

Online Motion Capture Marker Labeling for Multiple Interacting Articulated Targets

Qian Yu^{1,2}, Qing Li¹, and Zhigang Deng¹

¹ Computer Graphics and Interactive Media Lab, Department of Computer Science, University of Houston

² Institute of Robotics and Intelligent Systems, University of Southern California

Abstract

In this paper, we propose an online motion capture marker labeling approach for multiple interacting articulated targets. Given hundreds of unlabeled motion capture markers from multiple articulated targets that are interacting each other, our approach automatically labels these markers frame by frame, by fitting rigid bodies and exploiting trained structure and motion models. Advantages of our approach include: 1) our method is an online algorithm, which requires no user interaction once the algorithm starts. 2) Our method is more robust than traditional the closest point-based approaches by automatically imposing the structure and motion models. 3) Due to the use of the structure model which encodes the rigidity of each articulated body of captured targets, our method can recover missing markers robustly. Our approach is efficient and particularly suited for online computer animation and video game applications.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism - animation

1. Introduction

Motion capture has become one of the popularized methods for acquiring natural and subtle human motions in entertainment industry and academia. Currently optical motion capture systems, which require a number of optical markers attached to captured subjects, are widely used. Generally, a number of surrounding calibrated cameras are employed to reconstruct 3D positions of markers automatically. Based on how markers are identified from background, we can classify the optical motion capture systems into two types: active and passive. The active system uses LED markers that pulse in-sync with the cameras' digital shutters. More widely used passive systems work with the markers made of a strongly retro-reflective material and an illumination source co-located with each camera. The passive optical motion capture system can generate accurate 3D position data for the markers with possibility of a sporadic marker missing due to occlusions. However, in complex capture scenarios, e.g. capturing multiple interacting subjects, marker correspondences across frames often impose great challenges to the systems. In this paper, we propose an efficient online marker labeling approach for multiple articulated targets for

the passive optical motion capture system. Given hundreds (or even thousands) of unlabeled markers from the multiple captured targets that are interacting each other, our approach can automatically label these markers robustly. When the sporadic marker missing happens, our approach can reliably identify and estimate these missing markers. Figure 1 shows an input and an output of our approach.

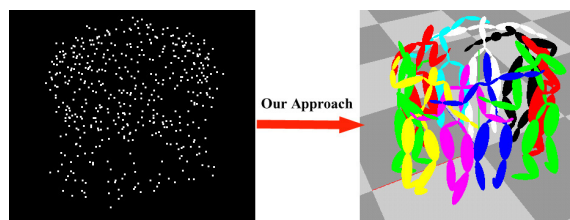


Figure 1: The left panel shows the snapshot of an input frame - ten persons are hugging. The right panel shows animation results after the markers are automatically labeled by our approach.

Comparing with the state of the art, our approach has the following distinctions: first, we aim to track and identify markers for multiple interacting targets. Tracking a sparse number of markers from a single articulated target is relatively simple because the markers often do not move much between consecutive frames due to a fairly high capture rate, e.g. 120 frames per second, as such, a closest point in the next frame is very likely to be the correct correspondence. However when an interaction among multiple targets exists (e.g., the dataset of ten people hugging in Fig. 1), the closest point criteria easily leads to wrong correspondences. Second, our algorithm is able to label markers online. The online feature is critical to various video games and computer animation applications. Third, our algorithm is tolerant to the marker missing. To our knowledge, this is the first work that can label motion capture markers online for the multiple interacting targets in complex capture scenarios.

In this approach, a captured subject is modeled as an articulated object composed of multiple rigid bodies that encode its spatial consistency. This spatial consistency provides important cues to evaluate the marker correspondences over time. Another important cue employed in our work is motion smoothness, which can significantly speed up the marker labeling process. Due to the noise of the motion capture and the model drifting, both the spatial consistency in the rigid body and the motion smoothness are needed to be updated continuously.

