Automatic Initialization for Skeleton Tracking in Optical Motion Capture

Tobias Schubert Alexis Gkogkidis Tonio Ball Wolfram Burgard

Abstract— **The ability to track skeletal movements is important in a variety of applications including animation, biological studies and animal experiments. To detect even small movements, the method should provide exact accurate estimates. Besides that it should not impede the mammal in its motion. This motivates the usage of a passive optical motion capture system. Thereby the main challenges are the initialization, the association of the unlabeled markers to their corresponding segment also across the frames, and the estimation of the skeleton configuration. While many existing approaches can deal with the latter two problems, they typically need a specific pose for initialization. This is rather unpractical in the context of animal tracking and often requires a manual initialization process. In this paper, we present an approach to reliably track animals and humans in marker-based optical motion capture systems with freely attached markers that is also able to perform an automatic initialization without any pre- or post-processing of the data. To achieve this, our approach utilizes a large database of previously observed poses. We present our algorithm and its evaluation on real-world data sets with an animal and humans. The results demonstrate that our initialization method performs accurately for the most kind of initial poses and our tracking approach outperforms a popular fully automatic skeleton tracking method especially with respect to the smoothness of the motion.**

I. INTRODUCTION

In many applications it is important to track the skeletal movement of an animal or a person. Such applications include animation, robotics, biomechanics, and animal experiments. The former typically requires realistic motion sequences and in the latter one usually needs to detect small changes in the movements to identify changes, for example after training, medication, or brain stimulation. In all of them, however, accuracy is one of the principal requirements. In addition the used motion capture system should be applicable for animal tracking. In particular it should not impede the animal in motion. The optical motion capture system with passive markers complies these requirements. In comparison to markerless approaches, the marker-based ones are normally more robust against occlusions and provide more exact position information at higher frame rates [14]. Furthermore the most skeleton tracking methods are only able to track the skeletal motion of humans.

A major drawback of passive optical motion capture systems is that they provide only unassociated three-dimensional positions of the detected markers. For an automatic labeling procedure the occlusion of the markers over longer periods of time and the labeling of reappearing markers is a challenging problem. Besides the marker are only attached to the skin or coat of the tracked animal or human and therefore move during the motion. Recently Meyer et al. [13] presented a skeleton tracking method for humans with a fully automatic labeling of the markers. This method allows to place the markers at arbitrary positions onto the human. It then requires the human to go into a so-called T-pose to initialize the tracker. Especially in the context of animal tracking tasks such initialization processes are highly unpractical as it is typically not possible to force the animal to stand still in a specific pose. Alternative solutions, which are able to initialize the system with arbitrary poses, require manually labeled marker frames or a specific minimal number of marker attached to each modeled segment. Thus, a system that can deal with a limited number of markers and can initialize the tracker without manual labeling and without a fixed pose are highly desirable.

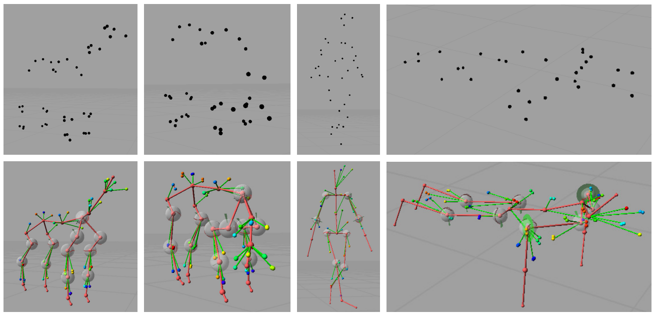

Fig. 1. Skeleton pose estimation. Top: Unlabeled marker positions given by the Motion Analysis software Cortex[5]. Bottom: Observed markers and the skeleton configuration obtained by our method.

