Fully Automatic Skeleton Tracking in Optical Motion Capture

Tobias Schubert Johannes Meyer Markus Kuderer Jörg Müller Wolfram Burgard

I. Motivation and related work

Recently, methods to accurately capture the motion of people gained increasing interest for variety of applications including interaction, animation, orthopedics, and rehabilitation. Compared to markerless approaches, marker-based methods are typically more accurate and are more robust against occlusions [8]. Major challenges in this context are to associate the observed markers with skeleton segments, to track markers between consecutive frames, and to estimate the underlying skeleton configuration for each frame. Existing solutions to this problem often assume fully labeled markers, which usually requires labor-intensive manual labeling.

In our previous work [7], we propose a fully automated method to initialize and track the skeleton configuration of humans from optical motion capture data. This method applies a flexible T-pose-based initialization that works with a wide range of marker placements without additional manual effort. To this end, we scale a standard human skeleton, based on Contini [3], to the person's size and align the skeleton to the person's limbs. After initialization we robustly estimate the skeleton configuration through least-squares optimization. Initialization methods without an underlying known skeleton structure were investigated by Ringer and Lasenby [9], Kirk et al. [5] and de Aguiar et al. [4]. These methods require a certain number of markers associated to each segment and an additional manual labeling step. Assuming known marker labels, several authors estimate the joint positions of the skeleton segments while taking into account skin movement artifacts [1, 2].

The contribution to this workshop are recent enhancements in skeleton tracking methods that build on our previous method [7]. First, using a large database of known skeleton configuration, we are able to mitigate the requirements during initialization. Instead of T-Pose initialization, we are able to initialize tracking during natural walking movements. Second, we update the association of markers to segments and the corresponding relative positions online during the tracking process in order to cope with initialization errors.

II. Skeleton Tracking

In this section we briefly recap our previous skeleton tracking method [7]. At each discrete time step t, we assume to receive a frame of data F_t that is a set of unlabeled 3D points $\{\mathbf{o}_{i,t}\}$. Each point is an observation of a marker $m \in M$, which we assume to be attached to a segment $s \in S$ of the skeleton. The goal of our method is to estimate the *skeleton configuration* C_t for each frame. We represent the skeleton configuration by the translation and the rotations of each segment. In general, the problem of skeleton tracking can be represented as a maximization problem, where the goal is to maximize the likelihood $\mathcal{L}(C_{1:t} \mid F_{1:t})$ of the skeleton configuration given the observations. In this representation, the association of markers to segments $\xi_t : M \to S$ and the labeling of the observations $\chi_t : F_t \to M$ are latent variables in our observation model. Thus, we marginalize over these latent variables to compute the likelihood

$$\mathcal{L}(C_{1:t} \mid F_{1:t}) = \sum_{\chi_{1:t},\xi_{1:t}} p(F_{1:t}, \xi_{1:t}, \chi_{1:t} \mid C_{1:t}). \quad (1)$$

We consider online tracking, therefore we perform a recursive estimation where we compute the most likely configuration C_t given the previous configurations $C_{1:t-1}$ as well as the previous marker labeling and segment associations. In a first step, we initialize C_0, ξ_0 and χ_0, assuming the person to stand in T-pose. As a contribution of this paper, we present a novel method that enables us to initialize a skeleton in natural walking motion without the T-pose requirement.

Since maximization of Eq. (1) is infeasible in practice, we split its computation in two steps in an EM-like fashion as outlined in Fig. 1. We estimate the most likely associations χ_t given the current frame F_t with fixed skeleton configuration C_t. Especially when a marker was occluded and reappears, the labeling based on the preceding frame is incomplete. We address this by associating the remaining observations to markers given the estimated skeleton configuration C_t. Given the resulting marker labeling, we then compute the most likely skeleton configuration C_t by means of optimization techniques [6]. In our previous work [7], we assumed a fixed association of markers to segments ξ_t and computed it once in an initialization phase. In this paper, we introduce a novel method to flexibly adjust this association online and demonstrate its improved tracking behavior.

III. Update of relative marker poses and segment association

Our experiments show that we can improve skeleton tracking by adjusting the association of markers to segments, and their relative poses, online during skeleton tracking. This step corrects errors that are introduced by inaccurate initialization. During online tracking, we consider the squared distance between the observations and the predicted marker positions, summed up over a number of previous frames. We then compute the position of each marker that minimizes

Jörg Müller is member of Department of Computer Science at the University of Southern California, USA. All other authors are members of the Department of Computer Science at the University of Freiburg, Germany.