The remainder of this paper is organized as follows: in Section 2, we introduce previous and related work. In Section 3, we introduce the structure models and the motion models, how to fit the rigid bodies, and how to deal with the marker missing in our approach. In Section 4, we show various experiment results by our approach. Conclusions and discussion are presented in Section 5.

2. Related work

Motion capture (active or passive) is an important technique for many graphics applications. Unlike active marker and magnetic systems, passive systems do not require users to wear wires or electronic equipment and thus is more widely applied. However, the passive motion capture systems can only acquire 3D positions of the markers without temporal correspondences between the markers in different frames. Assuming correct motion marker correspondences are given, significant research efforts on the motion capture are focused on novel human motion synthesis [RCB98, PB02, DN06] and motion data reusing and editing [WP95, Gle98, AF02, KGP02, BLCD02, DCFN06]. Some other work is focused on estimating joints' positions and skeleton (topology and parameterization) for one articulated body [SPB*98, KJF05, LR02, ATSO6].

In order to recover correct marker correspondences over time, the proximity of a marker's 3D locations in two consecutive frames has been extensively exploited, which leads

to the closest point association assumption. This assumption is reasonable when a capture rate is high (such as 120 frame/second) and limited interactions happen between different articulated bodies or subjects [RL02, HFP*00, GF05, KJF05]. In the work of [KJF05, ATSO6], a post processing method was proposed to separate the 3D trajectories of different markers into different clusters using the spectral clustering [NJW01] for one target. Yan and Pollefeys [YP06] proposed a similar correspondence matching method based on the spectral clustering to recover the kinematics chain for a single human subject. In the work of [RL02], a multiple hypothesis tracking algorithm was proposed to deal with marker tracking for a single subject. Most of the above proposed methods employ the assumption that the marker proximity between two consecutive frames determines the marker correspondences from a single target.

In addition to the proximity, rigidity within each articulated body is also an important cue to estimate the correct correspondences, especially in the case of a heavy interaction between objects and in the case of a low capture rate. The rigidity is often used to estimate parameters of skeleton and joints [KJF05, ATSO6]. The work of [HSD05, LR02] combines the rigidity and the proximity for the online tracking for a single human body or for the case with little interaction between different subjects. However, it has not been demonstrated that these approaches can handle complex motion capture scenarios where multiple articulated targets (e.g. humans or animals) are captured simultaneously, and dense interactions among the targets happen.

3. Our Approach

A direct output obtained from an optical motion capture system consists of 3D positions of each marker over time. In our approach, we apply two cues (rigidity and motion cues) to infer marker correspondences across frames. An articulated captured subject is approximately regarded as a combination of rigid bodies, and a spatial cue lies in the rigidity of articulated bodies. Ideally, relative positions of the markers on a rigid body keep fixed over time, in other words, standard deviations of distances between the markers on the rigid body are expected to be close to zero. Based on this observation, our approach automatically constructs a set of rigid bodies from marker motion data in a training stage. Furthermore, the rigidity of articulated bodies (termed as the *structure model*) is updated over time.

The second cue is the motion smoothness of a marker in a short time span, which is useful to narrow down legitimate candidates for the marker in the next frame. The motion smoothness can be conceptually regarded as a characteristic for each marker. For example, the motion of a marker on a human torso is very different with that of a marker on a human foot. Time variant motion characteristics of each marker (termed as the *motion model*) is also updated over time.

Before our online marker labeling approach starts, the

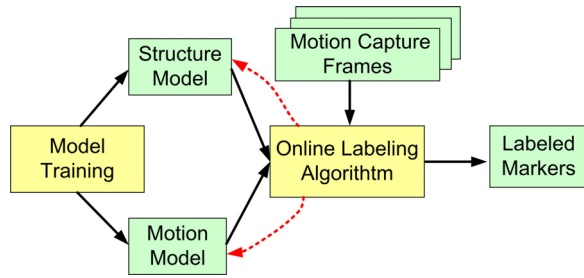


Figure 2: Schematic overview of this online marker labeling approach.

structure model (for each rigid body) and the motion model (for each marker) are initially trained. Once the training is done, the markers in a new motion capture frame are automatically labeled based on the trained structure models and the motion models that is updated at each time step. Figure 2 illustrates the procedure of our approach.

