2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Online Marker Labeling for Fully Automatic Skeleton Tracking in Optical Motion Capture

Johannes Meyer Markus Kuderer Jörg Müller Wolfram Burgard

Abstract— **Methods to accurately capture the motion of humans in motion capture systems from optical markers are important for a large variety of applications including animation, interaction, orthopedics, and rehabilitation. 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, especially when markers are temporally occluded during the movements. In this paper, we propose a fully automated method to initialize and track the skeleton configuration of humans from optical motion capture data without the need of any user intervention. Our method applies a flexible T-pose-based initialization that works with a wide range of marker placements, robustly estimates the skeleton configuration through least-squares optimization, and exploits the skeleton structure for fully automatic marker labeling. We demonstrate the capabilities of our approach for online skeleton tracking and show that our method outperforms solutions that are widely used and considered as state of the art.**

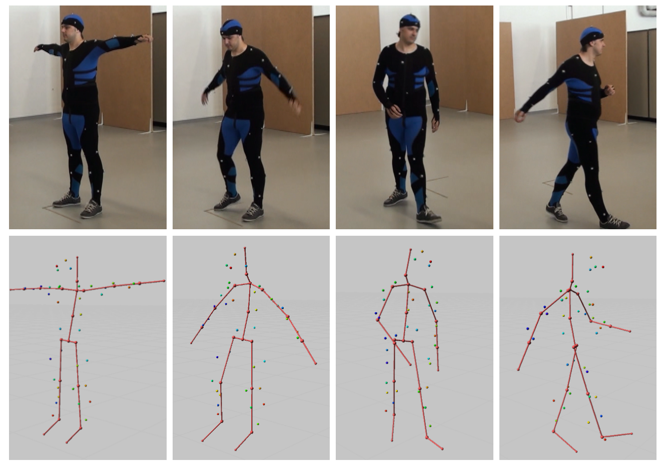


Fig. 1. Skeleton pose estimation. Top: Human with passive optical markers attached to the body. Bottom: Observed markers and the skeleton configuration as estimated by our fully automated method.

I. Introduction

Recently, methods to accurately capture the motion of people gained increasing interest for variety of applications including interaction, animation, orthopedics, and rehabilitation. The advantages of marker-based optical motion capture systems are that they provide automatic calibration procedures and precise position information about the markers at a high frame rate. Compared to markerless approaches, marker-based methods are typically more accurate and at the same time are more robust against occlusions [14]. Accordingly, they are perfectly suited to accurately capture even fast movements of people.

However, inferring the body pose in terms of the *skeleton configuration*, i.e., the global pose and the joint angles of the underlying skeleton, from raw 3D marker position data is a challenging task. Although the camera system provides accurate 3D marker positions, their association to the individual markers placed on the person is initially unknown. This data association problem is called *labeling* and is usually solved by analyzing the geometric structure of the set of detected 3D marker positions, which requires tedious and time-consuming manual work even with state-of-the-art software. Furthermore, markers are occasionally occluded by parts of the body or objects around the tracked person, which makes the labeling of the remaining and especially the re-appearing markers even more challenging. Finally, given the labeling of the markers, inferring the skeleton configuration requires to take into account that the markers are only attached to the skin or to clothes. Hence, the skin movement causes the markers to slightly move with respect to the bones during the activity, so that the skeleton configuration needs to be inferred in a robust way.

In this paper, we propose a novel fully automatic skeleton tracking technique. Our method is flexible and includes the labeling of passive markers, which can be placed on arbitrary positions on the human body, without any manual effort. We propose to use a parameterized standard human skeleton model, which we automatically adapt to humans of different size, and assume that each marker is attached to one of the limbs of this skeleton. Our approach to skeleton tracking jointly estimates the labeling of the observed markers and the skeleton configuration. In contrast to other approaches, our method exploits the information about the human skeleton during the labeling step and therefore provides a more informed data association.

