Estimation of Missing Markers in Human Motion Capture

Guodong Liu Leonard McMillan
Department of Computer Science, University of North Carolina at Chapel Hill
{liug, mcmillan}@cs.unc.edu

Abstract

Motion capture is a prevalent technique for capturing and analyzing human articulations. A common problem encountered in motion capture is that some marker positions are often missing due to occlusions or ambiguities. Most methods for completing missing markers may quickly become ineffective and produce unsatisfactory results when a significant portion of markers are missing for extended periods of time. We propose a data-driven, piecewise linear modeling approach to missing marker estimation that is especially beneficial in this scenario. We model motion sequences of a training set with a hierarchy of low-dimensional local linear models characterized by the principal components. For a new sequence with missing markers, we use a pre-trained classifier to identify the most appropriate local linear model for each frame and then recover the missing markers by finding the least squares solutions based on the available marker positions and the principal components of the associated model. Our experimental results demonstrate that our method is efficient in recovering the full-body motion and is robust to heterogeneous motion data.

1. Introduction

Motion capture, or mocap, is a prevalent technique for capturing and analyzing human articulations. Mocap data have been widely used to animate computer graphics figures in motion pictures and video games. An optical mocap system utilizes video cameras to track the movements of a set of reflective markers that are strategically attached to the actor's body. The 3D marker positions are estimated via the triangulation from multiple cameras. A marker is considered missing if it is not visible to at least two cameras. A major cause of marker missing is that a marker is occluded by props, limbs, bodies or other markers. It is also not unusual that positions of some markers can be missing for a long period of time. Although many methods have been developed to handle missing marker problem and are already in use in commercial mocap systems, most procedures require manual intervention and are not satisfactory with diverse motions, high percentage of missing markers, and/or extended occlusions. In this paper, we propose a data-driven approach that is especially appealing under these situations.

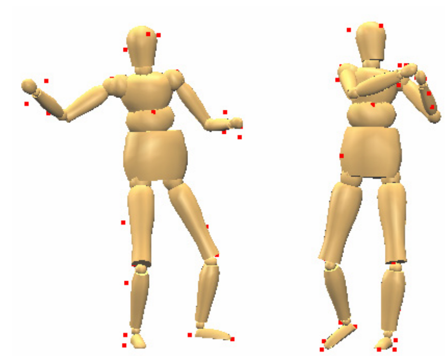


Figure 1. Example human poses constructed from the marker placements. The red dots represent the marker positions.

Coordinations as well as constraints between body parts are essential to produce valid human poses. As a result, the configuration space of human motions often exhibits low-dimensional local linearity. In particular, poses of a simple motion lie near a low-dimensional linear space with much fewer degrees of freedom than that of the original space. Such a local linear space can be described by the corresponding principal components obtained from principle component analysis (PCA). It is conceivable that, for a valid pose, its projection onto the principal component space can be determined by only a subset of markers.

Without assuming any skeleton model, we learn a global model and a hierarchy of local linear models from a training set that contains sufficiently representative motion sequences. We then take a two-step *coarse-to-fine* approach to estimating the missing marker positions of a new sequence. We use the global model to obtain a coarse estimation in the first step and refine the estimation via local linear models in the second step. Our method is very simple, fast and robust in recovering missing markers and estimating human motions on a *frame-by-frame* basis. Most importantly, our method allows different set of markers to be missing for a moderate-to-long period of time. In our experiments we demonstrate that our method can successfully estimate missing markers over a variety of motions from multiple subjects.

While many motion capture systems are used to acquire athletic or unusual motions for video games or motion pictures, we imagine that a system like ours would more likely be used in low-budget settings like capturing plausible avatars for VR or capturing interacting group behaviors where marker occlusion is more problematic.

2. Related work

Missing marker problem is commonly encountered in marker-based mocap systems. Interpolation methods [8, 16, 22] can effectively estimate the missing marker if a marker is missing for only a very short period of time, typically less than 0.5 second. Some commercial mocap systems [14, 20] also provide missing marker recovery solutions, more or less based on various interpolation techniques as well as the use of kinematic information with assumed skeleton models. Kalman filters [5, 21] have also been used to predict the missing markers in the current frame with the available temporal information. These methods, however, can quickly become ineffective and often require manual intervention when markers are missing for a long period of time, or missing from the very beginning.