3.1. Model Training and Updating

In this section, we describe how the motion models and the structure models used in our approach are trained and updated over time. In the training stage, we assume no interaction among multiple captured targets happens, and typically motion capture systems operate at a high sampling rate, e.g. 120 Hz, we found that the heuristic closest point rule can be effectively applied to find accurate marker correspondences if there is no abrupt motion or interaction among subjects. In other words, assume the 3D position of a marker M_i at time step $t - 1$ is represented as M_i^{t-1} , the correspondence of M_i at time step t is the one that is the closest to M_i^{t-1} .

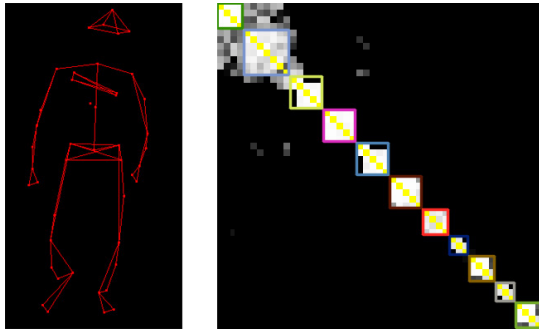


Figure 3: The left panel shows a skeleton of a captured subject. The right panel shows the standard deviation of marker-pair distances, where the brighter indicates the distance between two markers has smaller deviation, and different colors indicate different rigid bodies.

The rigidity of a rigid body can be encoded as the standard

deviations of marker-pair distances within the rigid body over time. Therefore, our method clusters the markers into a number of the rigid bodies, based on the standard deviations of the marker-pair distances. All motion frames in the training session are used to compute a distance matrix D and its standard deviation matrix A for each captured subject. D_{ij} is the mean of the distance between the i^{th} marker and the j^{th} marker, and A_{ij} is the standard deviation of the distance between the i^{th} marker and the j^{th} marker. Based on the standard deviation matrix, we select groups of markers which have small group-internal standard deviation to form a rigid body. Note that the number of rigid bodies is fixed for tracking once trained. The typical size (the number of the markers contained) of each rigid body is between three and eight in our experiments.

The standard deviation matrix of one captured subject in our dataset is visualized in Figure 3, where each rigid body is indicated by different colors, and the standard deviation is truncated to infinity if it is larger than a threshold. The distance matrix D will be used for validating the correspondences between consecutive frames, since the marker-pair distances within one rigid body should not be changed over a short time span. Both D and A are updated when a new frame is available as follows (Eq. 1 and 2):

$$D_{ij}^t = \left(\sum_{\tau=t-T+1}^t \sqrt{\|M_i^\tau - M_j^\tau\|^2} \right) / T \quad (1)$$

$$A_{ij}^t = \sigma(D_{ij}^t) \quad (2)$$

In the above equations, T is the size of a sliding window used to update the structure model.

Due to a high capture rate, the motion of each marker within a small time span can be modeled as the motion with constant velocity. We construct the motion model to reduce the number of legitimate candidates at next frame and filter out the possible noisy 3D markers (outliers) caused by camera occlusions. The case of marker missing is discussed in Section 3.3. The 3D position of the marker at the next frame (time t) is estimated by the Kalman filter as follows.

We denote x_k^t as a hidden state vector of a marker k at time t to be $[x, y, z, \dot{x}, \dot{y}, \dot{z}]$ (position and velocity in 3D). We consider a linear kinematic model of constant velocity dynamics:

$$x_k^{t+1} = A_k x_k^t + w_k \quad (3)$$

where A_k is the transition matrix, and we assume w_k to be a normal probability distribution, $w_k \sim N(0, Q_k)$. The observation $y_k^t = [x, y, z]$ contains the measurement of the marker's position in 3D. The observation model is represented as:

$$y_k^t = H_k x_k^t + v_k \quad (4)$$

where H is the measurement matrix and v_k is a normal probability distribution, $v_k \sim N(0, R_k)$.