Our approach initializes the skeleton tracking based on a T-pose executed by the person being tracked (see Fig. 1). Our method uses this initialization step to scale the skeleton to the person's size and aligns the skeleton to the person's limbs. During tracking, it then in an alternating fashion labels the observed markers and optimizes the skeleton configuration in an expectation-maximization-like procedure [6]. Thereby, it exploits both the marker positions of the most recent frame and the current skeleton configuration to obtain a consistent

All authors are with the Department of Computer Science at the University of Freiburg, Germany. This work has partially been partly supported by the German Research Foundation under grant number EXC 1086 and the Research Training Group 1103.

labeling and to reliably identify re-appearing markers that were temporally occluded.

We implemented our approach and evaluated it in extensive experiments. The results demonstrate the effectiveness and reliability of our approach and show that it outperforms *Cortex*, a state-of-the-art commercial solution for tracking articulated objects. Furthermore, we present results indicating that our method is highly efficient and enables online skeleton tracking on a standard desktop computer.

II. Related Work

The problem of inferring the structure and motion of objects based on the observation of markers attached to these objects has been studied intensively in the past. A common approach is to detect connected rigid bodies and to estimate the structure of the underlying skeleton in terms of the joints connecting these segments. Ringer and Lasenby [15] cluster observed 3D markers using the variance of the pairwise distance over all frames. Given the association of all markers to the segments, they determine their offsets, similar to our work. Similarly, de Aguiar et al. [5] and Kirk et al. [10] propose a method to cluster the observed markers to rigid bodies and estimate the center of rotation between the resulting limb segments.

Several authors estimate the joint rotation centers and the joint angles of a human skeleton from clusters of labeled 2D and 3D marker positions while taking into account skin movement artifacts [1, 2, 11]. Cerveri et al. [2] and Klous and Klous [11] even infer the skeleton structure and its configuration by calculating a statistics over all frames, which allows only offline processing of the captured data. In contrast, the method proposed by Aristidou and Lasenby [1] as well as our approach are able to track the skeleton online frame by frame without observing all frames first.

In many applications, for example when tracking humans, the underlying skeleton structure is known in advance. In particular, Contini [3] describes a standard human skeleton model, which we also use in the work described here. Zordan and Van Der Horst [18] propose a force based approach to fit such a human skeleton model to observed markers. Similar to our work, Xiang et al. [16] estimate the configuration of a human skeleton and utilize the knowledge about joint limits. One can either use medical studies [7] to determine these joint limits or learn them for individual humans by observing their motions [9].

All the approaches described above assume known marker labels to estimate the underlying skeleton model, or they solve the marker labeling independently of the skeleton estimation process. In contrast to these methods, we propose a marker labeling technique that exploits the knowledge about the current skeleton configuration. Similarly, Herda et al. [8] present a method to estimate the labeling of the markers in each frame and Yu et al. [17] aim to track multiple targets by fitting rigid bodies into the observed point cloud. Both approaches are able to deal with occluded markers but they require to manually label the markers in the first frame.

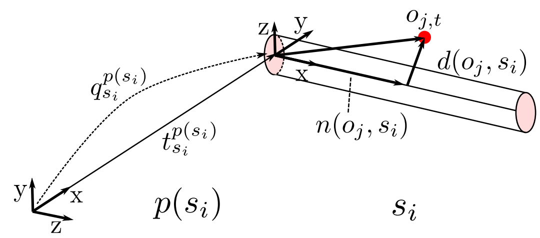


Fig. 3. This figure illustrates two linked joints s_i and its parent segment $p(s_i)$, their local coordinate systems and the transformation between the two segments. Furthermore, it illustrates an observed point and its normalized projection and distance to the segment s_i, which is used to compute the likelihood for the assignment of the observed marker to a segment.

In contrast, our approach performs labeling and skeleton pose estimation automatically without any user assistance.

In addition to the academic approaches described above, there are commercial solutions available for tracking articulated objects. We compare our work to *Cortex* developed by *Motion Analysis* [4] that provides online labeling of objects after a prior manual initialization and model training procedure. However, as opposed to our method, Cortex entirely separates the labeling from the skeleton tracking and therefore requires tedious manual post-processing to achieve results that are comparable to those obtained with our system.

III. Basics

The goal of this section is to give a formal problem definition and to present the foundations of online marker-based skeleton tracking.

