectories across non-overlapping uncalibrated cameras.

In addition to those difficulties with multi-target tracking in a single camera (e.g. cluttered environments, object occlusions, intersections, etc), there are many other challenges that are unique to multi-person tracking across non-overlapping uncalibrated cameras. To illustrate how difficult it is to track persons across non-overlapping uncalibrated cameras, Fig. 1 gives a typical scenario across two non-overlapping uncalibrated cameras. There exist *blind regions* where persons can not be observed by any of two adjacent cameras. In addition, the geometrical configuration of the cameras in the networks is unknown, and therefore recovering trajectory correspondences across cameras based on motion cues [10,11] is not applicable. Furthermore, compared with overlapping cameras, geometrical constraints (e.g. ground plane estimation [12,13] used in overlapped camera networks) are not available for multi-target tracking across non-overlapping cameras without calibrations. All of these factors lead to limited cues available for establishing the associations between persons. Although person appearance seems available and consistent, in real-world a person may appear very differently across cameras (e.g. front versus rear views, and left-to-right motions versus bottom-up motions. See Fig. 1), making multi-target tracking across non-overlapping uncalibrated cameras extremely difficult.

To achieve robust association across non-overlapping cameras, in this paper, we resort to human body part-aware correspondences of trajectories to build their associations across every two non-over-

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.11.038
Received 11 July 2016; Received in revised form 9 October 2016; Accepted 18 November 2016

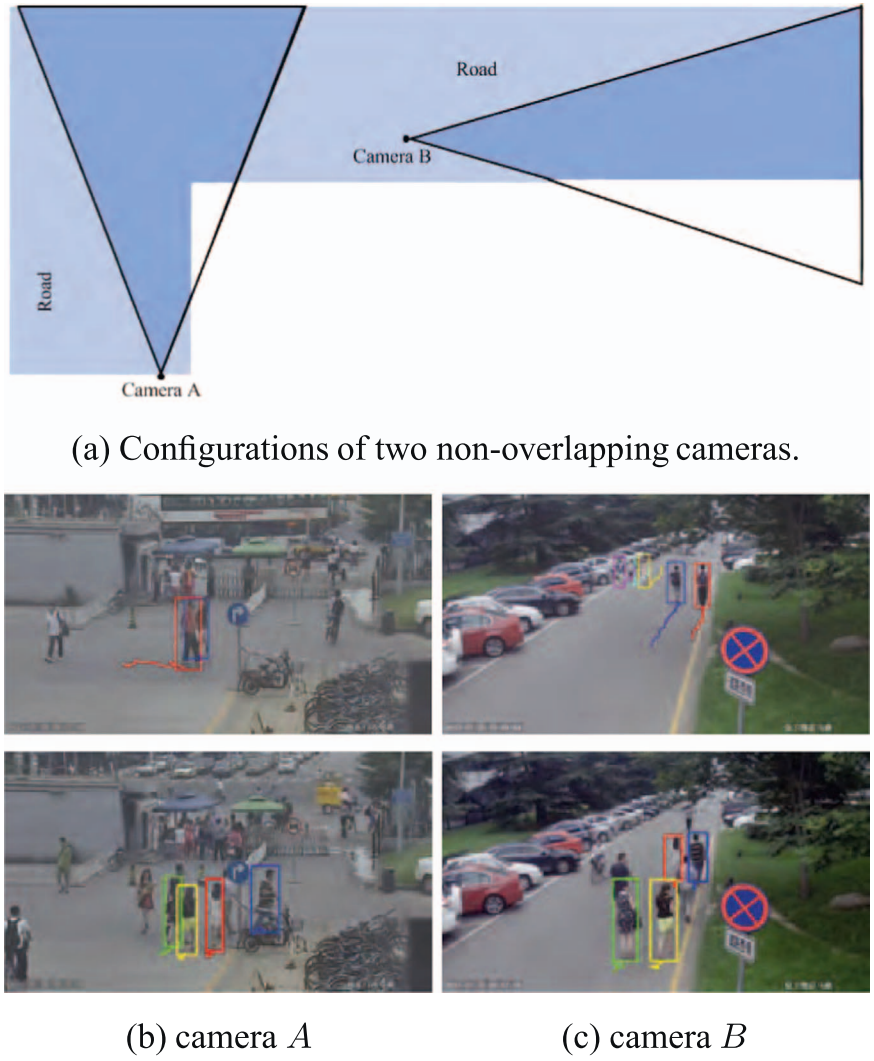

Fig. 1. Multi-person tracking across two non-overlapping uncalibrated cameras. (a) Configurations of two non-overlapping cameras.

lapping cameras of interest. Body part representations (i.e. head, left shoulder, right shoulder, left leg, right leg, etc) not only increase discrimination power for different appearances, but also consider the layout of two persons (e.g. white T-shirt with black skirt versus white shirt with blue jeans). To detect part, we adopt the pictorial structure model [14] to jointly capture the spatial relations between part locations and co-occurrence between part mixtures. We extend the original detector for a single frame to an image sequence to consider spatial-temporal cues. Instead of jointly infer part locations on multiple images which is intractable, we generate a set of pose proposals on each frame and select only one candidate that encourages the parts to have consistent appearance over the entire trajectory and share similar spatial configurations on neighboring frames. We formulate the proposal selection problem as a combinatorial optimization problem that can be efficiently solved using the Generalized Minimum Clique Graph (GMCP) algorithm [4].

In addition, to establish trajectories association between two non-overlapping uncalibrated cameras A and B, we also consider group activities of persons. In real scenarios, persons often walk in groups. Therefore, two close persons in A may also walk closely in B. These contextual cues, as well as the part-aware correspondences, are jointly described in a Markov Random Field (MRF) framework. In the MRF, we cast trajectories association problem as a multi-class classification task. We assign a label to each trajectory of camera B to denote its corresponding trajectory in A. Therefore the contributions of our work are mainly two folds: (i) We propose to utilize part-aware correspondence for trajectories association across non-overlapping cameras. The part locations are jointly detected on an image sequence incorporating spatial-temporal cues. Based on part representations, we consider both the appearance and spatial layout of two persons to achieve more robust association. (ii) We formulate the cross-camera trajectory association problem as a multi-class classification problem where each trajectory in camera A forms a class, and each trajectory in camera B needs to be classified into one of these classes. We define an energy function that encodes the part-aware appearance variations, group activities, mutual exclusion information between trajectories, etc. We use a Markov Random Field (MRF) model to effectively solve the multi-class classification problem with all these cues.

The rest of the paper is organized as follows. Section 2 briefly reviews related work of data association with different camera settings. We present our proposed approach in Sections 3 and 4 in details. Extensive experiments are conducted in Section 5 to verify the effectiveness of our approach. Finally, we discuss future work and conclude this paper in Section 6.

2. Related work

In this section, we briefly review previous works of multi-target tracking according to different settings of cameras. Particularly, we focus on data association models, not the re-identification models.