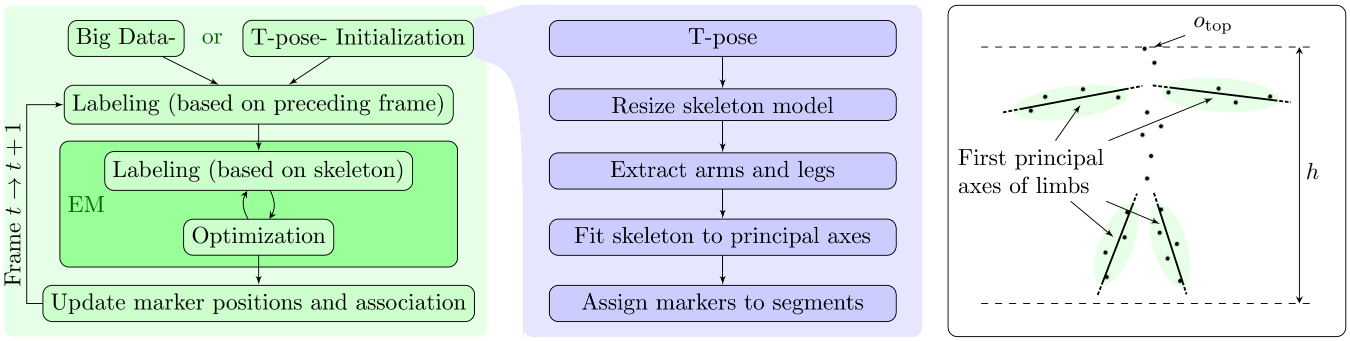

Fig. 1. Overview of the proposed method. On the right hand side we recap the T-pose initialization presented in Meyer et al. [7]. This T-pose initialization can be replaced by our Big Data approach. In each successive frame, most observed points are labeled based on the preceding frame by nearest neighbor association. Repeated optimization of the skeleton configuration and association based on the current skeleton estimate robustly labels the remaining points.

this distance error. In addition, we change the segment association, if this further reduces the mean distance error. This update step makes our method more robust against movements of the markers with respect to the segment, for example due to shifts of the markers that are attached to the person's clothing, or even changes in the segment the marker is attached to. This is especially beneficial for the robustness of long time tracking studies.

IV. Big Data Initialization

For the Big Data initialization we use a database of 50000 previously recorded natural skeleton configurations. For each of these configurations, we automatically generate 100 variants by rotation and scaling. Then we compute the corresponding association function and pick the 100 skeletons with least sum of the marker to segment distances. We start the previously described tracking algorithm with these configurations and subsequently erase hypotheses with high optimization errors.

V. Experimental Results

In this set of experiments, we present the effect of adjusting the marker associations and their relative poses online. We evaluated the presented algorithm on a set of four motion capture recordings of different test subjects and marker sets, each recorded with 100 Hz frame rate. Fig. 2 shows the performance λ_{50} of each data set, which is the mean squared value of the optimized cost function, i.e., a weighted sum of marker distances and joint limit costs, averaged over 50 frames. In this experiment, we used T-pose initialization. Using Big Data initialization, we obtain approximately the same performance results.

As a performance measure, we use the optimization error λ, which comprises the distance between observations and predicted marker positions as well as an additional term that represents natural joint limits. During online tracking,

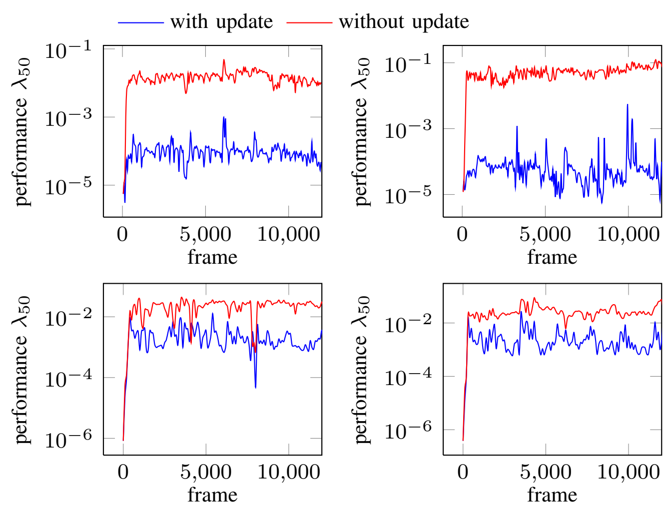

Fig. 2. Comparison of the performance of two different runs (coloumns) of two different persons (rows). The low performance values at the begining are due to the lack of movement in the T-pose initialization.

References

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

[2] P. Cerveri, A. Pedotti, and G. Ferrigno. Kinematical models to reduce the effect of skin artifacts on marker-based human motion estimation. *Journal of Biomechanics*, 38(11), 2005.

[3] R. Contini. Body segment parameters II. *Artificial Limbs*, 16 (1), 1972.

[4] E. de Aguiar, C. Theobalt, and H.-P. Seidel. Automatic learning of articulated skeletons from 3D marker trajectories. In *Second Int. Symposium on Advances in Visual Computing (ISVC), Part I*, 2006.

[5] A.G. Kirk, J.F. O'Brien, and D.A. Forsyth. Skeletal parameter estimation from optical motion capture data. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[6] 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.

[7] J. Meyer, M. Kuderer, J. Müller, and W. Burgard. Online marker labelimg for fully automatic skeleton tracking in optical motion capture. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2014.

[8] S. Obdrzalek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel. Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual Int. Conf. of the IEEE*, 2012.

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