Towards Kinematically Constrained Real Time Human Pose Estimation using Sparse IMUs

Deepak Nagaraj[a], Rhett Dobinson[a] and Dirk Werth[a]

[a] *AWS-Institute for Digitized Products and Processes, Uni-Campus Nord, Saarbruecken, Saarland, Germany*

Abstract

Real time human posture estimation using reduced number of sensors is a challenging and highly sought after problem. Various model-based methods have been developed over the years in this direction which utilize optical and/or inertial sensor data. Although these methods have proven effective in laboratory settings, their applicability in the real world is limited due to the difficulty in information gathering, high intrusiveness and higher cost. This non-position paper deals with a hybrid approach involving full-body inverse kinematics (IK) and deep learning in order to estimate physiologically feasible joint angles in real time, based on orientation information from 6 inertial measurement units (IMUs). IK is performed on a kinematically constrained 3D human body model, to obtain joint angles of the body model, given orientation data of 17 sensors attached to different bone segments of the body. A bidirectional recurrent neural network (bi-RNN) is then trained using a newly collected IMU dataset to regress from the orientation data of 6 sensors to the joint angles obtained from IK. The training converged to a mean squared error (MSE) of 5.98 degrees.

Keywords

MoCap, Recurrent Neural Network, Inverse kinematics (IK), Sparse IMUs

1. Introduction and Motivation

Automated and accurate human motion capturing and analysis is a key requirement in many applications such as motion rehabilitation, performance analysis of athletes, ergonomic assessment in the workplace, AR/VR applications, etc. Two main types of motion capturing systems being used are optical or ultrasound based line-of-sight (LoS) methods that require a fixed marker sensor structure, and inertial sensor systems. Although LoS methods capture motion with good accuracy in laboratory environments, they lose their applicability in cases of un-constrained free space movement as it is a prerequisite of these systems that locations and orientations of their sensing elements (i.e. cameras) are known and invariant. On the other hand, IMU consisting of an accelerometer, gyroscope and magnetometer, which detect position and motion by measuring physical quantities like acceleration, angular velocity and angular acceleration, are relatively robust in real world environments. The commercially available inertial sensor systems for full-body human motion capture consist of around 20 IMUs. In such systems, individual sensor orientation is obtained by combining accelerometer, gyroscope and magnetometer data in a sensor fusion framework, such as Kalman filtering [1]. The data from each IMU is applied to a human bio-mechanical model using kinematic equations to obtain the position and orientation of the body segments. Xsens [2], which uses 17 sensors to reconstruct posture, demonstrated a RMS difference of less than 5 degrees in joint angles when compared to an optical position measurement system. However, the requirement that the subject wears a suit of 17 sensors and other accessories makes the system highly intrusive, affecting normal movement.

Recently, several studies have been carried out towards achieving real time full-body pose estimation using a reduced number of IMUs. In particular, deep learning (DL) based methods have been investigated in order to approximate the mapping from raw IMU data to joint angles of a 3D body [3]. The major challenge in generalizing the system, as is the case in most DL approaches, is acquiring the vast amounts of data representative of all kinds of poses and tasks. This problem becomes even more pronounced if the 3D body model, whose parameters being used as ground truths for supervised DL, is not kinematically well constrained. To overcome this problem, [3] performs DL on a large synthetic dataset, which was collected via forward kinematics on virtual sensors placed on a novel body model, and predicting pose parameters from real IMU inputs in real time. However, because of the absence of noise and drift from virtual sensor data, the model performance was much better when trained only using synthetic data than when trained using real IMU data. We try to address this problem of better constraining the 3D body model and accordingly the pose estimation by utilizing the research in the field of Biomechanics towards modeling and analysis of human motion. IK can be performed using the sensor orientation data to obtain ground truth joint angles with respect to a sufficiently constrained bio-mechanical model. These joint angles, along with orientation data from a reduced number of real IMUs can then be used to train a DL model for pose estimation.