Assume the estimated position of a marker, M_i , at time t is \hat{M}_i^t , then the markers at time t whose Mahalanobis distance to \hat{M}_i^t is smaller than the threshold is regarded as legitimate candidates. Using this motion model, the number of legitimate candidates for each marker is significantly reduced. Furthermore, to make this computation tractable, if the number of candidates for a marker is larger than a pre-defined upper limit, the algorithm will only take the best K candidates (K is a pre-defined constant). Actually, in our experiments, we found that the average number of legitimate candidates per marker automatically chosen by the motion model varies from two to four. Note that a marker at time t can be selected into the candidate sets of several markers. In Section 3.2, we describe how to solve this candidate conflict issue by fitting the rigid bodies.

3.2. Fitting rigid bodies

In this section, we describe how to assign correct labels $\{L_i\}_{i=1}^N$ to markers $\{M_i^t\}_{i=1}^N$ in a new coming frame (time is t), given the assigned labels in the previous frames $\{L_i^j\}_{i=1}^N$ (here j is from 1 to $t-1$), and the constructed structure models D and A . First, based on the motion model and the labels of the markers in the previous frame, we construct a candidate set for each marker M_i^t . Second, since the number of candidates for each marker is usually small and each rigid body is only composed of several markers (typically from three to eight markers), we enumerate and evaluate all possible assignments and select the optimum assignment which achieves the maximum score. Once one rigid body has been selected to be satisfied, the algorithm updates the other unsatisfied rigid bodies (usually the neighbors) whose candidate sets are changed, and update the structure models and the motion models. In this way, the algorithm iteratively satisfies the current best rigid body and reduces the number of unlabeled markers. It is summarized in Algorithm 1.

In this procedure, we introduce a measurement (score) to evaluate how a rigid body is fit with a marker assignment. The structure information lying in each rigid body provides the best evaluation of the fitting since the distance between different markers in the rigid body does not change much between two consecutive frame. In other words, if the correct correspondence is selected, the distance matrix among the markers in that rigid body of the new frame should be consistent with the distance matrix D learned in the structure model. Once we have one possible marker assignment for the rigid body r in the new frame, we can compute the new distance matrix D_r' of the rigid body r . The score of this assignment is computed as follows (Eq. 5):

$$C = \sum_{i,j} \exp\left(-\frac{(D_r(i,j) - D_r'(i,j))^2}{2A_r(i,j)^2}\right) / N \quad (5)$$

where N is the number of links in the rigid body. If a marker in this rigid body r has no assigned candidate, $D_r'(i,j)$ will be assigned a constant large value.

Algorithm 1 Fitting Rigid Bodies

Input: The labels at the previous frames, 3D positions of all markers in time t , the trained rigid body models D and A

Output: The new labels of all markers at current frame, the updated structure model, and the updated motion models

- 1: Based on the motion model (Eq.3 and Eq. 4), create the set of candidates for each marker M_i^t at time t .
 - 2: Based on D and A , motion markers are grouped into small rigid bodies $\{ur_i\}_{i=1}^k$.
 - 3: $Status(ur_i)$ ="unsatisfied" and $NeedRecomputed(ur_i)$ =TRUE for $i = 1 \dots k$.
 - 4: **while** at least one rigid body is "unsatisfied" **do**
 - 5: **for** each rigid body ur_i **do**
 - 6: **if** $Status(ur_i)$ is "unsatisfied" AND $NeedRecomputed(ur_i)$ is TRUE **then**
 - 7: search for the best marker-label assignment within ur_i using an exhaustive search (Eq. 5)
 - 8: Store the optimum assignment as $OptAssign(ur_i)$, and its score as $MaxScore(ur_i)$
 - 9: $NeedRecomputed(ur_i)$ = FALSE.
 - 10: **end if**
 - 11: **end for**
 - 12: $r = \text{argmax}(MaxScore(ur_i))$ for all ur_i satisfying $Status(ur_i)$ ="unsatisfied"
 - 13: $status(r)$ ="satisfied"
 - 14: **for** each "unsatisfied" rigid body ur_i **do**
 - 15: **if** the candidate set of any marker of ur_i has intersection with the markers in $OptAssign(r)$ **then**
 - 16: Update the candidate set of corresponding markers of ur_i by removing common markers in $OptAssign(r)$.
 - 17: $NeedRecomputed(ur_i)$ = TRUE.
 - 18: **end if**
 - 19: **end for**
 - 20: **end while**
 - 21: Update the structure model D and A (Eq. 1 and Eq. 2).
 - 22: Update the motion model (Eq.3 and Eq. 4).
-

