Pattern Recognition 48 (2015) 1760–1772

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust multiple cameras pedestrian detection with multi-view Bayesian network

Peixi Peng [a], Yonghong Tian [a,*], Yaowei Wang [b,a], Jia Li [c], Tiejun Huang [a]

[a] National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, China
[b] Department of Electronic Engineering, Beijing Institute of Technology, China
[c] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, China

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form
14 September 2014
Accepted 7 December 2014
Available online 17 December 2014

Keywords:
Pedestrian detection
Multiple cameras
Multi-view model
Bayesian inference
Height adaptive projection

ABSTRACT

Multi-camera pedestrian detection is the challenging problem in the field of surveillance video analysis. However, existing approaches may produce "phantoms" (i.e., fake pedestrians) due to the heavy occlusions in real surveillance scenario, while calibration errors and the diverse heights of pedestrians may also heavily decrease the detection performance. To address these problems, this paper proposes a robust multiple cameras pedestrian detection approach with multi-view Bayesian network model (MvBN). Given the preliminary results obtained by any multi-view pedestrian detection method, which are actually comprised of both real pedestrians and phantoms, the MvBN is used to model both the occlusion relationship and the homography correspondence between them in all camera views. As such, the removal of phantoms can be formulated as an MvBN inference problem. Moreover, to reduce the influence of the calibration errors and keep robust to the diverse heights of pedestrians, a height-adaptive projection (HAP) method is proposed to further improve the detection performance by utilizing a local search process in a small neighborhood of heights and locations of the detected pedestrians. Experimental results on four public benchmarks show that our method outperforms several state-of-the-art algorithms remarkably and demonstrates high robustness in different surveillance scenes.

1. Introduction

Pedestrian detection is a key step in many video surveillance applications, such as pedestrian tracking, crowd analysis and event detection. In existing studies on pedestrian detection, the most challenging task is to accurately locate multiple pedestrians in crowded scenes with heavy occlusions. In this case, it is often difficult to detect the occluded persons from the 2D view obtained from a single camera.

Intuitively, a feasible solution to detect the occluded persons in a crowded scene is to use multiple cameras that can provide complementary information for the same scene. Here a latent hypothesis is that two pedestrians overlapped in some views may become separable in other views. Following this hypothesis, Sankaranarayanan et al. [1] proposed to project the foreground of each view onto the same ground plane by homography. As shown in Fig. 1(a), such projections were then fused and their intersections were assumed to correspond to locations of the probable pedestrians. By assuming that all pedestrians have the same height [2–4], the candidate locations were projected back to each view for pedestrian detection (as shown in Fig. 1(b)). Similar to [1], Khan and Shah [5] projected the foreground from each view to a reference view. Kim [6] utilized lines to represent foregrounds in each view and projected those lines from each view to the ground plane.

Generally speaking, these approaches can achieve promising results in those scenes with weak occlusions. However, they may fail when the scenes become extremely crowded. In this case, the projection of one pedestrian may falsely intersect with the projections from some other pedestrians and such intersections may lead to phantoms (as shown in Fig. 1(c)). Moreover, the assumption that all pedestrians have the same height may not always hold. When projecting back to the camera views, the changing pedestrians' heights, as well as the synthesis noise from camera parameters, may lead to inaccurate detection results in all views (as shown in Fig. 1(d)).

To address these problems, we propose a novel approach for pedestrian detection in multiple cameras by using multi-view Bayesian network (MvBN). The overall framework of the proposed approach is shown in Fig. 2. We first obtain a set of preliminary detection results using the existing multi-view pedestrian detection methods with some predefined parameters (e.g., heights of pedestrians). Such results can be

* Corresponding author.
E-mail address: yhtian@pku.edu.cn (Y. Tian).