A. Problem Definition

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 $\mathbf{o}_{i,t} \in F_t$ is the observed 3D position of a marker attached to one of the limbs of a person. The goal of our method is to estimate the *skeleton configuration* C for each frame, i.e., to estimate the global pose and the joint angles of the underlying skeleton. In particular, we use a human skeleton model that consists of 14 connected segments $s_i \in S$ having overall 45 degrees of freedom and that is based on medical data [3]. However, the 3D observations are subject to measurement noise and the markers are usually placed on the skin. During the movements of the body this induces non-deterministic variations of 3D position of the markers also with respect to the segments of the skeleton. Therefore we consider the full posterior probability $p(C \mid F)$ and estimate the maximum likelihood skeleton configuration of the skeleton configuration given the observations. This entails to *label* the observed points, i.e., to find the association function $\chi_t : F_t \to M$ that assigns each point to a marker label $m_j \in M$, which is one of the key challenges of the method presented in this paper.

B. Skeleton Model

In this work, we use a hierarchical skeleton model, where the pose of each segment s_i is defined by its local pose with respect to the coordinate system of its parent segment $p(s_i)$. In particular, we denote the quaternion that describes the

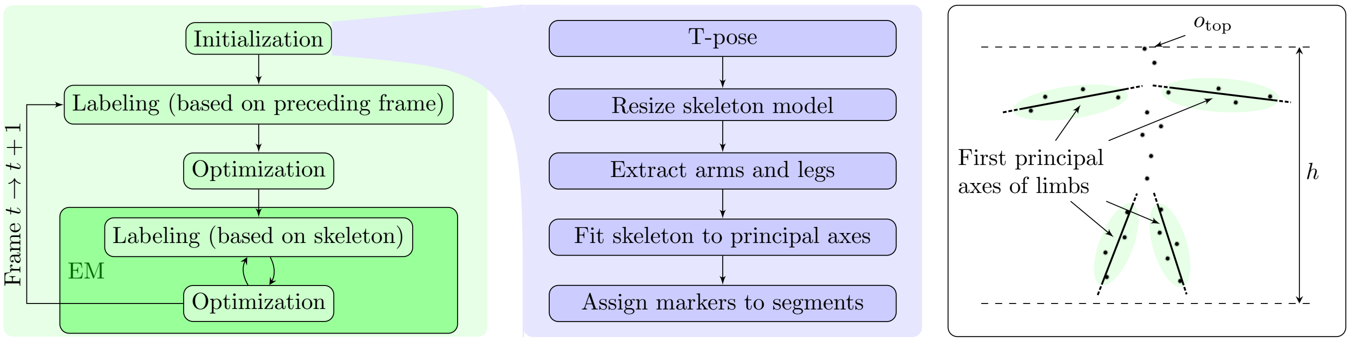


Fig. 2. Overview of the proposed method for automated skeleton estimation. Based on the observed markers in the first frame our method adjusts the skeleton model to the estimated height and fits the skeleton into the observed point cloud using the principal axes of arms and legs (middle and right). Furthermore, it initializes the marker labeling and their association to skeleton segments. Left: 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.

orientation of segment s_i in its parent's coordinate system at time step t as $\mathbf{q}_{s_i}^{p(s_i)}(t)$ and the position as $\mathbf{t}_{s_i}^{p(s_i)}(t)$, respectively. Fig. 3 shows an example of a segment s_i and its parent segment $p(s_i)$. Consequently, the global orientation $\mathbf{q}_{s_i}^{g}(t)$ of segment s_i at time step t is recursively given by

$$\mathbf{q}_{s_i}^{g}(t) = \mathbf{q}_{p(s_i)}^{g}(t) \odot \mathbf{q}_{s_i}^{p(s_i)}(t) \,, \tag{1}$$

where $\odot$ is the quaternion product. Similarly, the global position $\mathbf{t}_{s_i}^{g}(t)$ of segment s_i is recursively given by

$$\mathbf{t}_{s_i}^{g}(t) = \mathbf{t}_{p(s_i)}^{g}(t) + \mathbf{q}_{p(s_i)}^{g}(t) \odot \mathbf{t}_{s_i}^{p(s_i)}(t). \tag{2}$$