In this paper we present two methods for fully automatic initialization of an arbitrary pose. Firstly, a fast method, which gets along with simple k-means clustering. Secondly, we use a big data approach for initialization. For the first method we use only one frame of unlabeled marker positions to find the initial skeleton. In the big data initialization we use one frame to reduce the data set to a small number of skeletons, e.g., one hundred in our experiments, and then we select the one with best tracking performance over given sequence of frames. In many existing skeleton tracking solutions, like the work of Meyer et al. [13] initialization errors worsen the entire skeleton tracking result. Therein, they compute the marker to segment association and the relative positions of the markers to the corresponding joints only once in the initialization step and they fix them during

---

All authors are with the University of Freiburg, Germany. This work has partly been supported by the German Research Foundation under grant EXC 1086. The recording of the sheep data sets was approved by the administrative council in Freiburg and we adhere the Protection of Animals Act and the “Tierschutz-Versuchstierordnung”.

the whole algorithm. We present an update step which uses the Kalman filter [8] to correct the relative positions and the marker position variance to change the marker to segment association. Moreover this models the marker movement along the skin or coat and leads to smoother motion trajectories. We apply our method to the skeleton tracking of a sheep, which is increasingly used as a large animal model in biological studies. In order to test small movement changes of a sheep caused by external or environmental factors the skeleton tracking algorithm should work online and provide results in a timely fashion. While we use a sheep as the only animal in our experiments (see Fig. 1), our method can be easily extended to other kind of mammals.

II. RELATED WORK

In the past many researchers investigated the problem of inferring the skeleton motion out of marker data. Whereas there also exist markerless approaches (see Baak et al. [2]), currently their accuracy is not high enough for certain applications such as medical studies, where one wants to detect small movement changes [7]. Several years ago, Schwarz et al. [16] presented a method for skeleton tracking using the depth data of a Microsoft Kinect, geodesic distances and optical flow. Their method provides an pose error of $70.1 \pm 9.1\,\mathrm{mm}$ with respect to a marker-based approach, which they use as ground truth in their experiments. Thereby this pose error is defined as the average of the Euclidean distances between the identical joints. This appears to be too inaccurate given the medical applications that we envision.

One of the main challenges for an accurate skeleton tracking is the initialization. Many existing solutions require a specific initialization pose and are therefore unpractical for animal experiments, see Meyer et al. [13]. Another approach is to detect the regions of rigid body segments first and then compute the position of the joints, which connect these segments, see Ringer and Lasenby [15], Kirk et al. [9], and de Aguiar et al. [6]. These methods require a certain number of markers associated to each segment and an additional manual labeling step, which is a time consuming and tedious work. Accordingly, a method which robustly works for different marker placements and counts is desirable. In the most cases the underlying skeleton structure of the mammal to be tracked is only approximately known. In their previous work, Meyer et al. [13] used a standard skeleton structure given by Contini [4], which they scale to the actual size of the tracked person. Zordan and Van Der Horst [18] used a physical model to fit a predefined skeleton structure into the marker cloud. Additionally, they use known marker positions for the initialization. Other authors estimate the joint positions of a human skeleton, while taking into account skin movement artifacts [1, 3, 10]. Here, the latter two methods use a statistics over all frames and thus only allow offline analysis of the data. In contrast, our approach and the work of Aristidou and Lasenby [1] compute the skeleton configuration online, without the knowledge of future data.

The work presented in this paper extends the work of Meyer et al. [13]. In the following we present our algorithm, see Fig. 2. Thereby the problem formulation, the labeling steps and the optimization part stay the same as in Meyer et al. [13]. Thus we give only a short sketch of this parts and refer the interested reader to their paper for more detailed information.

III. PROBLEM FORMULATION

Our input data is a set $F_{1:t}$ of frames of unlabeled three-dimensional observations $z_{i,t} \in F_t$ at equidistant discrete time steps t. Thereby, each observation is the three-dimensional position of one visible marker $m \in M$ attached to the skin, cloth, or coat of the object to be tracked. The task in the labeling step is to find the association function $\chi_t : F_t \to M$, which assigns each observation to a marker. Each marker is assumed to be attached to a segment $s \in S$ of the skeleton and we need to compute the corresponding segment association function $\zeta_t : M \to S$. The aim is to infer the skeleton configuration C_t for each time step t, which is the global pose of each modeled joint.

