Online Marker Labeling for Automatic Skeleton Tracking in Optical Motion Capture

Johannes Meyer Markus Kuderer Jörg Müller Wolfram Burgard

I. Introduction

Recently, methods to accurately capture the motion of people gained increasing interest for various applications like animation, orthopedics, and rehabilitation. In particular, passive marker based motion capture systems provide accurate 3D position information of retroreflective markers that are attached to people. Inferring the poses of the underlying skeleton structure requires labeling of the markers, i.e., unique identification of each observed marker. Using currently available tools for this task entails time-consuming manual labeling. We present a novel method for fully automated pose estimation of the skeleton configuration of a person equipped with markers. In particular, our algorithm assigns markers to limbs, determines the correct labeling of the passive markers in each frame, and estimates the skeleton pose using optimization methods. The experimental evaluation shows that our method outperforms existing methods for skeleton pose estimation and is able to run online.

II. Related Work

Inferring the structure and motion of the underlying skeleton based on the observation of attached markers has been studied widely. Ringer and Lasenby [4] present a method to assign markers to limbs and to determine their offsets, similar to our work. However, to estimate the pose of the limbs they assume that all markers are visible and labeled correctly in all frames. Since this assumption does not hold in practice, Aristidou and Lasenby [1] estimate the position and the orientation of the limbs in the presence of occluded markers. The visible markers are implicitly labeled from an active marker system, which is not the case in common passive marker systems like the one we use. Herda et al. [2] also consider occluded markers but need to manually associate each marker to a joint in the first frame. In our work, we propose to estimate the assignment of markers to limbs, label the markers in each frame and infer the pose of the limbs without any user assistance needed.

III. Online Labeling for Skeleton Tracking

In our skeleton tracking method, we apply a human skeleton model consisting of 14 segments and having 45 degrees of freedom. We initialize the skeleton structure and the marker labeling from the person standing in the T-Pose. Thereby, we obtain the size of the skeleton by assuming that the person is standing on the floor and one marker is placed on top of the head. We infer the initial posture, i.e., the heading of the torso and the orientation of the stretched arms and legs, by extracting the limbs as main axes out of the recorded marker set. Afterwards, we align the skeleton model to that posture through least-squares optimization to obtain the initial model configuration.

The key idea of our automatic labeling approach is the combination of nearest neighbor tracking (NNT), which can usually label a big proportion of all markers, and labeling the remaining markers based on the model configuration. In particular, our approach attempts to label the observed markers using NNT in a first step. Based on this labeling, we apply a least-squares optimization of the model configuration using g^2o [3]. This works well in practice, as usually the bulk of all observations is successfully labeled by the NNT and only the reappearing markers remain unlabeled. These remaining unlabeled markers are then associated based on the predicted marker position given the optimized model configuration using the Hungarian method. Subsequently, for each frame, the pose estimate is refined by a second optimization step given all labeled markers.

IV. Experimental Results

We evaluated the presented algorithm on a set of various motion capture recordings of different test subjects and marker sets, each recorded with 100 Hz frame rate. The quality of our marker labeling procedure was evaluated on a typical sequence with various movements, which was manually labeled to achieve ground truth data. Our approach was able to label 98.3% of all markers correctly. This shows its strong performance compared to the built-in technology in Motion Analysis Cortex, which correctly labeled only 89.6%. One of the main advantages of our approach is that it can adaptively skip the optimization step in frames in which the NNT successfully labels all markers. In practice, we are skipping the optimization in up to 10 successive frames so that our approach can process all data online and still achieves to label 96.3% of the observations correctly.

References

[1] A. Aristidou and J. Lasenby. Real-time marker prediction and CoR estimation in optical motion capture. *The Visual Computer*, 29, 2013.

[2] L. Herda, P. Fua, R. Plänkers, R. Boulic, and D. Thalmann. Using skeleton-based tracking to increase the reliability of optical motion capture. *Human Movement Science*, 20(3), 2001.

[3] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2011.

[4] M. Ringer and K. Lasenby. A procedure for automatically estimating model parameters in optical motion capture. In *Proc. of the British Machine Vision Conference*, 2002.

All authors are with the Department of Computer Science at the University of Freiburg, Germany. {meyerjo, kudererm, muellerj, burgard}@informatik.uni-freiburg.de