The position $\mathbf{t}_{s_0}^{p(s_i)}(t) = \mathbf{t}_{s_0}^{g}(t)$ and orientation $\mathbf{q}_{s_0}^{p(s_i)}(t) = \mathbf{q}_{s_0}^{g}(t)$ of the root segment, which is the hip segment in our experiments, describes the global pose of the skeleton. The positions $\mathbf{t}_{s_i}^{p(s_i)}(t)$ of the remaining segments correspond to their parent's lengths and are therefore static and specified by the skeleton model. As a result, the *skeleton configuration* C consists of 14 orientations that correspond to the joint angles and the global position of the skeleton. This results in overall $14 \cdot 3 + 3 = 45$ degrees of freedom of C.

We assume each marker m_j to be rigidly attached to one of the segments of the skeleton and introduce the function $\xi : M \to S$ that maps each marker to a segment. The local position of a marker m_j in the coordinate system of its corresponding segment is denoted as $\mathbf{p}_{m_j}^{\xi(m_j)}$. Thus, the position of marker m_j in the global coordinate system at time step t is given by

$$\mathbf{p}_{m_j}^{g}(t) = \mathbf{t}_{\xi(m_j)}^{g}(t) + \mathbf{q}_{\xi(m_j)}^{g}(t) \odot \mathbf{p}_{m_j}^{\xi(m_j)} \,. \tag{3}$$

IV. Probabilistic Marker-based Skeleton Tracking

In general skeleton tracking, one aims at estimating the skeleton configuration $C_{1:t}$ from time step 1 to t given the frame of unlabeled, noisy 3D observations of markers $F_{1:t} = \{\mathbf{o}_{i,1:t}\}$. In particular, we consider the likelihood $\mathcal{L}(C_{1:t} \mid F_{1:t})$ of the skeleton configuration given the observations. However, the association of markers to segments $\xi_{1:t}$ and the labeling of the observations $\chi_{1:t}$ are latent variables in our observation model. Thus, to compute the likelihood

$$\begin{aligned} \mathcal{L}(C_{1:t} \mid F_{1:t}) &= p(F_{1:t} \mid C_{1:t}) \\ &= \sum_{\chi_{1:t}, \xi_{1:t}} p(F_{1:t}, \xi_{1:t}, \chi_{1:t} \mid C_{1:t}) \end{aligned} \tag{4}$$

we marginalize over these latent variables. Since the maximization of Eq. (4) is infeasible in practice, we rely on the popular EM algorithm, which iteratively determines the maximum likelihood skeleton configuration

$$C_{1:t}^{*} = \underset{C_{1:t}}{\operatorname{argmax}} \sum_{\chi_{1:t}, \xi_{1:t}} p(F_{1:t}, \xi_{1:t}, \chi_{1:t} \mid C_{1:t}) \,. \tag{5}$$

In particular, the EM algorithm consists of two steps. First, the E-step computes the expectation value of the latent variables $\mathbb{E}(\chi_{1:t}, \xi_{1:t} \mid C_{1:t}^{(k)}, F_{1:t})$ given the configuration $C_{1:t}^{(k)}$. Second, the M-step computes the configuration $C_{1:t}^{(k+1)}$ that maximizes the likelihood under these expectations.

However, evaluating all possible marker and segment associations over all frames is not feasible in practice. Therefore, we propose the following approximations:

1) We assume the association ξ to be static and only compute it once in the initialization phase.
2) We consider online skeleton tracking and therefore recursively compute the likelihood. Hence, we assume the latent variables of the preceding frame to be known so that the recursive E-step reduces to computing the expectation value $\mathbb{E}(\chi_t \mid C_{1:t}, F_{1:t}, \chi_{1:t-1})$ and the recursive M-step computes $C_t^{(k+1)}$.
3) We apply the Hungarian method for optimal assignment to efficiently approximate the expectation value of the latent variables in the E-step. This maximum likelihood assignment technique is also known as hard EM in the literature.