3.3. Missing Marker Auto-Recovery

In a motion capture process, not every marker position at every time instant is tracked correctly: marker missing happens sporadically. This is usually caused by the occlusions in the optical motion capture procedure. As such, when a marker is missing, most likely this missing will continue for several frames, even up to tens of frames.

Our approach has a function of auto-recovering the missing markers, even the markers are missing for continuous tens of frames. Its recovery procedure works as follows: when the Fitting-Rigid-Bodies algorithm (Algorithm 1) is close to finish, remaining "unsatisfied" rigid bodies are regarded as the rigid bodies enclosing the missing markers. Due to the high sample rate of the motion capture system, the displacement vectors of one marker to the other markers in a rigid body in a small time span is approximately fixed. These displacement vectors can be acquired from the previous frames of the rigid body. Based on this observation, if a marker in a rigid body is missing, given the other

markers in the rigid body and the previously acquired displacement vectors, we compute the average 3D position as an estimation of the missing marker. When the number of the missing markers of the rigid body is larger, the accuracy of this recovery will decrease. In addition, our auto-recovery algorithm can also deal with very noisy markers and ghost markers. The noise in inlier (healthy) markers is modeled with the Gaussian model between the markers' distance, as such, our algorithm automatically regards these very noisy markers or ghost markers as missing markers, because these markers will not be selected into candidate sets. Figure 4 illustrates the missing marker auto-recovery in this approach.

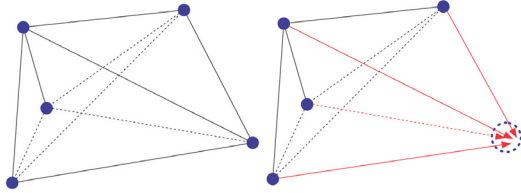


Figure 4: The left panel shows displacement vectors of a rigid body. In the right panel, red links indicate the average displacement vectors relative to the missing point. And the estimation of the missing marker is the average of predictions from known markers.

4. Results and Evaluations

The test data sets used in our experiments include two motion capture sequences. The first sequence with 1775 frame was captured when five captured persons dog piled in the middle, and each person was put 45 markers. The second sequence with 1467 frames was captured when ten persons hugged together, and each person was put 47 markers. Both sequences are captured at 120 frames/second and contain intensive interactions among multiple subjects, while marker sets used are different. Even with the aid of Vicon iQ motion capture processing package, ground-truth marker labels (of the test data sets) are obtained by painstakingly manual work, which is the major reason that we only have the limited test data sets for our experiments. Figure 5 shows snapshots of the test data sets.

The input data to our approach are motion capture frames that are composed of a set of unlabeled 3D points (except the first frame). Since our approach is designed as an online algorithm, the motion capture frames are fed into our approach frame by frame. It is assumed that the markers in the first frame for each person is correctly labeled, which is used to draw the skeleton of the human body. The marker tracking itself does not require any physical meaning of the markers. We further assume the captured human subjects have enough motion between different rigid bodies in the initial learning stage. Since the number of rigid bodies is fixed for the labeling once trained, the initial correspondence is also used

to validate the correctness of a rigid body clustering to avoid the case that no enough inter rigid body motion exists in the initial learning stage. The first 50 frames are used for the initial training. The assumption of no interaction among the subjects during the first 50 frames is satisfied for both sequences. And we use the closest point criteria to establish the correspondences in the first 50 frames to initialize the structure and motion models. If the closest point is not corresponding to the correct one, we manually adjust it.

