Detection and Tracking of Occluded People

Siyu Tang
tang@mpi-inf.mpg.de
Mykhaylo Andriluka
andriluka@mpi-inf.mpg.de
Bernt Schiele
schiele@mpi-inf.mpg.de

Computer Vision and
Multimodal Computing
MPI Informatics
Saarbrücken, Germany

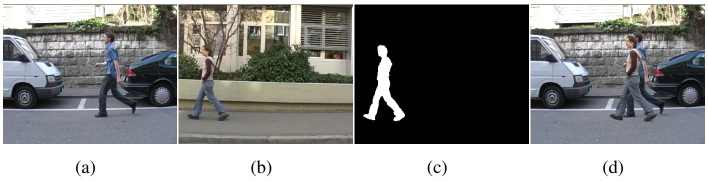

Figure 1: Procedure to synthetically generate training images for our double-person detector. (a) background person, (b) foreground person, (c) foreground person map, (d) generated synthetic training image.

We consider the problem of detection and tracking of multiple people in crowded street scenes. Several methods, i.e. tracking and 3D scene reasoning approaches [3, 7, 10], have been proposed to track people even in the presence of long-term occlusions. While these approaches allow to reason across potentially long-term and full occlusions they still require that each person is sufficiently visible at least for a certain number of frames. State-of-the-art approaches to people detection [5, 6] are able to reliably detect people under a variety of imaging conditions, people poses, and appearance, but their performance degrades when people become partially occluded. Careful reasoning about association of image evidence to detection hypotheses has been proposed in [4, 8, 9], but these approaches treat partial occlusion as nuisance and perform decisions based on the image evidence that corresponds to the visible part of the person, which makes them unreliable in cases of severe occlusions.

Here, we explore an alternative strategy, observing that in crowded street scenes most occlusions happen due to overlaps between people, we consider the joint evidence of both people. This is possible since overlapping people result in characteristic appearance patterns that are otherwise uncommon. Our approach builds on the powerful deformable part models (DPM [6]), which we extend in three ways. First we propose a new double-person detector that allows to predict bounding boxes of two people even when they occlude each other by 50% or more. Second, we propose a joint person detector, that is jointly trained to detect single- as well as two-people in the presence of occlusions. Last, we integrate the above joint model into a tracking approach to show its potential for people detection and tracking.

Double-person detector: We build the double-person model upon the DPM framework to detect the presence of two people and to predict the bounding boxes of both people, the occluding person as well as the occluded person. For training, we synthetically generate two-people samples (Fig. 1) based on the TUD training data [1]. The synthetic images are ideal for training as they come with accurate occlusion-level estimates. We demonstrate experimentally that our double-person detector significantly outperforms a single-person detector in the presence of severe occlusions (Fig. 2).

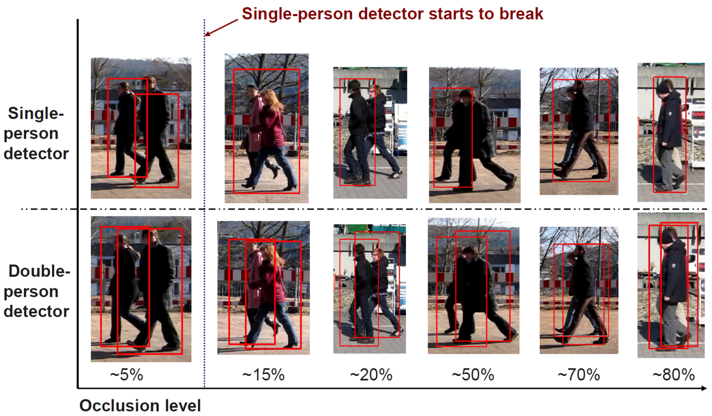

Figure 2: Qualitative comparison of single- and double-person detectors with occlusion.

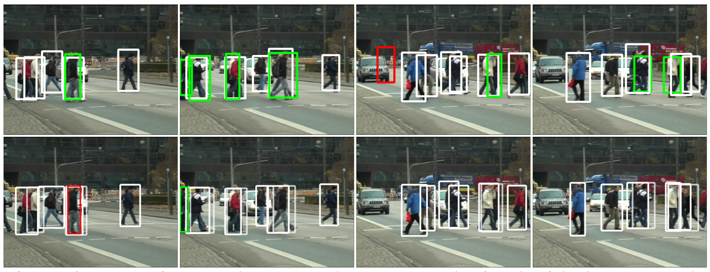

Figure 3: Detection results at equal error rate obtained with the approach of [4] (top) and our joint detector (bottom) on the TUD-Crossing [1] dataset. False-positive detections are shown in red and missing detections in green. One of the two bounding boxes predicted from the two-person detection is shown with the dotted line.

Multi-Person Detection: The joint person detector is again built upon the DPM-approach where the role of the different components is now to differentiate both between single and two people as well as between different occlusion levels among two people. Similarly to double-person detector we initialize the double-person components with training examples corresponding to gradually increasing levels of occlusion. For the single-detector components we rely on the standard initialization based on the bounding box aspect ratio. Our experiments on TUD-Crossing [1] dataset confirm the benefit of the joint detector in the realistic scenes (Fig. 3).

Multi-Person Tracking: We compare the performances of a single-person and the joint detector in the context of multiple people tracking. To that end we apply the people tracking-by-detection approach of [2] without modification both to the output of the single-person and the joint detectors on TUD-Crossing [1] dataset. The tracker based on the joint detector is able to correctly track people even in cases of severe occlusions clearly showing the potential of using our joint detector as the basis for multi-people tracking in scenes with many people and in the presence of severe occlusions.

[1] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *CVPR'08*, .

[2] M. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estimation and tracking by detection. In *CVPR'10*, .

[3] A. Andriyenko, K. Schindler, and S. Roth. Discrete-continuous optimization for multi-target tracking. In *CVPR'12*.

[4] O. Barinova, V. Lempitsky, and P. Kohli. On detection of multiple object instances using hough transform. In *CVPR'10*.

[5] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark. In *CVPR'09*.

[6] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *PAMI'10*.

[7] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *ECCV'08*.

[8] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *CVPR'05*.

[9] X. Wang, T.X. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. In *ICCV'09*.

[10] B. Wu and R. Nevatia. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors. *IJCV*, 75:247–266, November 2007.