A. Skeleton Model

We use a predefined skeleton model, which we scale only once in the initialization step and afterwards we fix the lengths of the individual bones. Unfortunately, a detailed analysis of the relative bone lengths, similar to Contini's work for humans [4], is not available for animals. A good approximation can be obtained by measuring the bone length together with a proper analysis of the marker position data. For example, our sheep skeleton model consist of 17 segments. Although, it is not satisfactory to model each vertebra in the backbone, it suffices to give a good approximation of the problem, while keeping the optimization dimension low. The skeleton model is hierarchical, namely each skeleton configuration can be described uniquely by a three-dimensional rotation for each segment plus a three-dimensional translation vector of the root segment. All together, we describe the skeleton configuration of a sheep by $54 = 3 + 3 \cdot 17$ degrees of freedom. Thereby we represent the rotations using unit quaternions.

B. Probabilistic Skeleton Tracking

There is a lot of uncertainty hidden in the problem formulation, which we have to take into account. The observed positions $z_{i,t}$ are affected by measurement noise. The markers are attached to the skin, cloth or coat of the object and thus move slightly and non-deterministically with respect to the corresponding segment. This suggests a probabilistic problem formulation. More precisely, we want to get the most likely skeleton configurations $C^*_{1:t}$ given the marker observations $F_{1:t}$. This means in formulas that we have to solve

$$C^*_{1:t} = \arg\max_{C_{1:t}} P\left(C_{1:t} \mid F_{1:t}\right). \tag{1}$$

This probabilistic formulation comes with several hidden variables. These are the marker association function $\chi_t : F_t \to M$ and the segment association function $\zeta_t : M \to S$. Marginalization over all the hidden variables leads to

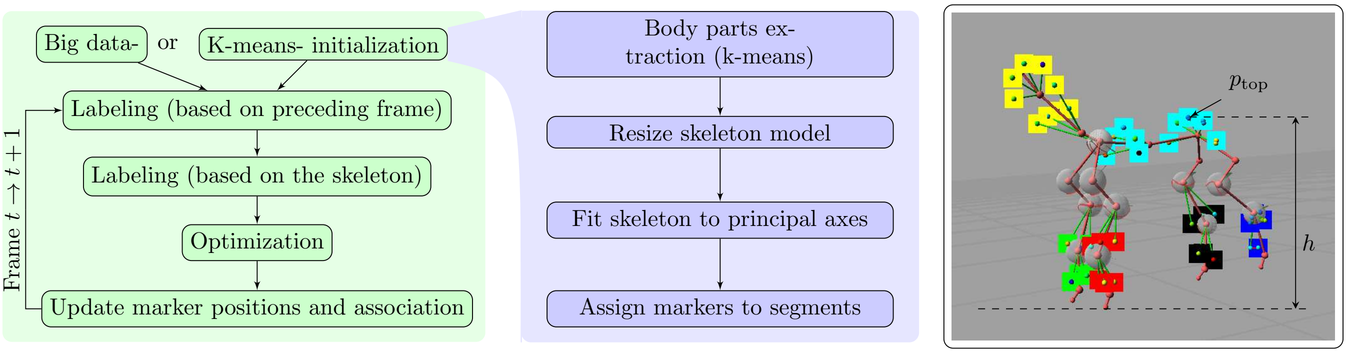

Fig. 2. Overview of the proposed method. On the right hand side we present the k-means initialization. In this paper we also present a big data approach to solve this problem. This approach assumes that there is a sufficiently dense skeleton data base. In each successive frame, most observed points are labeled based on the preceding frame by nearest neighbor association. Association based on the current skeleton estimate robustly labels the remaining points.

an optimization problem, which is infeasible in practice. Besides, we are aiming at an online skeleton tracking method and thus, we assume that the skeleton configuration and the association function only depend on past data. Meyer et al. [13] showed, that a good approximation for online skeleton tracking can be achieved as follows. Given the skeleton configurations $C_{1:t-1}$ and association functions $(\chi_{1:t-1}, \zeta_{1:t-1})$, we first compute the most likely association functions (χ_t, ζ_t) and then the most likely skeleton configuration C_t^* by optimizing $P(C_t \mid \chi_t, \zeta_t, F_t)$. As an extension to the work of Meyer et al. [13], we allow the markers to shift across the body and thus consider a time-dependent segment association ζ_t. We furthermore compute the initial skeleton configuration C_1 and the initial association functions (χ_1, ζ_1) in the initialization phase. State of the art skeleton tracking methods of humans require the tracked person to stand still in an initialization pose, as the T-pose [13], which, as already mentioned, is unpractical for animal tracking. In the following, we therefore present methods for the initialization of arbitrary poses.