Fig. 2 illustrates the resulting algorithm for online skeleton tracking. In the initialization phase, we determine the initial skeleton configuration as well as the association function ξ. For every incoming frame at time F_t, we first label the observations based on the labeling of the preceding frame and optimize the skeleton configuration to obtain an initial guess for the EM skeleton estimation. We then iteratively find the most likely labeling χ_t given the current skeleton configuration (E-step) and optimize the configuration C_t (M-step). In the following, we describe the individual steps of our algorithm for automatic skeleton tracking in detail.

V. T-Pose Initialization

In this section, we present the initialization step of our method to estimate the skeleton configuration of a person given the 3D position of observed markers that are attached to the body, as illustrated in Fig. 2.

The goal of the initialization step is to estimate the initial skeleton configuration C_0, to determine the initial labeling χ_0, i.e., to assign a marker label to each observed point in the first frame of observations F_0, and to determine ξ_0, i.e., to assign each marker label to one of the skeleton segments. We assume that the person initially stands in the T-pose, in which the person stands on the floor, stretches both arms and legs, and holds the arms approximately horizontally sidewards as shown in Fig. 1. Furthermore, we assume that there is one marker placed on top of the person's head. Fig. 2 (middle) illustrates the initialization process.

First, we obtain the height of the person by extracting the uppermost marker observation. Given the height, we scale the skeleton model, i.e., the local position vectors $\mathbf{t}_{s_i}^{p(s_i)}$ of all segments $s_i \in S \setminus \{s_0\}$, to match the size of the observed person. By considering the anatomy of humans, we identify the subset of points that belong to the legs and to the arms and compute their first principal axes, as illustrated in Fig. 2 (right). We align the skeleton model to these axes through least-squares optimization to obtain the initial skeleton configuration C_0. Note, that our initialization method is robust against deviations from the perfect T-pose due to the optimization-based alignment.

We initialize the bijective association function $\chi_0 : F_0 \to M$ by assigning each observed point $\mathbf{o}_{i,0}$ to the marker m_i. For each observed point $\mathbf{o}_j$, we compute the projection $n(\mathbf{o}_j, s_i)$ and distance $d(\mathbf{o}_j, s_i)$ to the corresponding segment $s_i = \chi_0(\mathbf{o}_j)$, normalized by the segment's length and radius, respectively. Thus, as illustrated in Fig. 3, points with the value of n and d in the interval $[0, 1]$ form a tube that has approximately the size of the corresponding body part. We define the likelihood of a point $\mathbf{o}_i$ to belong to a segment s_i as

$$\mathcal{L}(s_i \mid \mathbf{o}_i) = f(n(\mathbf{o}_j, s_i))\, f(d(\mathbf{o}_j, s_i)) \,, \tag{6}$$

where f is a function that has high values in the interval $[0, 1]$ and converges to 0 outside this interval. As a result, we assign high likelihood to points located inside the tube that represents the body part corresponding to segment s_i. In particular we choose

$$f(x) = \phi((x + \theta_1)\theta_2)\, \phi((1 - x + \theta_1)\theta_2) \tag{7}$$

with

$$\phi(x) = \frac{0.5x}{\sqrt{1 + x^2}} + 0.5. \tag{8}$$

The parameters θ_1 and θ_2 determine the convergence properties of f. Then, we assign each point to the segment it is most likely attached to:

$$\xi(m_j) = \operatorname*{argmax}_{s_j \in S} \mathcal{L}\left(s_j \mid \chi_0^{-1}(m_j)\right) \,. \tag{9}$$

Finally, we compute the initial estimate of the relative position $\mathbf{p}_{m_j}^{\xi(m_j)}$ of all markers with respect to their corresponding segments.

To resolve the ambiguity between the two possible headings of the person in T-pose, we maintain both hypotheses at first and dismiss one hypothesis as soon as it gets unlikely due to the joint limits of the skeleton (Sec. VI-C).

VI. Joint Labeling and Skeleton Estimation

Skeleton tracking aims at maximizing the likelihood of the skeleton configuration C_t given the unlabeled, noisy observations of markers in an online fashion. As illustrated in Fig. 2, we initialize the labeling based on the preceding frame in a nearest neighbor association. Given the initial skeleton configuration, we simultaneously perform the labeling of marker observations and the estimation of the skeleton configuration through an EM procedure, which alternately updates the labeling based on the skeleton configuration and estimates the skeleton configuration given the labeling.

