A Deep Learning Approach to 3D Human Walking Pose Estimation From SensFloor Capacitive Signals

Anonymous Author(s)

Abstract

While camera-based systems are the dominant approach for human pose estimation, they face several challenges in terms of privacy and occlusion effects. These issues are of particular importance in domains such as elderly care, where it could be useful to monitor the residents to conclude about their health conditions or backtrack incidents. To assess an alternative for camera based pose estimation, this paper aims to predict walking poses using SensFloor: a capacitive-based floor which registers movement activity. In this context we analyze whether it is possible to use the floors low resolution signal to fulfil this task and to what extent certain joint positions can be predicted correctly.

For this purpose, we collected synchronized data of SensFloor signals combined with a video capture. To extract human poses from the video as reference labels for training, we use MediaPipe's 3D pose estimation model. The SensFloor signals together with the labels are used for supervised training of an LSTM neural network. To localize the poses on the floor, we use a Kalman filter which processes the same signals.

Our results demonstrate, that it is possible to predict human walking poses using the proposed methods, establishing a proof-of-concept for an alternative way of activity monitoring.

Keywords

SensFloor, Human pose estimation, Deep Learning, LSTM, Kalman Filter, Capacitive Floor, MediaPipe

Code: https://github.com/sensfloor

1 Introduction

2 Methods

[1]

2.1 Training Setup

We used a CNN-LSTM architecture for training our model. It is based on the model from . Instead of rotating the step, we use a CNN to account for the Translation invariance. Figure STH shows the full architecture. Instead of giving the model all signals of the floor, we construct a Region of Interest (ROI) which is a part of the floor containing the most signals in a defined sequence. A sequence is a collection of frames. For each frame we have a state of the floor with certain signals. Our model receives the signals in the ROI frame by frame and produces a vector with continuous values as logits. These logits are given into our Loss function which can be mathematically described in the following We substract the predictions from the reference labels, square and multiply the weights of the landmarks. The link loss is identical to the one in We assume an upright position, this is why the upper body and the head will only rotate and the focus is more on the knee and feet Mediapipe extracts 33 Joints, we only used 13, excluding the fingers, face expression, heels and foot index. For each epoch we calculate the following metrics: Mean joint position error (MJPE), percentage correct keypoits (PCK) with a 5 and 10 threshold

We assume only one person is walking on the floor

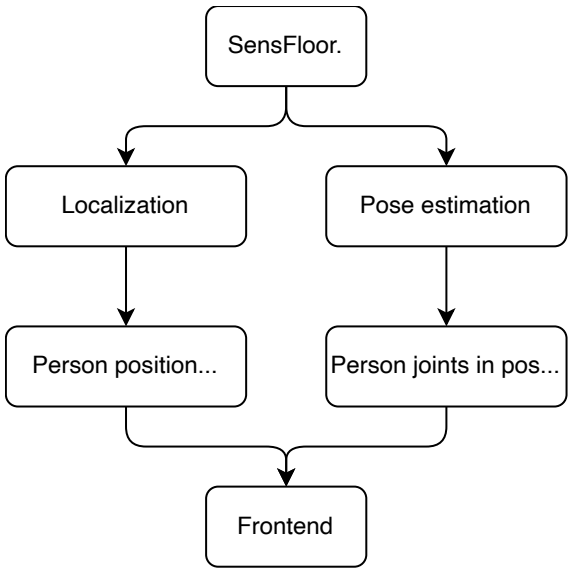

Figure 1: Test caption

3 Results

4 Discussion and Conclusion

References

[1] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. 1993. Reasoning About Naming Systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (November 1993), 795–825.