IV. INITIALIZATION METHODS

In the initialization step we want to get the initial skeleton configuration C_1 given the initial unlabeled observations F_1. We obtain the bijective initial marker association function $\chi_1 : F_1 \to M$ by assigning each observed position $z_i \in F_1$ to a marker $m_i \in M$. As soon as we have the initial skeleton configuration, we get the initial relative position matrix $R_1 \in \mathbb{R}^{3\times|S|}$ of each marker to the segments base. In contrast to the work of Meyer et al. [13], we consider a time dependent relative position matrix R_t, so that we are able to get rid of initialization errors and to model the shift of markers over the coat or cloth more exactly. We will explain this in more detail in Section VI. With help of the initial skeleton configuration, we infer the initial segment association function ζ_1, which assigns each marker to a limb of the skeleton configuration. We do this in a similar fashion as in the work of Meyer et al. [13], who compute the likelihood $L(s_j \mid z_i)$ that a point z_i belongs to a segment s_j. Thus, it remains to find the initial skeleton configuration C_1. In the following we propose two initialization methods to solve this problem.

A. *Big Data Initialization*

For the first initialization method we assume that a sufficiently dense database of skeleton poses is available for the corresponding skeleton. In our current system, we use scaled and rotated versions and translate them into the set of observed positions. This leads to extensive initial skeleton configurations $\{C_{1,i}\}$, which we have to reduce appropriately to keep the algorithm fast. For each of these initial configurations we compute the segment association function $\{\zeta_{1,i}\}$, as in [13]. Next, we compute the number of segments to which at least one marker is attached. Since we assume a proper distribution of the markers over the whole body, we presume that a low number of occupied segments indicates a bad initialization, so that we can immediately dismiss most of the skeleton configurations. Next, we consider for each skeleton configuration $C_{1,i}$ the likelihood that the set of observed positions $\{z_j\}$ belongs to its associated segments $\{\zeta_{1,i}(\chi_1(z_j))\}$, i.e.,

$$\prod_{z_j \in F_1} L\left(\zeta_{1,i}\left(\chi_1(z_j)\right) \mid z_j\right), \tag{2}$$

where we compute the individual likelihood $L\left(\zeta_{1,i}\left(\chi_1(z_j)\right) \mid z_j\right)$ as Meyer et al. [13]. Next, we reduce the set $\{C_{1,i}\}$ to the one hundred skeletons with highest probability. Up to now, we only used the first frame F_1 of the data. We start our skeleton tracking algorithm with these one hundred skeletons and eliminate the ones with highest performance values in comparison to the others. The problem of this initialization method is that we need a database of all possible skeleton poses. One might get this by sampling the rotation matrices of the predefined skeleton model. Due to the high dimensionality of the sampling space, e.g., 54 for the sheep, this leads to a huge amount of skeletons in the database, where most of the data will never be attained. Another approach is to restrict the data set to realistic postures. While one can use the method of Meyer et al. [13] for the generation of human skeletons, a dense data set of realistic poses for animals is not available.

B. K-means Initialization

The second initialization method gives an adequate and fast initialization for many poses an animal can adopt. Moreover, we can use it to generate a data set of natural animal postures. As mentioned above, we fit a predefined skeleton model to the marker cloud. We only assume, that the sheep stands on its feet and that the head and each feet has attached at least one marker to it. Firstly, we scale this model to the size of the tracked object. In practical experiments we found out that the highest marker p_{top} on the trunk of the animal provides a robust feature for estimating the height h of the animal. To determine this marker, we separate the markers attached to the head and neck from the ones attached to the rest of the body. We consider the set of markers, which are lower than the center of mass, and use the k-means algorithm [17] to cluster them into four or five groups. If we obtain four groups, the head is at least as high as the trunk while it is below the trunk if we find five clusters. Given the cluster for the legs, we can identify the position of the head and also find the trunk (see right image of Fig. 2). Finally, we use least square optimization to fit the skeleton model to the clusters. This leads to the desired initial skeleton configuration C_1. We compute the segment association function ζ_1 with help of the likelihood $L\left(s_j \mid z_i\right)$.