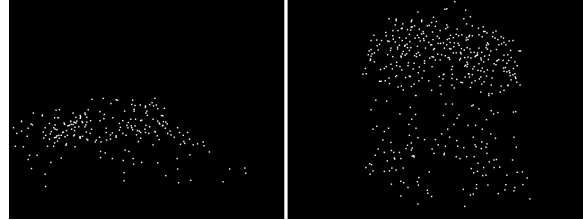


Figure 5: Snapshots of our datasets. The left shows a snapshot of the five-person dog pile dataset. The right shows a snapshot of the ten-persons hug dataset.

In order to evaluate our method, we compared our approach with the closest point based approach that assumes the correct correspondence is the closest point in the next frame. To make the situation more challenging, we down-sampled the capture rate from 120 frames/second to 60 frames/second and so on. The down-sampling is meaningful in practice, because usually 30 frame/second is enough to generate smooth animation and a low capture rate requires less data bandwidth from an optical capture system. Comparison results are shown in Figure 6. Due to the use of the structure and motion models, our method actually fits the new input data into the trained model instead of associating it with the previous frame. Thus our method is able to recover the correct labels even when there exist some errors in previous frame. In contrast, the closest point method will lose tracks once errors happen. As shown in Figure 6, the labeling results of our approach is significantly better than those of the closest point-based approach. This gap is even enlarged if down-sampling is applied. Figure 7 shows several comparison results.

In order to evaluate the accuracy of missing marker auto-recovery provided by this approach, we conducted the following experiments: several markers were randomly removed in the middle of motion capture sequences (up to twenty continuous frames), then this approach was applied to identify and recover these missing markers. We tested our approach in various ways by removing markers on different parts of human bodies and varying the number of continuous missing frames. Figure 8 shows recovery errors in terms of "error over max distance in the rigid body". The max distance in the rigid body is defined as the longest marker-pair

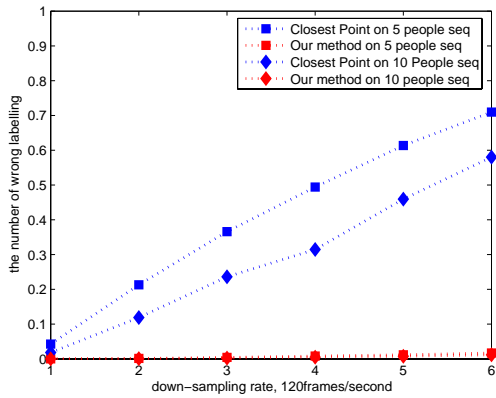


Figure 6: Comparison results of our approach with the closest point based approach with/without down-sampling.