A. Labeling based on the Preceding Frame

We initialize the labeling of every incoming frame of observations based on the labeling and the skeleton configuration of the preceding frame. Due to the high frame rate of motion capture systems usually most of the markers only move a short distance between two consecutive frames. In our approach, we initially label these markers through nearest neighbor association given the preceding frame.

In particular, we optimally associate the labeled observations of frame F_{t-1} to the unlabeled observations of frame F_t given the spacial distance as a cost function. Additionally, we define the threshold $\theta_{\text{max}}^{\text{PF}}$ as an upper limit for a valid association to avoid a wrong labeling in cases where one marker disappears and another marker appears at the same time. We determine the optimal one-to-one assignment using the Hungarian method [12]. This method assigns each observation of F_t to one observation of F_{t-1} such that the sum of the distances is minimized. The resulting labeling function is $\chi_t(\mathbf{o}_{i,t}) = \chi_{t-1}(\mathbf{o}_{j,t-1})$ for each pair $(\mathbf{o}_{i,t}, \mathbf{o}_{j,t-1})$ that has been assigned by the Hungarian method. Note, that we can adapt the thresholds $\theta_{\text{max}}^{\text{PF}}$ online for each marker individually based on statistics over the previous frames to account for a changing velocity of the markers.

B. Labeling based on the Skeleton Configuration

If a marker was occluded during some frames and reappears, or if it moves more than the threshold $\theta_{\text{max}}^{\text{PF}}$, we cannot label it based on the preceding frame. However, the current skeleton configuration C_t provides a prediction of the position of each marker, given by the global marker positions $\mathbf{p}_{m_j}^{g}(t)$. According to Sec. VI-A, we use the Hungarian method [12] to assign all remaining observations to markers that have not already been labeled based on the preceding frame. Here, we assign only observed points to markers in a radius of $\theta_{\text{max}}^{\text{S}}$ to be more robust against outliers.

C. Skeleton Estimation

Given the full or partial labeling χ_t of the marker observations of the current frame F_t and the relative position $\mathbf{p}_{m_j}^{\xi(m_j)}$ of each marker $m_j \in M$ with respect to its corresponding segment, we estimate the skeleton configuration C_t. In particular, we estimate the maximum likelihood skeleton configuration of $p(C_t \mid F_t)$ taking into account the uncertainty of observations and the joint limits of the skeleton. Due to the optical measurement process and skin effects, we assume Gaussian noise on the 3D observations of markers with respect to the segments of the skeleton. Thus the M-step translates to optimizing the skeleton configuration with respect to the mean squared error of the marker distances and the joint limits of the skeleton.

For each observation $\mathbf{o}_{i,t}$, we compute the reprojection, i.e., the estimated global position given C_t, of the marker assigned to this observation $\mathbf{p}^g_{\chi_t(\mathbf{o}_{i,t})}(t)$ by evaluating Eq. (3). We aim to find the configuration of the skeleton that minimizes the quadratic reprojection error, which is the quadratic distance between this reprojection and the actual observation. To improve the performance of the skeleton pose estimation, we include penalties for joint configurations that are outside of certain limits defined by considering the natural configurations of the skeleton, forcing for example the knee to move in one degree of freedom. Thus, the overall optimization function at time step t,

$$f(C_t) = \sum_{i \in I_t} \left\| \mathbf{p}^g_{\chi_t(\mathbf{o}_{i,t})}(t) - \mathbf{o}_{i,t} \right\|^2 + l(C_t)\,, \qquad (10)$$

sums over the set I_t of labeled marker observations and takes into account the joint limit cost function $l(C_t)$.

We determine the skeleton configuration C_t using the optimization framework g^{2}o [13], which provides modular and flexible gradient descent optimization. Particularly, we formulate the skeleton estimation problem as a graph, in which the skeleton configuration is contained in a variable node, the relative marker positions are contained in fixed nodes, and the components of the error function $f(C_t)$ are computed in unary and binary edges. The gradient of $f(C_t)$ with respect to the skeleton parameterization C_t is computed by means of numerical differentiation. This optimization procedure usually converges after a few iterations, even during fast movements and when initialized with the skeleton configuration C_{t-1} of the preceding frame.