V. SKELETON TRACKING

After we have initialized our method the subsequent challenge is to resolve the data association problem, see [13] for more details.

A. Labeling

One wants to estimate the most likely associations χ_t given the current frame F_t with fixed skeleton configuration C_t. Due to the high frame rate of the motion capture system the markers do not move much and one can associate the marker from χ_{t-1} using nearest neighbor association. Thereby we obtain the optimal assignment with help of the Hungarian method [11]. Especially when a marker reappears after an occlusion, 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. An alternative solution would be to use a multi-hypothesis tracking method, which we left for future research.

B. Optimization

Once we know the association functions χ_t and ζ_t, we can compute the most likely skeleton configuration. For this, we have to solve

$$\arg\max_{C_t} P\left(C_t \mid \chi_t, \zeta_t, F_t\right). \tag{3}$$

For each observation $z_{i,t} \in F_t$ and associated marker $\chi_t\left(z_{i,t}\right) \in M$ we can use the relative marker position R_t to obtain the predicted marker position $p_{i,t}\left(C_t\right) = p\left(z_{i,t}, \chi_t, \zeta_t, R_t, C_t\right)$ dependent on the skeleton configuration C_t. We try to find the skeleton configuration, which minimizes the distance between the predicted marker position and the observed position. In order to improve the performance of our method we include a quadratic joint limit cost term, which penalizes abnormal movements, e.g., those exceeding the joint limits of the involved joints. In sum, we realize the maximization in Eq. (3) by minimizing the optimization function

$$f(C_t) \quad = \sum_{z_{i,t} \in F_t} \left\| p_{i,t}\left(C_t\right) - z_{i,t} \right\|^2 + l\left(C_t\right), \tag{4}$$

where $l(C_t)$ stands for the quadratic joint limit cost term. Finally, we obtain the skeleton configuration C_t using the optimization framework g^2o [12]. After computing the skeleton configuration one can use the observed positions to update the relative positions of the markers.

VI. UPDATE OF RELATIVE MARKER POSES AND SEGMENT ASSOCIATION

Our experiments demonstrate that we can improve the skeleton tracking of Meyer et al. [13] by adjusting the association of markers to segments and their relative poses online during skeleton tracking. This step corrects errors introduced by an inaccurate initialization. A first naive approach would be to compute the relative position matrix R_t using the current skeleton and the marker positions $\{\chi_t^{-1}(m)\}_{m \in M}$. In order to incorporate the uncertainty in measurements and the skeleton estimation we use the Kalman Filter [8]. In our current implementation, we use a zero motion model, i.e., $R_t = R_{t-1} + \epsilon_t$, where ϵ_t is a zero-mean Gaussian with covariance r_t. As measurement we use for the marker m_i the relative position of the measurement, i.e., $\chi_t^{-1}(m_i)$ to the base of their corresponding segment $\zeta_t(m_i)$. Unfortunately, this approach cannot handle association errors and leads to a catastrophic labeling performance. In reality the relative positions of a marker slightly oscillates around a fixed position. To model this we introduce another relative position matrix $\widehat{R}_t$ and force R_t to stay near to it. In order to get rid of initialization errors and to model displacements of markers we again use a zero motion Kalman Filter with the same measurement, but a lower confidence assigned to it. Thus R_t models the fast movement changes of the marker along the skin, cloth or coat due to the muscle movement and $\widehat{R}_t$ guaranties a robust labeling performance and changes noticeable only due to shift of markers or initialization errors. In addition, we change the segment association ζ_t by investigation of the marker variance. This is especially beneficial for the robustness of long time tracking studies.

VII. EXPERIMENTAL RESULTS

The data sets we use in the experiments were recorded with a Motion Analysis motion capture system with ten Raptor-E cameras at 100 Hz for the human and 300 Hz for the sheep. We test our results on an Intel® Core™ i7-950 CPU with 3.07 GHz with eight threads. A video illustrating our method can be found at http://www2.informatik.uni-freiburg.de/ tobschub/research.html.