Herda et al. [9] used a simplified human skeleton and marker positions from the immediately previous frames to predict the missing markers. If only a few isolated markers are missing over a long period of time, their positions can still be inferred from the neighboring markers which share the kinematic relations with the missing markers. However, the skeleton information must be known *a priori* in order to apply this method. Hornung and Sar-Dessai [11] proposed to utilize more markers in a mocap set up and assemble neighboring markers into a rigid clique. Markers in a clique have fixed inter-marker pairwise distances. When a marker is missing, its position can be recovered through the distance constraints imposed by the markers within the same clique. This approach may become uneasy to use when many markers are missing so that the clique is unable to be formed from the available markers.

There have been studies on exploiting the correlations among features in mocap data for motion synthesis based on a few control signals or motion estimation from incomplete information. Among them, Grochow and colleagues [7] developed a style-based inverse kinematics method in which a global nonlinear dimensionality reduction technique, Gaussian Process Latent Variable Model (GPLVM) [12], was used along with a pre-specified kinematic model. Their method worked well with a small homogenous data set, but might not be suitable for a large heterogeneous motion data set. The knowledge of skeleton information could present a challenge in some motion capture systems as well. Our method is a data-driven approach and requires no assumption about skeleton and domain information. Our piecewise linear approach also allows us to model data reasonably well even for a large heterogeneous motion data set.

Local linear models have been used to model data that show local linearity [2, 6, 10, 18]. Recently, Chai and Hodgins [4] presented a method that could quickly find the nearest neighbors of the current frame with only a few marker positions available. They constructed a local linear model from these neighbors and then reconstructed the full pose of the frame by conducting an optimization in the space constrained by the model. Their method was very effective in reconstructing high quality human motions from a few control signals, i.e. markers. However, a set of control signals and the skeleton information must be known before a mocap session so that the whole motion database can be scaled to fit with this particular actor and the scaled control signals can be computed and stored for the later neighbor search.

Liu et al. [13] proposed a piecewise linear approach to estimating human motions from a pre-selected set of most informative markers, i.e. principal markers. They characterized human motions as a collection of low-dimensional local linear models. Given a frame with only positions of principal markers, they classified the frame to the most appropriate local linear model and used the corresponding mapping function to recover the positions of the non-principal markers. Our method builds upon a common modeling infrastructure as theirs. However, this paper addresses a different problem -- filling missing markers, whereas they attempted to identify a smaller marker set based on sampling real motions and used them to recover the full-body motion. With their method, the principle marker set has to be pre-selected and fixed before the modeling process. This may not be realistic considering that any set of markers can be missing during different periods of time in a motion capture session. Our method, on the other hand, allows arbitrary markers to be missing for a considerable period of time while still being able to recover their positions using all the available marker positions.

3. Proposed method

There are two essential components in the process of missing marker estimation (Figure 2): modeling training data and estimation of missing markers for new sequences. Training data set contains sufficiently representative examples of motions. We take a two-stage modeling approach. In stage 1, we model motion data as a single global linear model, represented by the principal components. In stage 2, we model motion data in a refined fashion by a collection of local linear models, which together form a model hierarchy. Given a new sequence with missing data, we first fill in missing marker positions with the approximations derived from the global linear model. Next, for each frame with initially filled-in values, we identify the most appropriate local linear model via a classifier trained in the modeling stage and make a refined estimation for the missing markers by obtaining a least squares solution from the known markers and the principal components associated with the local linear model.

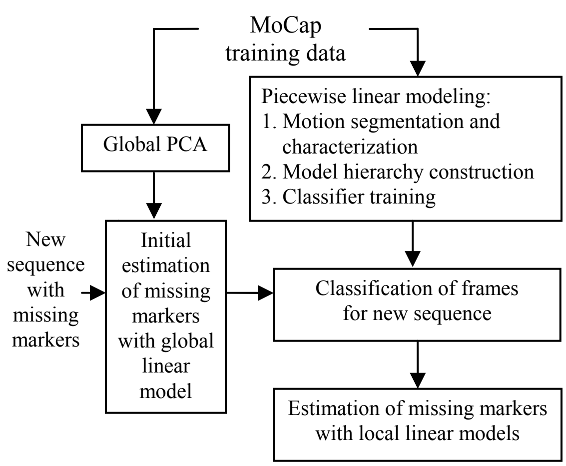


Figure 2. Motion data modeling and missing marker estimation process.