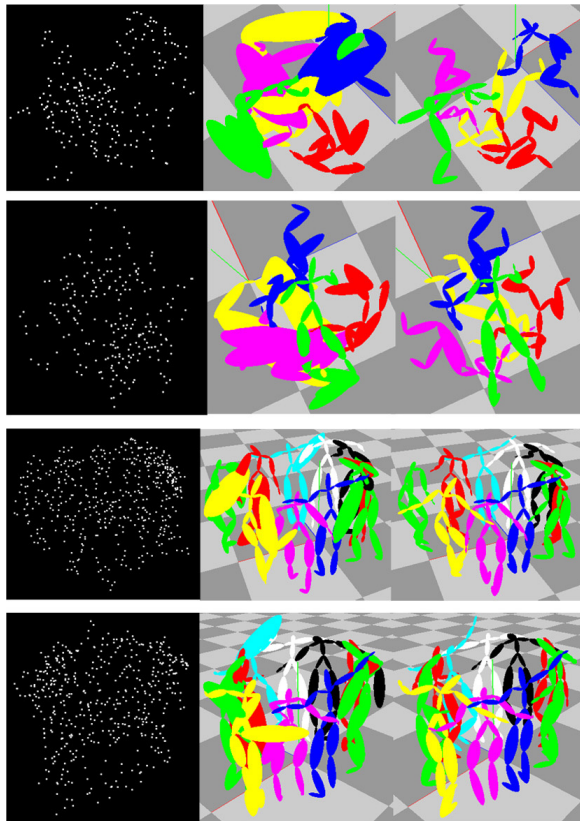


Figure 7: Frame comparison results of our approach with the closest point based approach. The left panel shows plotted 3D marker data, the middle panel shows the labeled results by the closest point approach, and the right panel shows the labeled results by our approach.

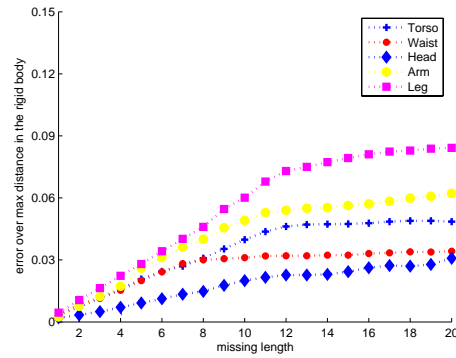


Figure 8: Missing marker recovery error of this approach, in terms of "error over max distance in the rigid body".

distance in the rigid body that the missing marker belongs to.

All testings were done on a PC Workstation (Intel Xeon CPU 3.00GHz, 4.00GB of RAM). On an average, our algorithm without any code optimization can process (label) one motion capture frame per captured subject in 227 millisecond.

5. Discussion and Conclusions

In this paper, we propose an online motion capture marker labeling approach for multiple articulated targets that are interacting each other. An articulated target is adaptively decomposed into a group of rigid bodies, then motion models (for each marker) and structure models (for each rigid body) are constructed and updated over time. After the models are constructed, this online approach can automatically label hundreds of incoming unlabeled markers frame by frame.

The advantages of our approach are as follows. First our method is an online algorithm which requires no user interaction once a tracking starts. Second, our method automatically imposes two important constraints of structure and motion for motion capture marker labeling. In other words, our method is designed to embed the new observations into the structure model instead of associating them with the previous markers. Thus our method is able to recover the correct correspondences even when certain measurement noises or errors exist in the previous frame as long as the correct one is covered by the candidate set. Third, due to the use of the structure models which encode the rigidity of each articulated body of captured targets, our method can recover the missing markers robustly.

Although our approach generates significantly better results than the closest point-based approach, our approach still has some limitations: first, our approach requires an initial learning stage, where motion capture sequences have no

interaction among the captured subjects. Second, if most of markers in a rigid body are missing in a motion capture sequence, our approach cannot recover those missing markers reliably. Third, the running efficiency of our approach can not achieve real-time performance on a single personal computer if the number of captured subjects increases. This is largely due the exhaustive search for matching each rigid body. To alleviate this issue, we need to improve the efficiency of the procedure of searching for the optimal correspondence of each rigid body. As a future work, we plan to look into how to efficiently impose local motion smoothness instead of using any global motion model for this speedup.

In the future, we plan to investigate whether introducing specific human motion models will improve this approach. Since the markers on a human subject are not moving independently, we can remove the candidates which conflict with reasonable human motion. Moreover, we are interested in looking into sample-based methods to avoid enumerate the labeling for each rigid body. This becomes crucial when we have a denser marker layout for each rigid body. We also plan to optimize the algorithms used in our approach to improve this online approach to achieve real-time performance. For example, introducing the GPU based parallel computation for each rigid body on each subject would be a promising way. Lastly, comprehensive evaluations on more complex motion capture scenarios would be necessary.

Acknowledgements

This research has been funded by new faculty research start-up fund at University of Houston. We also would like to thank Vicon for helping with the motion capture data and software support. Lastly, we thank Chang Yun for his proof-reading.