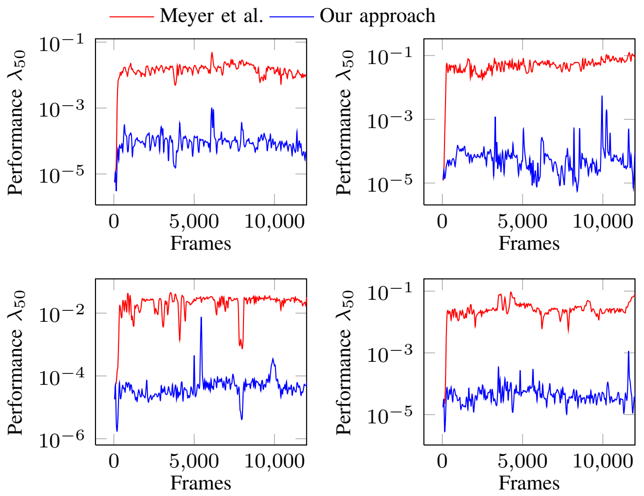

Fig. 3. Comparison of the performance of two different runs (columns) of two different persons (rows). The low performance values at the beginning are due to the lack of movement in the T-pose initialization.

A. Update Step

First of all, we present the effect of adjusting the marker to segment association and their relative poses online. For reasons of comparability we present the effect for human motion sequences. We evaluated the presented algorithm on a set of four motion capture recordings of different test subjects and marker sets. As a performance measure, we use the optimization error $\lambda_N = \frac{1}{\min\{t,N\}} \sum_{s=1}^{\min\{t,N\}} f(C_{t-s+1})$. Fig. 3 shows the performance λ_{50} of each data set. In this experiment, we use the T-pose initialization.

Next, we show that the update step leads to smoother trajectory estimates. To achieve this, we compare the smoothness of our method with the one of Meyer et al. [13]. In Fig. 4 we plot the velocity of the left elbow. We obtain similar results with all other joints and motion sequences in our data set.

B. Big Data Initialization

To analyze our big data initialization method, we apply it to the human motion sequences. Our database consists of about 170 000 skeleton configurations coming from human while walking, sitting, jumping, and even playing basketball. We created these skeletons poses with the skeleton tracking method of Meyer et al. [13] and regard them in the following as ground truth, since the real poses are not available. To speed up the initialization we only use every fifth skeleton. We test our approach on 103 arbitrary chosen initial frames out of the 170 000, where we do not use the skeletons two seconds before and after the frame we want to initialize. We use the mean distance of each joint to the ground truth as accuracy measure. Table I shows the number of frames for which the mean distance of the initialization to the ground truth is below the given threshold in the first row. For example in 81.55 % of all frames the assumed initial skeleton is less the 10 cm apart from the ground truth. This shows that the initialization method works in general. Its accuracy might be improved with the help of a dense and exhaustive database or using a particle filter with a more advanced motion model. In total, the initialization takes 388.6 ± 206.4 seconds, where it takes 204.6 ± 52.2 seconds to find the hundred starting hypotheses and 184 ± 194.9 seconds to find the hypotheses with the best performance. Up to a parallelization of the code we do not use any further code optimization.

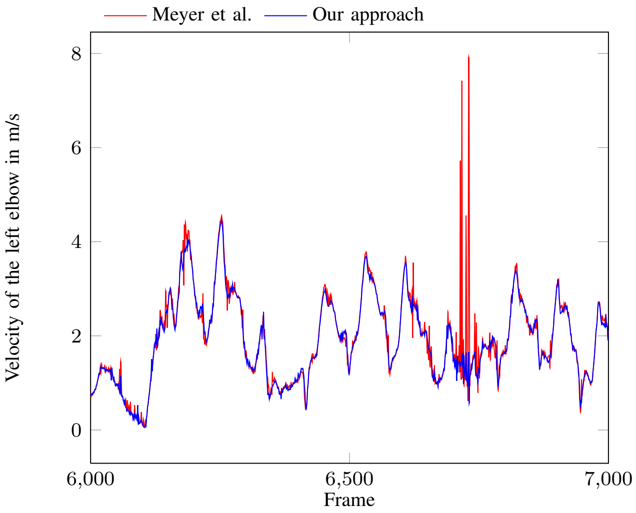

Fig. 4. Smoothness of the velocity of the left elbow obtained with our method in comparison to that of Meyer et al. [13].

