Real-time 3D Human Pose Estimation on Smartphones: Evaluating Approaches for Mobile Health Applications

Modern digital health applications aim to support patients in the treatment of illnesses or the compensation of impairments. Particularly in the case of physical complaints, training exercises are employed, which have to be practiced by the user during the duration of therapy. While physiotherapists traditionally demonstrate the correct execution of exercises and point out mistakes to patients, in digital health applications automated error detection and correction can be achieved via human pose estimation (HPE).
To provide sensible feedback on exercise performance, patient's estimated movements are compared to the optimal form. For this and similar healthcare and training use cases, user's movements need to be tracked and evaluated in *real-time*, to provide the user with live feedback, in *3D*, for ambiguity robustness and error detection in all directions, and *on-device*, to improve privacy and latency and to ensure network connection independence.
Pose estimation on mobile devices entails multiple challenges, particularly resource constraints in computing power and memory. Especially the limit in processing speeds results in a trade-off between accuracy and latency, as we aim to minimize errors, while maximizing the number of processable frames per second. Furthermore, 3D HPE from a monocular input, such as a smartphone camera, is posed with difficulties due to occlusions and depth ambiguities. Using monocular 2D images, 3D keypoints can be estimated either directly or indirectly, using 2D-to-3D lifting approaches.

To advance the field of real-time 3D HPE on mobile devices, the three main goals of this paper are the recording of a testset, the creation of a comparative analysis regarding existing direct and indirect strategies, and the design of a new approach, that combines the learnings from the preceding analysis.
In order to test, how well existing approaches generalize to unseen data, we will capture a dataset with multiple subjects, using the multi-camera motion capture system at MaD Lab. This testset will include several challenges, such as occlusions and varying lighting conditions.
While there already are analyses, comparing pose estimation strategies in general, so far, no comprehensive analysis exists for HPE approaches that target mobile devices. Therefore, we will conduct an analysis, comparing existing approaches on a suitable common dataset, using meaningful error metrics, such as the Mean Per Joint Position Error and Mean Per Joint Angular Error. While some existing approaches provide implementations, others have to be implemented in order to be able to analyze them.
Using the results of the comparative analysis, a new approach to mobile HPE will be developed and implemented. The foregoing analysis helps with crucial design decisions, like the commitment to a direct or indirect estimation approach. This implementation will utilize a framework to ensure mobile, embedded and edge device support.

Project Information

Type: Master's Thesis
Advisors: Jonas Müller (M.Sc.), Dr. Eva Dorschky, and Prof. Dr. Björn Eskofier
Student: Tobias Schöner
Start – End: June 1, 2024 – November 30, 2024