VII. Experimental Results

We evaluated the presented algorithm on various motion capture recordings of different test subjects and marker sets. Each data set was recorded with a Motion Analysis motion capture system with ten Raptor-E cameras at 100 Hz frame rate. Our method was able to process data online at 100 Hz on an Intel® Core™ i7-2600K CPU with 3.40 GHz.

A. Initial association of markers to limbs

In a first set of experiments, we evaluated the initialization method (see Sec. V) that associates observed markers to the skeleton segments. Therefore, we compared the association of our method to manually labeled ground truth data. For markers that were located in the joint area between two segments, for example the elbow or the knee, we allowed the association to both segments. As parameter for the association of markers to limbs in Eq. (7) we chose $\theta_1 = 5$ and $\theta_2 = 0.2$. In 24 experiments, our approach associated 99.1% of overall 969 markers to the correct segment.

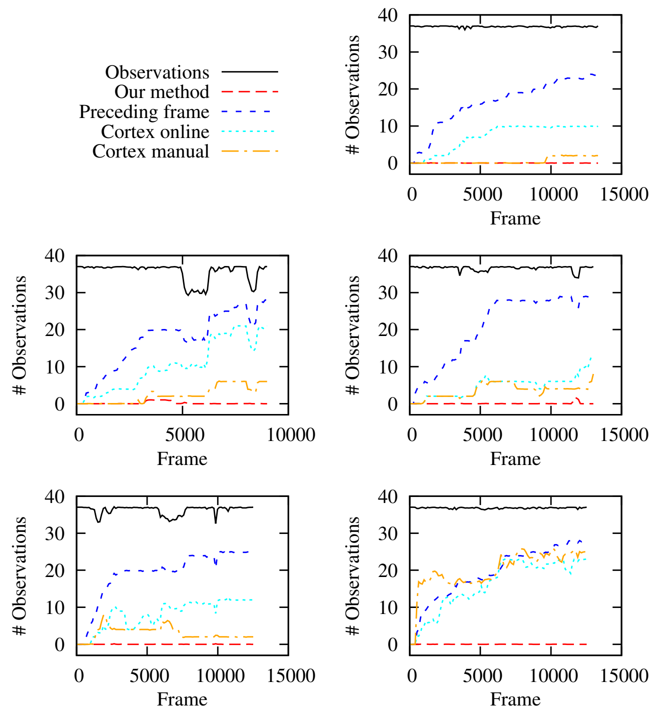


Fig. 4. The number of observed markers (solid) and the number of wrongly or not labeled markers (dashed) of our proposed method compared to the commercially available software Cortex and to a baseline method.

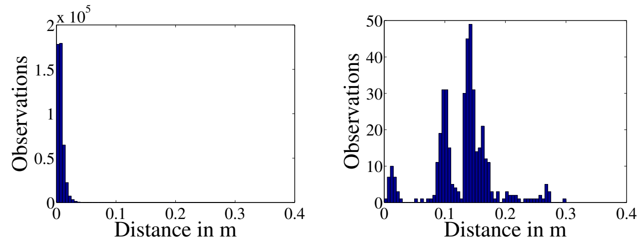
B. Performance of the marker labeling

A key challenge during the estimation of the skeleton pose of a human is to associate the observed points to the correct marker labels in each frame. To evaluate and compare the labeling performance of our method, we manually annotated five publicly available datasets[1] of overall 10 min motion capture data, including walking, sitting on a chair, stretching arms and legs, and jumping. Fig. 4 shows for each dataset and for our method as well as the following three alternatives the number of observed markers (*Observations*) and the number of markers that were either not or falsely labeled:

- *Preceding frame:* Our method with labeling based on the preceding frame only.
- *Cortex online:* Motion Analysis Cortex without manual post processing.
- *Cortex manual:* Motion Analysis Cortex where each dataset was manually post-processed for 1 h.