http://dx.doi.org/10.1016/j.patcog.2014.12.004
0031-3203/

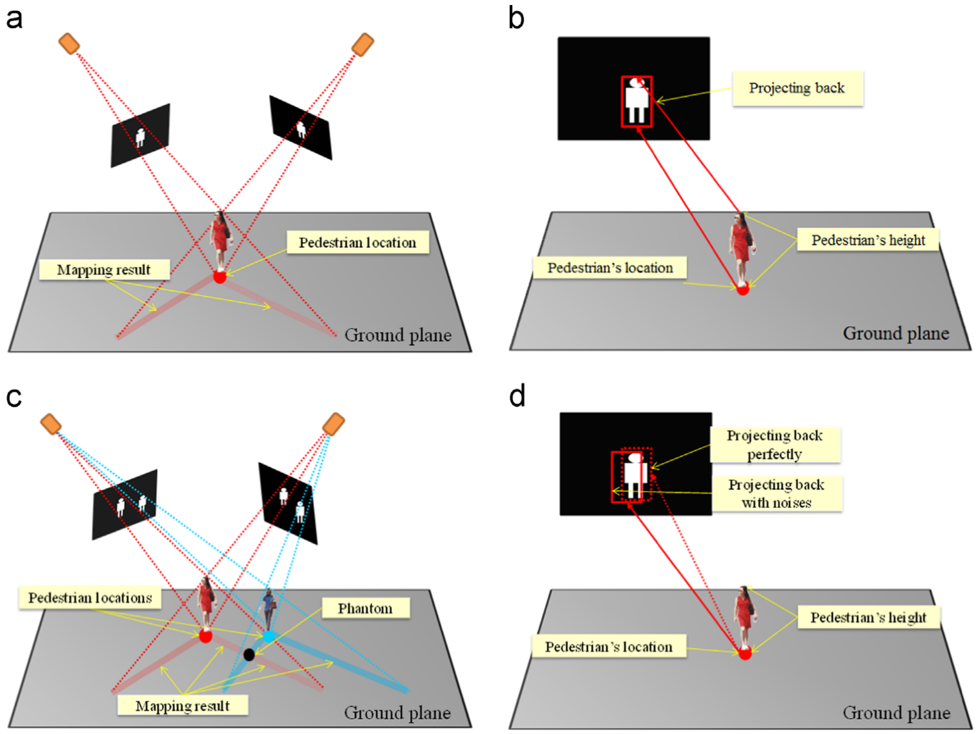

Fig. 1. (a) Homography-based multi-camera pedestrian detection method. (b) Projecting the location back with a predefined height to generate a detection result in the view. (c) The projection of one pedestrian falsely intersects with the projections of other pedestrians, consequently leading to phantoms. (d) An example of the synthesis noises.

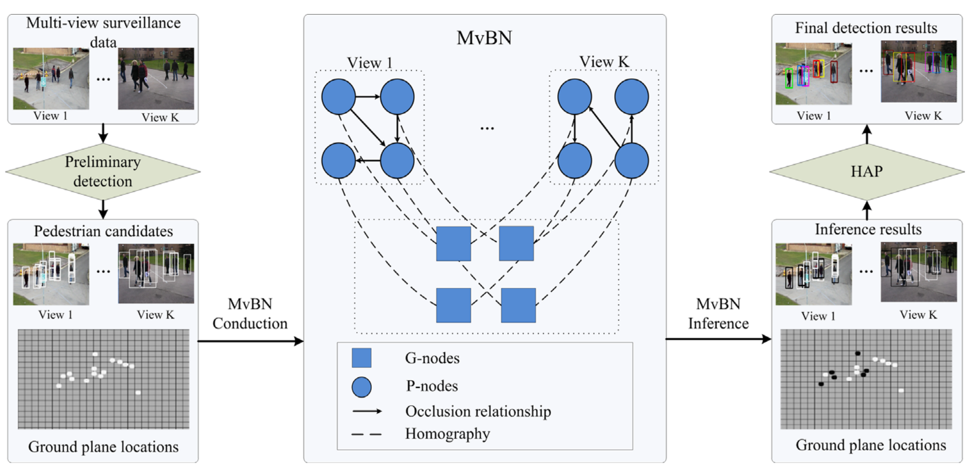

Fig. 2. The system framework of our approach. In the framework, the pedestrian candidates in all views and the corresponding locations on the ground plane are obtained by existing multi-view pedestrian detection methods with some predefined parameters (e.g., heights of pedestrians). Then, a Bayesian network is used to model the occlusion relationship between all candidates in each single view, while multiple Bayesian networks can be further combined to form a MvBN by considering the homography from the ground plane to the camera views. Thus phantoms (the white pedestrian candidates in Inference results) can be efficiently removed by inferring the G-nodes in the MvBN. Finally, the HAP is used to further refine the final detection results in each view, making the proposed method adaptive to diverse pedestrians' heights and robust to synthesis noises.

represented as the pedestrian candidates in all views and the corresponding locations on the ground plane. After that, a Bayesian network is used to model the occlusion relationship among all candidates in each camera view, and then multiple Bayesian networks can be further combined to form a MvBN by the homography from the ground plane to camera views. The MvBN includes two kinds of nodes that represent the pedestrian candidates (i.e., P-nodes) and the locations on the ground plane (i.e., G-nodes), respectively. The edges between P-nodes are used to model the occlusion relationship, while the edges between P-nodes and G-nodes represent the homography from the ground plane to different camera views. In other words, the MvBN is actually composed of G-nodes and several Bayesian networks, where G-nodes are used to combine the inference results from different Bayesian networks. Since phantoms are always concurrent with occlusions, such phantoms can be efficiently removed by inferring the G-nodes that demonstrate the highest probabilities of occluding the corresponding P-nodes.

Note that the preliminary work about MvBN have been published in [7]. But that version could not handle camera calibration noise and diverse pedestrians' heights. In order to solve this problem, a height-adaptive projection (HAP) method is proposed in this study. The HAP is used to further refine the detection results by utilizing a local search process in a small neighborhood of heights and locations of the detected pedestrians. In addition, more extensive experiments are conducted in this paper to demonstrate the effectiveness of the proposed approach. We test our approach on four public benchmarks,