Throughout the paper, we treat each pose of motion sequences as a data point represented by a $3m$-dimensional column vector, $\mathbf{y} \in R^{3m}$, containing 3D marker positions (x, y and z coordinates) of m markers. Thus a motion sequence with N pose instances can be represented by a $3m \times N$ data matrix $\mathbf{Y}=[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N]$, where $\mathbf{y}_i$ is a column vectors of marker positions ($i=1,\ldots, N$).

3.1. Global linear modeling

Global PCA modeling is the first and the coarser stage of the modeling process. In this stage, a single linear model is constructed by applying PCA to the whole training set. We compute the principal components by performing Singular Value Decomposition (SVD) [15] on data matrix $\mathbf{Y}$ and form a $3m \times d$ matrix $\boldsymbol{P}$ with its d columns being the leading d principal components. The principal component matrix $\boldsymbol{P}$ as well as the mean vector will be used to calculate the initial estimates of the missing markers.

3.2. Piecewise linear modeling

Piecewise linear modeling has 3 components: motion segmentation and characterization; model hierarchy construction; and classifier training. We briefly describe each component as follows.

3.2.1. Motion segmentation and characterization. Clustering individual poses without considering temporal relationship among adjacent poses may lead to poor identification of local linear models because poses arising from different behaviors may be clustered together, while temporally adjacent frames may be clustered into different clusters. This clustering approach tends to result in unnecessary model transitions that may cause jerkiness during the motion reconstruction phase. To address this problem, we first segment motion sequences into short segments of single and simple behaviors using the probabilistic PCA (PPCA) segmentation method [1, 19]. Particularly, we assume a multivariate Gaussian distribution for poses coming from one distinct behavior and use PPCA technique to estimate the distribution. A local change in the distribution along a motion sequence forms a cutting point for the segmentation, and thus indicating a transition from one local linear model to another. After segmentation, a feature vector can be extracted from the mean vector and the covariance matrix of each motion segment as in Liu et al. [13].

3.2.2. Model hierarchy construction. We construct a hierarchy of local linear models by performing divisive clustering on feature vectors of segments with distance metric being Euclidean norm. All of the branches keep splitting recursively to two children based on K-means method until all the clusters at those branches satisfy a preset distance tolerance. We assign the same local linear model ID to the motion segments whose feature vectors are clustered into the same cluster and label each frame of a segment by its model ID. For each local linear model, we also compute PCA using the poses associated with that model. The resulting mean vector and the leading $\boldsymbol{e}$ principal components are later used to estimate missing marker positions from the positions of the known markers.

3.2.3. Classifier training. We use Random Forest [3] to classify frames of a new sequence into different local linear models that are extracted from the training set. Random Forest (RF) is a powerful classification tool that grows and combines decision trees into predictive models. The overall prediction is determined by voting over all the trees in the forest, with the class having the most votes being chosen. For each frame labeled with a model ID, we retrieve all of its marker values and use them as input variables for training the RF classifier.

3.3. Missing marker estimation

We take a two-step, coarse-to-fine approach to estimating the missing marker positions of new motion sequences. In the first step, we apply the global PCA model computed from the training set to obtain a coarser estimation of the missing markers. Then a frame, with missing markers filled with the initial estimates in step 1, is classified into the most appropriate local linear model via RF classifier. We next find the least squares solution to the projection of the frame onto the principal component space associated with the identified local linear model. Finally we transform them back to the original marker space to obtain the estimates of the missing marker positions.

3.3.1. Estimation with the global linear model. Given a frame with all the known markers correctly labeled and the rest markers missing, we compute a least squares approximation of the missing marker positions from the known marker positions using the global PCA model. We first retrieve the $3k\times 1$ position vector of k known markers, $\boldsymbol{f}$, and obtain a centered vector $\boldsymbol{s}$ by subtracting from $\boldsymbol{f}$ the corresponding part of mean vector of the global PCA model. Then we form a $3k\times d$ matrix $\boldsymbol{U}$ from the leading principle component matrix $\boldsymbol{P}$ by taking the entries corresponding to the known markers, and a $3(m-k)\times d$ matrix $\boldsymbol{V}$ by taking the remaining entries. Let a $d\times 1$ vector, $\boldsymbol{w}$, be the projection of a frame on the leading d principal component axes, we compute a least squares solution to $\boldsymbol{w}$ according to $\boldsymbol{U}\,\boldsymbol{w} = \boldsymbol{s}$ and estimate the $3(m-k)\times 1$ position vector of missing markers, $\boldsymbol{x}$, as $\boldsymbol{x} = \boldsymbol{V}\,\boldsymbol{w}$. The least squares solution to $\boldsymbol{w}$ is

$$\boldsymbol{w} = \boldsymbol{U}^T(\boldsymbol{U}\,\boldsymbol{U}^T)^{-1}\,\boldsymbol{s},$$

and thus

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{U}^T(\boldsymbol{U}\boldsymbol{U}^T)^{-1}\boldsymbol{s}.$$