References

- [AF02] ARIKAN O., FORSYTH D. A.: Interactive motion generation from examples. In *ACM Transaction on Graphics* (2002), vol. 21, ACM Press, pp. 483–490.
- [ATS06] AGUIAR E. D., THEOBALT C., SEIDEL H.-P.: Automatic learning of articulated skeletons from 3d marker trajectories. In *Proc. of Int'l Symposium on Visual Computing (ISVC06)* (2006), pp. I: 485–494.
- [BLCD02] BREGLER C., LOEB L., CHUANG E., DESPANDE H.: Turing to the masters: Motion capturing cartoons. *ACM Transaction on Graphics* 21, 3 (2002), 399–407.
- [DCFN06] DENG Z., CHIANG P. Y., FOX P., NEUMANN U.: Animating blendshape faces by cross-mapping motion capture data. In *Proc. of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (2006), pp. 43–48.
- [DN06] DENG Z., NEUMANN U.: eface: Expressive facial animation synthesis and editing with phoneme-level controls. In *Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Vienna, Austria, 2006), Eurographics Association.
- [GF05] GUERRA-FILHO P. G.: Optical motion capture: Theory and implementation. *Journal of Theoretical and Applied Informatics (RITA)* 12, 2 (2005), 61–89.
- [Gle98] GLEICHER M.: Retargetting motion to new characters. In *Proc. of ACM SIGGRAPH '98* (1998), ACM Press, pp. 33–42.
- [HFP*00] HERDA L., FUA P., PLANKERS R., BOULIC R., THALMANN D.: Skeleton-based motion capture for robust reconstruction of human motion. In *CA '00: Proceedings of the Computer Animation* (2000), IEEE Computer Society.
- [HSD05] HORNING A., SAR-DESSAI S.: Self-calibrating optical motion tracking for articulated bodies. In *VR '05: Proceedings of IEEE Conference on Virtual Reality* (Washington, DC, USA, 2005), IEEE Computer Society, pp. 75–82.
- [KGP02] KOVAR L., GLEICHER M., PIGHIN F.: Motion graphs. *ACM Transaction on Graphics* 21, 3 (2002).
- [KJF05] KIRK A., J.F. O. B., FORSYTH D.: Skeletal parameter estimation from optical motion capture data. In *IEEE CVPR 2005* (2005).
- [LR02] LASENBY J., RINGER M.: A procedure for automatically estimating model parameters in optical motion capture. In *BMVC'02* (2002).
- [NJW01] NG A. Y., JORDAN M. I., WEISS Y.: On spectral clustering: Analysis and an algorithm. In *Proc. of Neural Info. Processing Systems (NIPS)* (2001).
- [PB02] PULLEN K., BREGLER C.: Motion capture assisted animation: texturing and synthesis. In *ACM Trans. on Graph.(Proc. of ACM SIGGRAPH'02)* (2002), ACM Press, pp. 501–508.
- [RCB98] ROSE C., COHEN M. F., BODENHEIMER B.: Verbs and adverbs: Multidimensional motion interpolation. *IEEE CG&A* 18, 5 (1998), 32–40.
- [RL02] RINGER M., LASENBY J.: Multiple hypothesis tracking for automatic optical motion capture. In *ECCV02* (2002), pp. 524–536.
- [SPB*98] SILAGHI M.-C., PLANKERS R., BOULIC R., FUA P., THALMANN D.: Local and global skeleton fitting techniques for optical motion capture. *Lecture Notes in Computer Science 1537* (1998), 26–40.
- [WP95] WITKIN A., POPOVIĆ Z.: Motion warping. In *Proc. of ACM SIGGRAPH '95* (1995), ACM Press, pp. 105–108.
- [YP06] YAN J., POLLEFEYS M.: Automatic kinematic chain building from feature trajectories of articulated objects. In *IEEE CVPR '06* (Washington, DC, USA, 2006), IEEE Computer Society, pp. 712–719.