3D Human Pose Tracking inside Car using Single RGB Spherical Camera

Pramod Murthy, Onorina Kovalenko, Ahmed Elhayek,
Christiano Gava, and Didier Stricker

DFKI - German Research Center for Artificial Intelligence
{pramod.murthy,onorina.kovalenko,ahmed.elhayek
christiano.gava,didier.stricker}@dfki.de
http://av.dfki.de

Abstract. The recent progress in Deep Learning methods in computer vision has resulted in improved Advanced Driver Assistance Systems (ADAS). The goal of ADAS is not only to assist drivers, but also to alert them before dangerous driving maneuvers. ADAS often do not capture human motions to augment it into the driving context. In this work, we attempt to propose a system to track 3D motion of humans (especially drivers) inside a car using single RGB spherical camera. We use a CNN based 3D human pose tracking system to track driver and passengers poses. Finally, we illustrate accurate results in real-time on recorded driving scenes.

Keywords: 3D human pose tracking, CNN, human activity, deep learning

1 Introduction

Over the last decade various novel driver assistance technologies were invented for safe driving experience [9]. These Advanced Driver Assistance Systems (ADAS) technologies provide information such as navigation, lane keeping, and collision avoidance in poor lighting conditions etc. without augmenting drivers behavior in the driving context [7]. These systems need to detect potential hazards on the road by simultaneously monitoring driver and road, which requires pose tracking, object detection [8] and data fusion [5,7,6]. Face detection and gaze estimation is still unarguably difficult problem for extreme head poses [10] or in challenging lighting conditions [9,4].

The recent success of deep learning methods in computer vision have resulted in novel methods of markerless human motion capture. Often RGB-D cameras are used to capture human motion in 3D. However, RGB-D cameras do not provide reliable depth measurement in outdoor scenes (eg. sunlight interference, and glass in front of camera), have lower resolution and not widely acceptable as color cameras. In this work, we track humans especially drivers poses and monitor the body and head movements using RGB spherical camera. For this purpose, we equip a car with a single RGB spherical camera device to capture

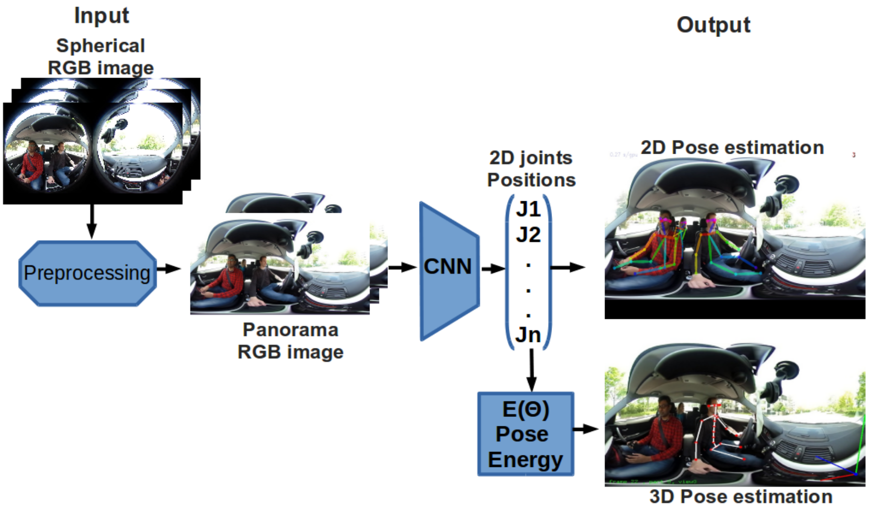


Fig. 1. Proposed system using spherical camera to track passenger inside car.

the driving context inside the car. Usage of only one spherical camera allows to monitor scene both inside and outside the car.

2 Proposed 3D pose tracking system

The overview of the proposed system to estimate 3D human pose inside the car is as shown in Fig. 1. We capture the scenes using single RGB spherical camera and stitch the two views to get the panoramic pixel map. We apply the CNN models to get the 2D joint positions for every person as described in Section 2.1. These 2D joints are combined with energy based model to estimate the final 3D pose of the human (Section 2.2). We build on works of 2D pose CNN regression [2,1] and combine with our skeleton fitting method proposed by [3].

2.1 2D joints detection

We selected two discriminative approaches: single person convolutional part heatmap regression [1] and multi-person 2D pose estimation using part affinity fields [2]. The convolutional part heatmap regression model is applied to each person individually by cropping the person centered bounding box. Thereafter, 2D pose is estimated for individual person separately. We also applied the recent multi-person 2D pose estimator over the complete frame which regresses heatmaps and the part affinity field. Further, these regression heatmap outputs and part affinity fields are combined to estimate the final 2D joint positions along with its confidence measure for every person in the scene.

