Multi-camera Pedestrian Detection with a Multi-view Bayesian Network Model

Peixi Peng
pxpeng@jdl.ac.cn
Engineering Laboratory for Video Technology, Peking University

Yonghong Tian
yhtian@pku.edu.cn
Engineering Laboratory for Video Technology, Peking University

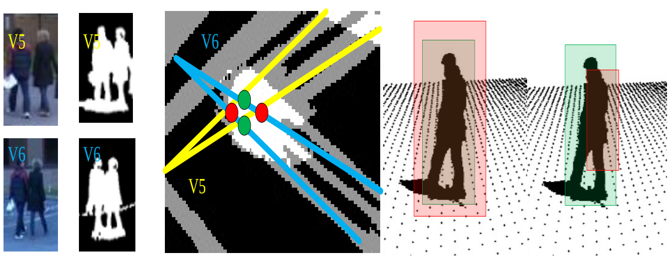
Yaowei Wang
ywwang@jdl.ac.cn
Department of Electronic Engineering, Beijing Institute of Technology

Tiejun Huang
tjhuang@pku.edu.cn
Engineering Laboratory for Video Technology, Peking University

In recent years, more and more cameras are widely deployed for video surveillance in a cooperative manner. In such scenarios, multiple-pedestrian detection has become an essential technology for many applications such as crowd behaviour analysis. Often, occlusions among pedestrians will complicate the detection process and make it difficult for the system to accurately detect the pedestrians after heavy occlusion. In this sense, the availability of multi-view information will make pedestrian detection easier and more accurate..

In this paper, we estimate the occupancy possibility of each location that can then be used to predict the occurrence of a pedestrian in this location. We integrate occupancy possibility of all views together as the final occupancy possibility on the ground plane. The general method [1], intersecting the view lines from multi-cameras on the ground plane, yield satisfying results when people are well-separated in multi-views. When occlusion becomes more frequently, these approaches will cause many "phantom" phenomena. Phantoms are the intersections of viewing rays at locations that are not occupied by any pedestrians (as shown in Figure 1(a)), which has also been reported in previous work [2].

To address this problem, we first classify the phantoms in a single view into two categories. The first-class phantoms are those who occlude some pedestrians (e.g., the right one in Figure 1(b)). Often, the phantoms of this kind are generated due to the projection of inaccurate foreground extraction results on the ground plane. In this case, if these phantoms are directly treated as detection results, the matching degree with the foreground masks should be much less than the pedestrians which are occluded by them. In order to reduce the first-class phantoms in the multi-view projection, the key point is to make the detection results best match the foreground masks using the occlusion relationship among phantoms and pedestrians. On the other hand, the second-class phantoms denote those that are occluded by pedestrians, despite they can also match the foreground masks well (e.g., the left one in Figure 1(b)). The reason for generating the phantoms of this kind is usually due to the non-invertible mapping from 3D world coordinates to 2D image coordinates. These phantoms always be occluded by pedestrians mostly. Thus to reduce the second-class phantoms, we need to estimate the non-occluded parts for each phantom. Hence, our method proposed in this paper try to reduce phantoms by analysing the occlusion relationship among potential pedestrians at different locations in all views.



(a) (b)

. Figure 1: (a) An example of phantom: locations (red circles) are occupied by phantoms and locations (green circles) are occupied by pedestrians; (b) The first-class (left) and second-class (right) phantoms, where the red ones denote phantoms

By summarizing the two cases above, we can conclude that the key problem to reduce the possible phantoms in the multi-view projection is to effectively model and utilize the occlusion relationship among potential pedestrians at different locations in all views. It is notable that a 1st phantom in one view may be the 2nd phantom in another view and vice versa. Considering it, a multi-view Bayesian network (MBN) is proposed in this paper. In general, a MBN is constructed with the locations on the ground plane and several single Bayesian networks (SBNs), where each SBN is used to characterize the potential occlusion relationship of all locations in a single view, while the locations on the ground plane is used to establish the correspondence among all SBNs through the geometric constraints among cameras (See Figure 2). Moreover, we also model the "subjective supposing" node states (SSNS) as a set of Boolean parameters of MBN, which are then used to denote whether a pedestrian occurs at the locations. In fact, SSNS can distinguish the 1st phantoms with the 2nd pedestrians. During MBN inference, we can estimate the part of the pedestrian at this location which are not occluded by other pedestrians based on SSNS. A learning algorithm is then proposed to estimate the SSNS parameters of the MBN, by finding such a configuration that the final occupancy possibility can best explain the image observations (i.e., foreground masks) from different views. The overall framework of our method is shown in Figure 2.

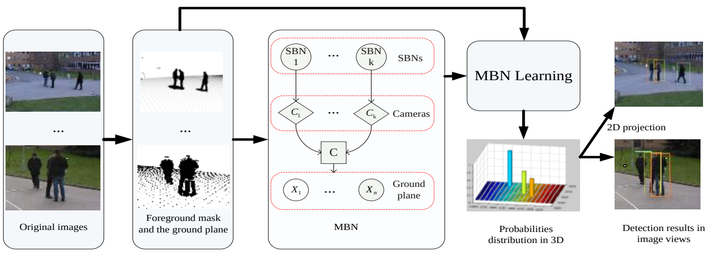


Figure 2:The framework of our approach

Implementation details of MBN model and SSNS learning is described in the paper. The experimental results on PETS2009S2L1 and APIDIS benchmark datasets demonstrate the effectiveness of our method compared with other state-of-the-art methods [2] [3][4].

References

[1] A. C. Sankaranarayanan, A. Veeraraghavan, and R. Chellappa. Object detection, tracking and recognition for multiple smart cameras. *Proc. of the IEEE,* 96(10):1606–1624, 2008.

[2] W. Ge and R. T. Collins. Crowd detection with a multiview sampler. In *Proc. of ECCV*, pages 324–337, 2010

[3] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):267–282, 2008.

[4] Alahi A., Jacques L., Boursier Y., and Vandergheynst P., "Sparsity Driven People Localization with a Heterogeneous Network of Cameras,"*Journal of Mathematical Imaging and Vision,* vol.41,pp.39-58, 2009.