TABLE I
BIG DATA INITIALIZATION

Data set	Poses	5cm	7cm	10cm	15cm
1	9	0	5	9	9
2	12	3	5	12	12
3	12	0	3	7	11
4	8	1	3	5	6
5	6	4	4	4	5
6	7	1	6	7	7
7	17	3	6	11	13
8	16	7	14	15	15
9	11	3	7	9	9
10	5	1	3	5	5
Total	103	23	56	84	92
%	100%	22.33%	54.37%	81.55%	89.32%

C. K-Means Initialization

We recorded the sheep data sets in a barn of $3 \times 5\,\mathrm{m}^2$ area. We located three cameras in three of the corners of the barn and the remaining seven cameras uniformly along the boundary of the barn at a height of about 2.5 m. We mounted hook and loop fasteners to the limbs, body and head of the shaven sheep and attached the passive marker on it. We recorded seven data sets of about four and a half minutes in total with one sheep and we tested our algorithm on these data sets.

To evaluate our initialization method we run it for each frame of our data sets. Next, we compute the mean distance of the feet joints to the corresponding ones of the ground truth skeleton. We again use the skeletons given by our skeleton method as ground truth. Table II contains the number of frames, where this distance is below the given threshold (see first row). For example, if we accept an initialization error of at most 5 cm our method is able to find an adequate initial skeleton in 48.27 % of all frames.

TABLE II

K-means initialization

Data set	Total Frames	3cm	5cm	7cm	10cm
1	18000	4947	14185	15319	16538
2	10000	566	4675	6392	9347
3	12000	2150	3051	5062	7070
4	7000	158	980	1891	3495
5	11000	1811	8137	9305	10440
6	10000	57	1250	2981	8251
7	15000	1827	7786	9641	11811
Total	83000	11516	40064	50591	66952
%	100%	13.87%	48.27%	60.95%	80.67%

VIII. Conclusion and Future Work

In this paper we presented a robust approach to tracking movements of people and animals in an optical motion capture system. In contrast to many other previous approaches, our method does not rely on a manual process or a specific pose for initializing the tracking process. Our methods has been implemented and tested on different data sets including human and sheep motion. In practical experiments we demonstrate the robustness of our approach. In our future work we want to apply our method to other kind of animals as well. Therefore, will extend our approach towards a fully automatic procedure for the determination of the lengths of the bones. Furthermore, we plan to use a particle filter with an improved motion model in the big data initialization. This will further improve the robustness of our approach.

IX. Acknowledgment

We want to thank Johannes Meyer, Markus Kuderer and Jörg Müller for fruitful discussions and suggestions.

References

[1] A. Aristidou and J. Lasenby. Real-time marker prediction and CoR estimation in optical motion capture. *The Visual Computer*, 29:7–26, 2013.

[2] A. Baak, M. Müller, G. Bharaj, H. P. Seidel, and C. Theobalt. A data-driven approach for real-time full body pose reconstruction from a depth camera. In *ICCV 2011*, pages 1092–1099, 2011.

[3] 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):2228 – 2236, 2005.

[4] R. Contini. Body segment parameters II. *Artificial Limbs*, 16(1):1–19, 1972.

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

[6] 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*, pages 485–494, 2006.

[7] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester. Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & Posture*, 39:1062–1068, 2014.

[8] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Fluids Engineering*, 82:35–45, 1960.

[9] 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)*, volume 2, pages 782–788, 2005.

[10] 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.

[11] H.W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2): 83–97, 1955.

[12] 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)*, pages 3607–3613, 2011.

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

[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*, pages 1188–1193, 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.*, pages 1–10, 2002.

[16] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab. Human skeleton tracking from depth data using geodesic distances and optical flow. *Image and Vision Computing*, 30(3):217–226, 2012.

[17] H. Steinhaus. Sur la division des corps matériels en parties. *Bull. Acad. Polon. Sci. Cl. III. 4*, pages 801–804, 1956.

[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*, pages 245–250, 2003.