2.2 3D pose estimation

In the previous section, we discussed two approaches of approximating 2D joint positions [2,1]. $E_{BP}(\Theta)$ measures the similarity between these 2D joint positions and the current pose. The best 3D pose which fits these positions is estimated by optimizing the following energy:

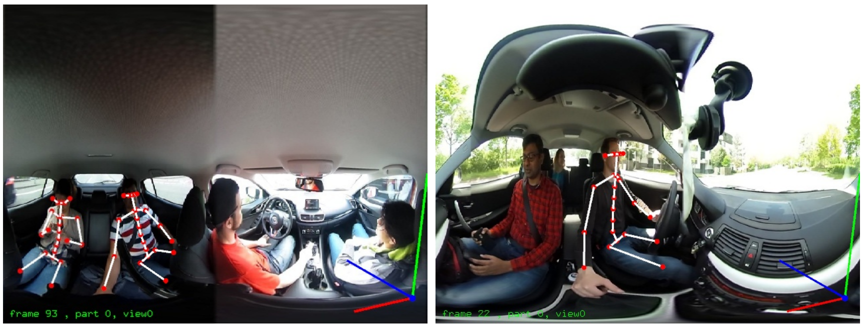
$$E(\Theta) = E_{BP}(\Theta) - w_l E_{lim}(\Theta) - w_a E_{acc}(\Theta) \tag{1}$$

$E_{lim}(\Theta)$ limits the impossible poses while E_{acc} prevents over-fitting of skeleton model to the 2D position by penalizing too strong accelerations. The weights $w_l = 0.1$ and $w_a = 0.05$ are kept constant in all experiments; see [3] for details.

3 Results

We recorded two large sequences (i.e. over 20 minutes) with a spherical RGB camera from two different positions inside the car; see Fig 2. To this end, *RICOH Theta S* is used. We did a qualitative evaluation of our results as shown in Fig 2. Please see our accurate tracking results in the supplementary video [1].

The outputs of multi-person 2D pose detector architecture combined with energy based skeleton fitting model outperformed the convolution heatmap regression model both in-terms of speed and accuracy. We also observed that 2D pose estimation of joints, especially knees, were noisy due to large fish eye distortions inherent spherical images. However, the distortion was mitigated by using local projection of every person onto smaller part of sphere resulting in a less distorted input image.



(a) (b)

Fig. 2. 3D pose tracking of passengers with the 3D skeleton superimposed for (a) passengers and (b) driver

[1] Demo video: `http://goo.gl/qhGZyG`

4 Conclusion

We propose a novel system of 3D human pose tracking inside the car with single RGB spherical camera. Our accurate tracking results illustrated robustness to strong distortion, noisy 2D joint detection and illumination changes inside the car. In future, we will evaluate the system quantitativly over more sequences and extend to estimate drivers precise head pose and gaze direction. With the help of our human pose information, the driving behavior can be understood which will lead to safer driving experience in semi-automatic (or fully autonomous) cars.

References

1. Bulat, A., Tzimiropoulos, G.: Human pose estimation via convolutional part heatmap regression. In: European Conference on Computer Vision. pp. 717–732. Springer (2016)
2. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: CVPR (2017)
3. Elhayek, A., de Aguiar, E., Jain, A., Thompson, J., Pishchulin, L., Andriluka, M., Bregler, C., Schiele, B., Theobalt, C.: Marconiconvnet-based marker-less motion capture in outdoor and indoor scenes. IEEE transactions on pattern analysis and machine intelligence 39(3), 501–514 (2017)
4. Jain, A., Koppula, H.S., Raghavan, B., Soh, S., Saxena, A.: Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In: ICCV. pp. 3182–3190 (2015)
5. Jain, A., Koppula, H.S., Soh, S., Raghavan, B., Singh, A., Saxena, A.: Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture. arXiv preprint arXiv:1601.00740 (2016)
6. Jain, A., Singh, A., Koppula, H.S., Soh, S., Saxena, A.: Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In: ICRA. pp. 3118–3125. IEEE (2016)
7. Laugier, C., Paromtchik, I.E., Perrollaz, M., Yong, M., Yoder, J.D., Tay, C., Mekhnacha, K., Nègre, A.: Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety 3(4), 4–19 (2011)
8. Redmon, J., Farhadi, A.: Yolo9000: Better, faster, stronger. arXiv preprint arXiv:1612.08242 (2016)
9. Rezaei, M., Klette, R.: Look at the driver, look at the road: No distraction! no accident! In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 129–136 (2014)
10. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: CVPR. pp. 3476–3483 (2013)