Overall, our approach was able to correctly label 99.6% of all markers whereas *Cortex online* correctly labeled only 79.8% of the data without manual post processing. After one hour

[1] http://www.informatik.uni-freiburg.de/~kudererm/



(a) correct associations (b) wrong associations

Fig. 5. Histogram of the movement of markers between two consecutive frames for correctly associated markers (a) and incorrectly associated markers (b) when assigned by the Hungarian method in a typical capture. Note the substantially distinct scales of the y-axes, which shows that the bulk of the markers are correctly associated based on the preceding frame.

of manual post processing for each dataset, *Cortex manual* correctly labeled 93.6% of the data. Obviously, labeling of markers based on the preceding frame only is not sufficient, as occluded markers are not correctly labeled when they reappear. Thus, our approach provides the best performance in fully automatic marker labeling for skeleton tracking.

C. Data Association Threshold Experiments

Labeling based on the preceding frame is sensitive to the parameter $\theta_{\text{max}}^{\text{PF}}$, which determines the maximum distance for optimal data association in the Hungarian method. We can adjust this parameter for each marker online using statistics over the previous frames. Thus, we can adapt the threshold to changing velocities of individual markers. To initialize the thresholds, we evaluated the movements of markers in a typical capture of natural movements. Fig. 5 shows a histogram of correctly (left) and incorrectly (right) associated markers based on the preceding frame with $\theta_{\text{max}}^{\text{PF}}$ set to ∞. Note, that the scales of the two figures deviate largely, which indicates that most of the markers are labeled correctly. All correctly associated markers moved less than 5 cm compared to the preceding frame. The incorrectly labeled observations are mostly due to occlusions. Therefore, in all of our experiments we chose the initial value $\theta_{\text{max}}^{\text{PF}} = 5\,\text{cm}$.

D. Ambiguity of the Initial Pose

As described above, our method initially tracks both potential headings of the person until the constraints in the joints and especially in the knee joints disambiguate the situation. To evaluate the robustness of our approach to determine the heading we evaluated its performance on 24 datasets with two different humans and six different marker setups. In all of the 24 datasets, our method chose the correct heading after 1.78 sec in average, with a standard deviation of 0.93 sec.

VIII. Conclusions

In this paper we presented a fully automatic method to estimate the underlying skeleton configuration of a human based on the position of markers perceived in a motion capture system and freely attached to the human. Our method uses the popular EM algorithm to compute the most likely skeleton configuration by estimating the unknown association of observations to marker labels in each frame. In contrast to existing approaches, which typically require tedious manual post processing, our method solves the estimation of the marker labeling without any user intervention. In an extensive set of experiments we demonstrate that our method outperforms even a commercially available and state-of-the-art method for skeleton pose estimation.

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] Motion Analysis Corp. *Cortex*, 2013. URL http://www.motionanalysis.com/html/industrial/cortex.html. Version 4.0.0.1387.

[5] 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.

[6] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1977.

[7] R. Hepp and R. Debrunner. *Orthopädisches Diagnostikum*. Thieme, 7 edition, 2004.

[8] 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.

[9] L. Herda, R. Urtasun, P. Fua, and A. Hanson. Automatic determination of shoulder joint limits using quaternion field boundaries. *Int. Journal of Robotics Research*, 22, 2003.

[10] 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.

[11] M. Klous and S. Klous. Marker-based reconstruction of the kinematics of a chain of segments: a new method that incorporates joint kinematic constraints. *Journal of Biomechanical Engineering*, 132(7), 2010.

[12] H.W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1), 1955.

[13] 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.

[14] 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.

[15] 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.

[16] Yujiang Xiang, Salam Rahmatalla, Jasbir S Arora, and Karim Abdel-Malek. Enhanced optimisation-based inverse kinematics methodology considering joint discomfort. *Int. Journal of Human Factors Modelling and Simulation*, 2(1), 2011.

[17] Q. Yu, Q Li, and Z. Deng. Online motion capture marker labeling for multiple interacting articulated targets. *Computer Graphics Forum*, 2007.

[18] V.B. Zordan and N.C. Van Der Horst. Mapping optical motion capture data to skeletal motion using a physical model. In *Proc. of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2003.