MocapNET: Ensemble of SNN Encoders for 3D Human Pose Estimation in RGB Images

Ammar Qammaz [1,2#*] and Antonis A. Argyros [1,2]

[1] Computer Science Department, University of Crete

[2] Institute of Computer Science, FORTH

Presenting author: Ammar Qammaz, email: ammarkov@ics.forth.gr
* Corresponding author: Ammar Qammaz, email: ammarkov@ics.forth.gr

ABSTRACT

We present MocapNET [1], an ensemble of Self Normalizing Neural Network encoders that estimate the 3D human body pose based on 2D joint estimations extracted from monocular RGB images. MocapNET provides an efficient divide and conquer strategy for supervised learning. It outputs skeletal information directly into the BVH format [2], which can be rendered in real-time or imported without any additional processing in most popular 3D animation software. The proposed architecture achieves 3D human pose estimations at state of the art rates of 400Hz using only CPU processing.

Our approach utilizes training data from the Carnegie Melon University Motion Capture dataset [3]. By randomizing and perturbing the data we create millions of unique 3D poses that depict the recorded motions under a great variety of different views. We also propose a novel 2D data representation that encodes the relative structure of 2D point clouds in the form of a matrix that contains signed normalized distances. The combination of the above building blocks allows us to formulate a training scheme where we can train neural networks that learn to associate 2D projections of articulated human bodies to their full 3D configuration.

The independent encoder formulation we propose is a novel solution that inherently decomposes a 105 degree of freedom problem that is very hard to tackle to 105 one-dimensional problems which can be solved in parallel. It also provides an end-to-end neural network solution to this problem completely bypassing the need for inverse kinematics or iterative optimization techniques.

REFERENCES

[1] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. InAdvances in neural information processing systems,pages 971–980, 2017.

[2] Maddock Meredith, Steve Maddock, et al. Motion capture file formats explained.De-partment of Computer Science, University of Sheffield, 211:241–244, 2001.

[3] Carnegie Mellon University. CMU graphics lab motion capture database.http://mocap.cs.cmu.edu/, 2003. Accessed: 2017-06-01.s