2. Method

The experimental methodology is as shown in figure 1. The left most component depicts the 17 Xsens sensors placement on the body. Given sufficient data pertaining to various movements, IK is performed on the data using orientation values of all 17 sensors to obtain joint angles of the OpenSim body model. Then, a bi-RNN is trained to regress from orientation values of 6 sensors (encircled in green) to the joint angles obtained from IK. During evaluation, the output (joint angles) from the bi-RNN, corresponding to test dataset, is visualized using OpenSim.
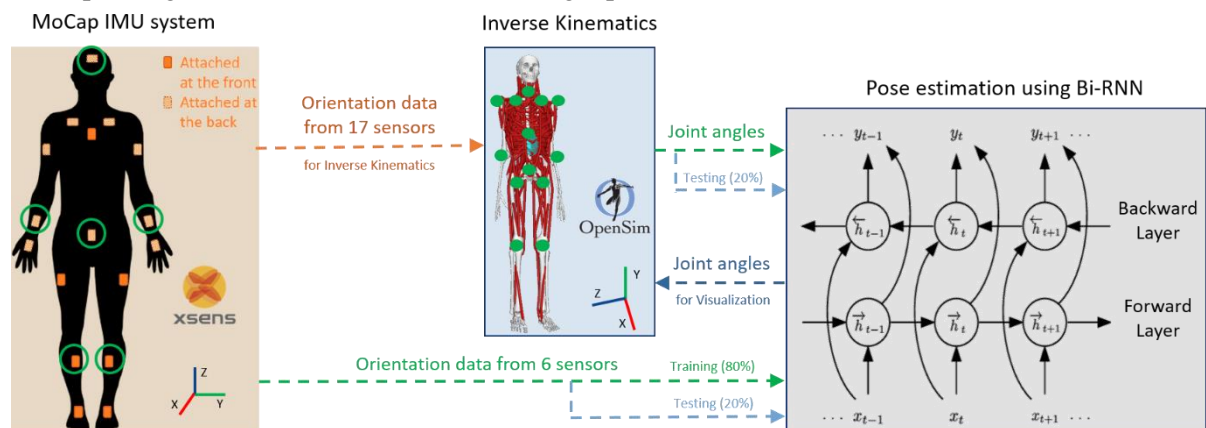

Figure 1: Proposed approach involving Inverse Kinematics and Deep Learning

2.1. Data Collection

In the scope of the project BauPrevent (more details in Acknowledgements section), wearable IMU data collection has already been carried out. Data was obtained from construction workers wearing IMU sensors undertaking their usual tasks, as the project focuses on preventive health, specifically for the construction domain. 7 hours of data from fifteen different subjects was collected resulting in 143 different motion sequences and close to a million IMU data frames. Various activities were performed targeting specific types of stresses as well as practical tasks from their typical daily routine. The following types of activities were recorded: lifting and carrying weights, lifting weights up onto a workbench, overhead fastening, plastering, grouting, sweeping, painting, hanging wallpaper and assembling/disassembling scaffolding. We utilized the commercially available Xsens [2] system that consists of 17 IMUs in total. The raw IMU data was collected along with computed information such as sensor and bone-segment orientation and joint angles however, we only use sensor orientation data as described in next sections.

2.2. Inverse Kinematics

OpenSim [5], the most widely used open-source software package for biomechanical analysis, was used in order to obtain the joint-angle ground truths with respect to the chosen body model by carrying out IK. OpenSim through its OpenSense API feature, provides an offline measurement-scaling IK pipeline, where orientations of virtual sensors placed on model bone segments are fitted to the raw sensor orientations. The model-defined anatomical joints, that represent the movements, were then extracted from OpenSim MOT motion files. In order to represent physiologically feasible joint angles effectively and efficiently, we chose the body model demonstrated in [4] for our study, which was available as an OpenSim model. The model, in XML format, contains all the information needed for the biomechanical description of the human body, including body-segments, kinematic constraints (joints) and dynamic constraints(i.e. muscles). Since OpenSim models can be freely edited, the number of joint parameters can be adjusted according to constraint requirements. In our case this was constraining the model for reduced sensor DL. This required using least number of joint parameters possible whilst maintaining the most important degrees of freedom (DOF) and allowing effective posture representation. Our resulting body model reduced body joint parameters (DOFs) from 165 to 26, relating to the 13 joints (green dots) depicted on the 3D kinematic body model in figure 1; the considered body model [4] has very detailed spine with 126 DOFs having magnitude in the range of 0 to 4 degrees mostly (as observed), of which only 4 DOFs are considered. The 26 parameters thus obtained are the ground truth values for training the bi-RNN.
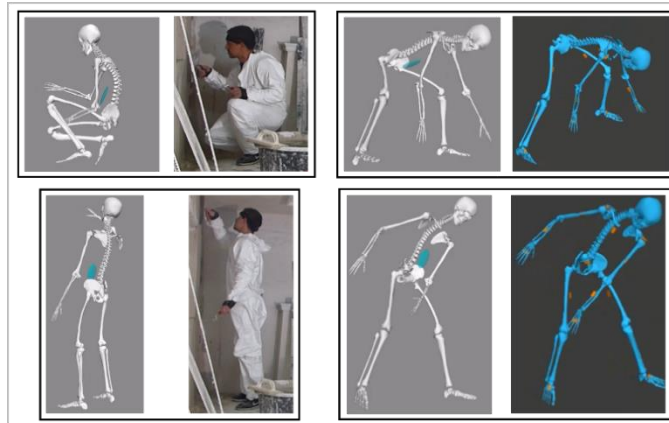
2.3. Deep Learning

