



REPORT



Submitted by:

Tharun V. P

NOVEMBER 18, 2019

Problem Statement

Camera function is low : 1) there are many misrecognitions other than human body 2) If the brightness is insufficient or the distance from the camera is 5M or more, the human body recognition power becomes weak=Omissions of human body detections occur frequently and fails to detect fall.

Background

In general, pose estimation algorithms follow either a top-down approach where the person selector algorithms runs in the first stage followed by the key point extraction stage or a bottom-top approach where initially the key points are extracted followed by the generation of human body skeleton using metrics like PAF, confusion matrix etc. and techniques like Bipartite Mapping. The open pose algorithm ideally belongs to the later.

Approach 1:

- In order to tackle the problem for misrecognitions (i.e.) false detection of other objects as humans, a structured object detection algorithm like Mask-RCNN trained on the Coco dataset could be used in the primary stage.
- The output of the Mask-RCNN is basically the ROI (humans) marked by bounding boxes. This segmented portions could be demarcated and further fed to the pose estimation algorithm for anatomical key point extraction.

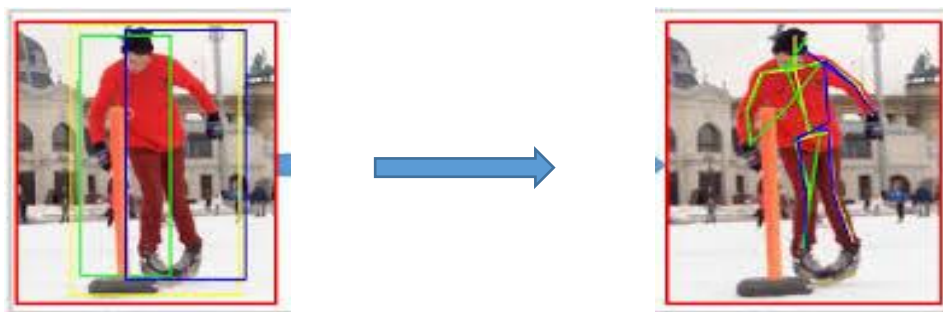


Figure 1. **Left:** Person detection using RCNN. **Right:** Pose estimation of detected person.

Things that could go wrong:

- This approach could produce significantly improved results in terms of reducing misrecognitions but may not be invariant to the scale and orientation of the ROI captured out of the image (i.e.) there is chance that pose estimation algorithm could fail as for the demarcated ROIs could be of varied dimensions and orientation.

Approach 2:

- In order to handle the variations in the orientation and scale of the captured ROI(Human) from the input image, image processing based transformation like Perspective transform can be used .

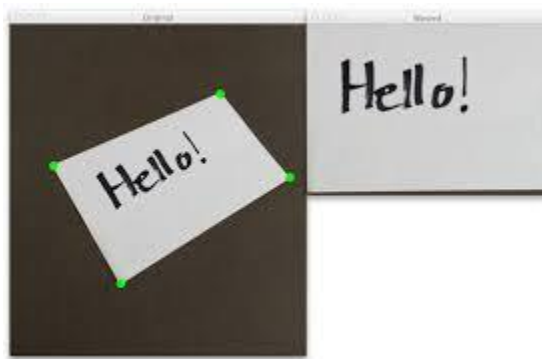


Figure 2. Result of perspective transform. The image of the disoriented white sheet is brought back into perspective.

Things that could go wrong:

- In case of occlusions or target person missing in the captured image due to distortions, the approaches above might have reduced accuracy.

Approach 3:

- Introducing person tracking algorithms like Deepsort in the primary stage could highly improve the accuracy for it could prevent scenarios where

target person might go missing from the ROI due to occlusions or camera view might be partially blocked.

- Since Deepsort technique by default involves object detection in the initial stage, the ROI (human) with bounding box could be extracted for pose estimation.



Figure 3. Person tracking implemented through deep sort. It can be seen that even in consecutive frames all the people in the image are tagged with the same id as that of the previous frame.

Conclusions:

- Approach 1 brings in an improvement in accuracy for it ensures to a large extent that the target person is detected every time before pose estimation is conducted. With the inclusion of Perspective transform in Approach 2, it can be ensured that disorientation and scale variance of the ROI extracted through Mask-RCNN is corrected before performing pose estimation. Finally, Approach 3 happens to be an optimized solution considering it can handle the tracking of the target person real-time even in scenarios of occlusions and feed distortions thereby ensuring there is no any misrecognition.

Suggestions for Improvements:

- A setup with two cameras can be installed instead of one thereby ensuring complete coverage of the room is obtained and there would be better insights of the inmates' actions

- A combination of an RGB and thermal camera could be used with a model trained on RGB images and thermal images respectively, this way the low lighting scenarios can be handled by the thermal model and the normal scenarios can be handled by RGB model.
- 360° motion based camera could also enhance the performance and at the same optimize power utilization.

References:

1. <https://towardsdatascience.com/human-pose-estimation-simplified-6cfd88542ab3>
2. <https://towardsdatascience.com/people-tracking-using-deep-learning-5c90d43774be>
3. <https://www.ijert.org/human-fall-detection-using-convolutional-neural-network>
4. <https://arxiv.org/pdf/1909.13423.pdf>
5. <https://arxiv.org/abs/1602.00134>
6. <https://medium.com/@laanlabs/real-time-3d-car-pose-estimation-trained-on-synthetic-data-5fa4a2c16634>
7. <https://arxiv.org/pdf/1812.08008.pdf>