Bicycle Pose Estimation From Monocular Video

Gonzalez, Benjamin
Supervisor 1
B.GonzalezToledo@tudelft.nl

Moore, Jason
Supervisor 2
J.K.Moore@tudelft.nl

Seth, Ajay
Supervisor 3
A.Seth@tudelft.nl

February 13, 2025

1 Description

Traditional methods to analyse the motion of bodies, such as humans or vehicles, rely on Motion Capture systems or plenty of sensors. These methods offer great precision and reliability, however, they are only available in specialised laboratories or instrumented vehicles, which limits the analysis of 'in-the-wild' data. For this reason, during the last years, and supported by the constant performance increase of computer vision techniques, the scientific community have been developing tools to extract motion data from videos. This method is called 'Pose estimation' and has received great attention in the last ten years, mainly focused on human motion [5, 3, 7]. However, there is no publicly available implementation of this methodology for estimating the motion of a bicycle. Therefore, we propose a project to implement similar methodologies into bicycle motion analysis, which will allow researchers to analyse bicycles in-the-wild and enhance their understanding.

2 Objective

Extract bicycle's kinematic data from monocular videos and compare the performance of different training approaches.

3 Suggested approach

To carry this project, the suggested approach consists in a three-stage pipeline, based on [1, 2]:

1. Use computer vision algorithms to detect keypoints on the bicycle (see Figure 1).
2. Convert 2D keypoints into 3D point cloud with different methodologies (see Figures 2 and 3).
3. Compare the performance of the different methodologies.

3.1 Known methodologies

From [7], it is known that three different methodologies to go from 2D keypoints to 3D cloud of points are:

- Heatmap from depth estimation.
- Train the algorithm with motion capture data.
- Train the algorithm with synthetic data.

References

[1] Kevin Gildea, Daniel Hall, and Ciaran Simms. "Forward dynamics computational modelling of a cyclist fall with the inclusion of protective response using deep learning-based human pose estimation". In: *Journal of Biomechanics* 163 (2024), p. 111959. ISSN: 0021-9290. DOI: 10.1016/j.jbiomech.2024.111959.

[2] Daniel Hall, Kevin Gildea, and Ciaran Simms. "Drainage Troughs as a Protective Measure in Subway–Pedestrian Collisions: A Multibody Model Evaluation". In: *Applied Sciences* 14.22 (2024). ISSN: 2076-3417. DOI: 10.3390/app142210738.

Figure 1: Example of annotated keypoints of a bicycle.

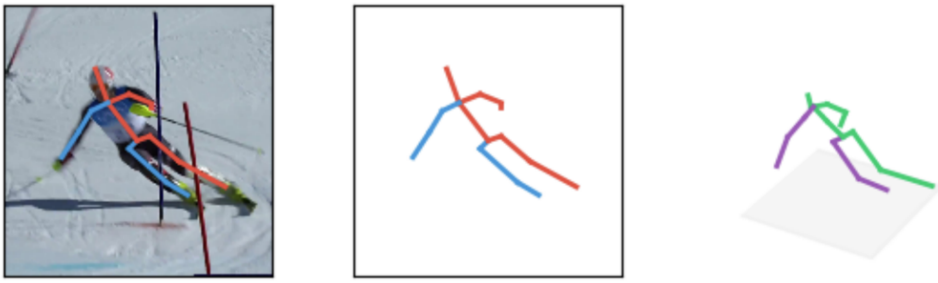

Figure 2: Example of human pose estimation from monocular video. Modified from [4].

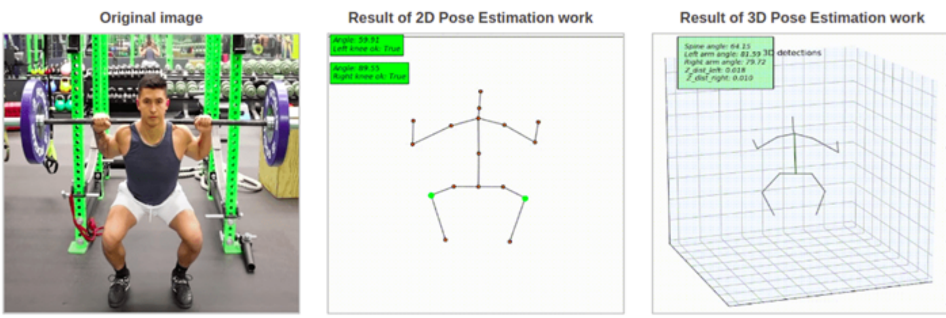

Figure 3: Example of human pose estimation from monocular video. Extracted from [6].

[3] Jiefeng Li et al. "CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmark". In: *arXiv preprint arXiv:1812.00324* (2018).

[4] Julieta Martinez et al. "A simple yet effective baseline for 3d human pose estimation". In: *CoRR* abs/1705.03098 (2017). arXiv: 1705.03098. URL: http://arxiv.org/abs/1705.03098.

[5] Dario Pavllo et al. "3D human pose estimation in video with temporal convolutions and semi-supervised training". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019.

[6] Liubov Zatolokina. *Human Pose Estimation Technology in Fitness Rehab Therapy Apps*. https://mobidev.biz/blog/human-pose-estimation-technology-guide. Nov. 2024. (Visited on 02/13/2025).

[7] Wentao Zhu et al. "MotionBERT: A Unified Perspective on Learning Human Motion Representations". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.