The model architecture and the training methodology employed is similar to that demonstrated by [3]. A two-layer bi-RNN, each with 512 LSTM units, was trained using the sensor orientation and joint parameter data discussed in sections 2.1 and 2.2. The sensor placement was similar to that of [3], however, unlike their approach, which used sensor orientation as rotation matrices and acceleration as additional input, only sensor orientation as quaternions was used in our study; provided that the acceleration data is often too noisy, is already incorporated through sensor fusion, and that the OpenSim API doesn't require it for carrying out IK. The model was trained with sensor orientations from the 6 sensors each, consisting of 4D quaternions (resulting 24 values) as input, and 26 OpenSim model joint parameters from IK as output. The 143 different motion sequences with varying number of time-steps were divided into batches of 300 time-steps. Training and test data was split with the ratio of 0.8 to 0.2 and then shuffled. As depicted in figure 2, the mean squared error converged on training and test sets to 3.30 degrees and 5.98 degrees respectively after 89 epochs.
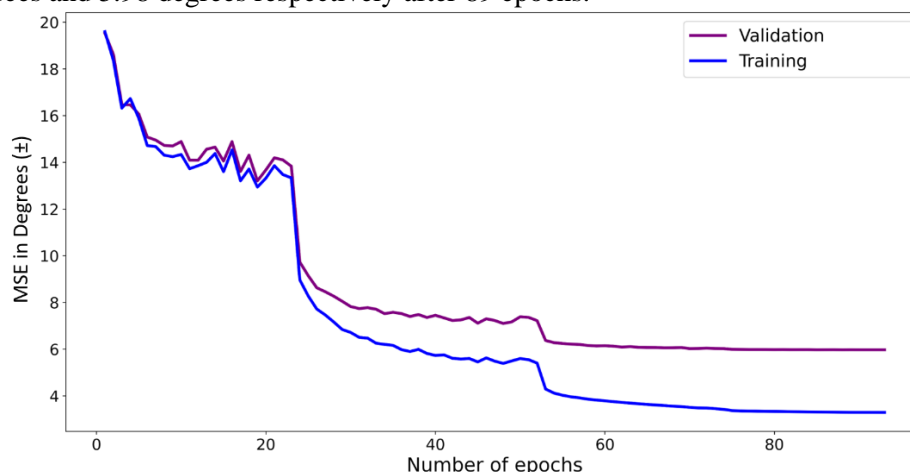
Figure 2: Convergence of MSE during training the bi-RNN

Figure 3 depicts the visualization through OpenSim, the output of the bi-RNN model (white skeleton) on test data in comparison with the ground truth (person), and that with the output from IK (blue skeleton). Because of the kinematic constraints of the body segments, despite reducing the number

of DOFs of spine and the lower extremities of limbs, the posture matches reasonably well with ground truths and IK output.

Figure 3: Comparison of the resulting postures with the ground truth and output from IK

3. Conclusion and Outlook

In this paper, we present the first step towards developing a hybrid approach involving full-body IK and deep learning for accurate estimation of physiologically feasible joint angles in real time, based on orientation information from 6 inertial measurement units (IMUs). Recent advances in bio-mechanical modeling is taken into account to constrain the output of the deep learning model. The approach has shown promising results. However, it must be extensively tested with further dataset, especially on the movements not only pertaining to construction work. In this direction, we are in the process of collecting raw dataset from the 6 IMU hardware system developed by us. Furthermore, as proposed in [6], diverse bi-RNNs are to be trained with variable window size and random input sequences, forming ensemble of models for estimating poses more accurately and robustly.

Acknowledgements

References

[1]   X. Yun, and E. R. Bachmann, Design, Implementation, and Experimental Results of a Quaternion-Based Kalman Filter for Human Body Motion Tracking, Robotics, IEEE Transactions on 22, 1216–1227 (2007).

[2]   M. Schepers, M. Giuberti, and G. Bellusci, Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing, (2018), DOI=10.13140/RG.2.2.22099.07205.

[3]   Y. Huang, M. Kaufmann, E. Aksan, M. Black, O. Hilliges, and G. Pons-Moll, Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time, ACM Trans. Graph. 37 (2018) 185:1–185:15., (2018) DOI=10.1145/3272127.3275108.

[4]   S. Schmid, K. A. Burkhart, B. T. Allaire, D. Grindle, and D. E. Anderson, Musculoskeletal full-body models including a detailed thoracolumbar spine for children and adolescents aged 6–18 years, Journal of Biomechanics 102, 109305, (2020).

[5]   S. Delp, F. Anderson, A. Arnold, P. Loan, A. Habib, C. John, E. Guendelman, and D. Thelen, OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement, Biomedical Engineering, IEEE Transactions on 54, 1940–1950, (2007).

[6]   D. Nagaraj, E. Schake, P. Leiner, and D. Werth, An RNN-Ensemble Approach for Real Time Human Pose Estimation from Sparse IMUs, Association for Computing Machinery, New York, NY, USA. DOI=10.1145/3378184.3378228, (2020).