

# ARC '16

مؤتمر مؤسسة قطر  
السنوي للبحوث  
QATAR FOUNDATION  
ANNUAL RESEARCH  
CONFERENCE



Towards World-class  
Research and Innovation

## Social Sciences, Arts and Humanities Pillar

<http://dx.doi.org/10.5339/qfarc.2016.SSHAPP3413>

### Pose Detection for Partially Occluded Persons in Spectator Crowds

Arif Mahmood<sup>1</sup>, Muhammad Shaban<sup>1</sup>, Haroon Idrees<sup>2</sup>, Nasir M Rajpoot<sup>1</sup>, Mubarak Shah<sup>2</sup>

<sup>1</sup>Qatar University, QA

<sup>2</sup>University of Central Florida, USA, US

Email: [rfmahmood@gmail.com](mailto:rfmahmood@gmail.com)

In recent years, vision based solutions have shown improvement in performance for scenes containing single or few persons for the tasks of person detection, tracking, and action recognition. Dense crowd analysis is the next step which actually helps to solve more useful real world problems. It is crucial for surveillance, space and infrastructure management of large events such as political, religious, social, and sports gatherings.

Visual analysis of a dense crowd is significantly difficult as compare to a single or few person analyses due to a set of challenges including severe occlusion, low resolution, and perspective distortion. Such environment also offers a set of special constraints such as person visibility is strongly dependent on the position of other persons or a persons actions can also be inferred from the actions of the surrounding people.

Person pose detection in densely crowded scenes is very challenging task but also very useful for the higher level tasks like person tracking, action recognition and activity classification etc. Many automatic person pose detection methods are proposed in literature but for only single or few persons. These algorithms expect visibility of full body and therefore try to fit in all body parts. The body parts which are occluded are also forced to fit in resulting in incorrect detection (Figure 1). We present a pose detection method for partially occluded persons using the extra constraints available in the dense crowd videos. We present our results on S-Hock spectator crowd dataset. It consists 15 videos each contains 929 frames recorded by five different cameras in four ice hockey matches. The annotations (face and head boundaries) for each person in each frame are also available.

In S-Hock dataset all videos were recorded using fixed cameras. Which means we can easily calculate the expected person height and width in pixels using the intrinsic and extrinsic parameters of the camera. We use state of the art face detector to get an initial bounding box around the face of each person. We use the expected person height and width along with person's face bounding box to get initial person boundary. In a crowded environment, a

**Cite this article as:** Mahmood A, Shaban M, Idrees H, Rajpoot NM, Shah M. (2016). Pose Detection for Partially Occluded Persons in Spectator Crowds. Qatar Foundation Annual Research Conference Proceedings 2016: SSHAPP3413 <http://dx.doi.org/10.5339/qfarc.2016.SSHAPP3413>.

person is usually occluded by other persons therefore initial boundaries have significant overlap with other persons boundaries and we use this fact to correct the initial boundary of each person.

Figure 1 Results of the current state of the art human pose detection algorithm of Yong and Ramanan 2013 articulated on spectator crowd dataset Conigliaro 2015 shock. Only head (green) and arms (magenta and cyan) skeleton elements are displayed. The algorithm performed excessive errors.

Figure 2 Horizontal vs. Vertical occlusion. Left: two examples of horizontal occlusion. Green shaded regions show the bounding box of the person of interest. Right: two examples of vertical occlusion.

Figure 3 Left: ground truth vs estimated bounding box comparison. Right: horizontally occluded region and its edge representation. We check the curvature of the longest curve segment. The person corresponding to concave curvature is foreground and the person corresponding to convex curvature is in the background. For the background person we do not try to detect his occluded arm while pose detection.

First we adjust the horizontal lower boundary of a person using vertical occlusion (Figure 2). To detect the vertical occlusion we use the detected face position and define the vertical occlusion as if two detected faces horizontally overlap each other more than a certain threshold (25%) and their vertical distance is less than the person expected height then upper person is occluded by the lower one. If a person is vertically occluded by another person then we limit its lower horizontal boundary to the center of face of person below (Figure 2). Second, we shrink the vertical boundaries of a person bounding box using the horizontally adjacent persons. If any person's initial vertically boundary overlap more than fifty percent with the face of horizontally adjacent person then this boundary will be limited by the center of that face (Figure 2).

Figure 4 Results of Yong and Ramanan 2013 articulated pose detection algorithm on partially occluded persons. Since the algorithm is not able to automatically detect occluded body parts, therefore the algorithm forcefully tries to fit in the legs (red) and torso (yellow) and makes obvious mistakes.

Figure 5 After detecting the partially visible body parts of the three persons we apply the partial trained model consisting of head (green), upper torso (yellow) and arms (cyan and magenta) for pose detection. The partial model has performed quite well on the partially visible bodies and forcefully fitting of missing body parts is avoided.

After bounding box corrections we compared our estimated bounding boxes with the available ground truth and got very close match (Figure 3). In person segmentation problem, we not only need to locate the person but also separate it from the other persons and the background. For this we need to identify that which person is at the front and which is at the back in the occluded region. We divide the occlusion issue into vertical and horizontal occlusions. First, vertical occlusion occurs when a person sits in front of other person and it is normal in spectator crowds. It means the person with lower head position will be in foreground. So we sort the persons using their vertical head position and mark the person as foreground in occluded region who has lower head position. Second, horizontal occlusion occurs due to the side by side appearance of persons. It is more challenging to mark a person as foreground or background in horizontally occluded region. To cope with this challenge we find edges in the occluded region and fit a curve in these edges. If the curve is concave with respect to the right person then this person is in foreground and other in background and vice versa (Figure 3).

After having enough information about the person's location, occluded region, and boundary, the next step is to apply a state of the art pose detector to get the person pose. We use an approach similar to and train different body parts models (Figure 5) including full body, upper body or head with arms which use the persons boundary and position information as prior. At time of detection, detection model is selected based on the size and shape of person's corrected bounding box for precise pose estimation. This model selection step helps to avoid the forceful detection of occluded parts. At the end, all detections of the pose detectors are reasoned holistically with respect to other persons positions. The proposed approach has demonstrated significant improvement over the full body detector when applied to dense crowds.

In future we will use the extracted pose information from consecutive frames of a video sequence of a crowd for the purpose of action or activity recognition. Pose information will enrich the optical flow based action or activity recognition algorithms offering them excessive information. Pose information is also required to localize the motion performed by different persons in dense crowds and thus action assignment will be improved.

#### Acknowledgment

This work was made possible by NPRP grant number NPRP 7-1711-1-312 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.