Such an initial estimate of missing markers from one global model may be too coarse, especially when the database is a large heterogeneous motion data set where various types of motions are included. So it is crucial to use the local linear models to refine the estimation result of this step.

3.3.2. Estimation using the local linear models. Once we fill in the missing marker positions of a frame with the estimated values in step 1, we classify this updated frame, consisting of *full* marker positions, to the most appropriate local linear model by the Random Forest classifier. We then use the mean vector and the principal component matrix associated with the local linear model to estimate the missing marker positions through a least squares solution method as described in step 1.

When modeling time series data, an inherent drawback of piecewise linear modeling approach is temporal discontinuity at the transitions between two different linear models. We incorporate a mixture of local linear models associated with the previous poses to smooth out the jerkiness at the transitional poses. Let $\boldsymbol{s}_t$ be a centered pose vector containing the 3D positions of available markers at time t, we define matrices $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ associated with the ith model in the same way as matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in Section 3.3.1, we estimate the positions of the missing markers, $\boldsymbol{z}_t$, as

$$\boldsymbol{z}_t = \Sigma_i\, w_i\, \boldsymbol{V}_i\, \boldsymbol{U}^{\mathrm{T}}_i\;(\boldsymbol{U}_i\,\boldsymbol{U}^{\mathrm{T}}_i)^{-1}\;\boldsymbol{s}_t,$$

where $w_i = r_i\,/\,(h+1)$ is a weight for the ith model, r_i is the number of poses classified to the ith model among the prior h poses and current pose. Two complement parts of the leading principle component matrix, $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$, are eigenvector matrices corresponding to the known markers and missing markers respectively. Basically, we want to put more weights on the model that is favored by more of the current pose and its previous h poses. In our experiments, h=10-30 works well.

4. Experiments

4.1. Design

We evaluated our method with the mocap data from Carnegie Mellon University's Graphics Lab motion capture database available at http://mocap.cs.cmu.edu.

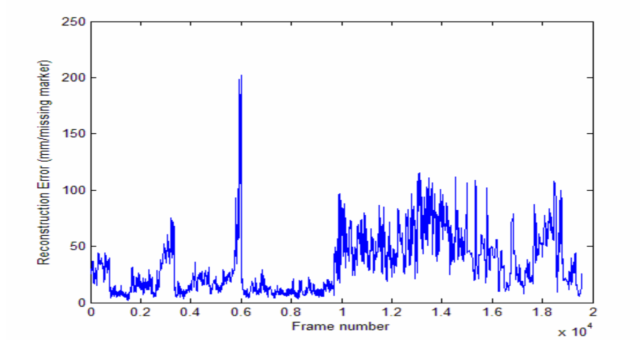
Typical motion data are captured in an absolute world coordinate frame. Our model, however, describes relative motions in a model-rooted frame. Therefore, a normalization step is required. CMU mocap system used a 41-marker setup. We choose the marker located at the C7 vertebrae as the origin. Our z-axis coincides with the z-axis (the *up* axis) of the original world coordinate system. We project a vector pointing from the left shoulder marker to the right shoulder marker onto the horizontal plane and use the projected vector as the x-axis. The cross product between z and x axes produces the y-axis.

We divided data into a training set and a testing set. The training set consisted of 132 sequences with total 151,882 frames collected from 21 subjects. We included a variety of motions (i.e., walking, running/jogging, golfing, soccer kicking, Salsa dancing, jumping, cartwheel, climbing steps), as well as different styles of the same motion from different subjects. Segmentation of the training sequences yielded 271 segments with length varying from 128 to 3,670 frames (mean: 560; standard deviation: 425; median: 440). Hierarchical clustering of segments according to their feature vectors produced 65 clusters, i.e., 65 local linear models. We retained the leading 15 principal components to approximate the poses of each local linear model.

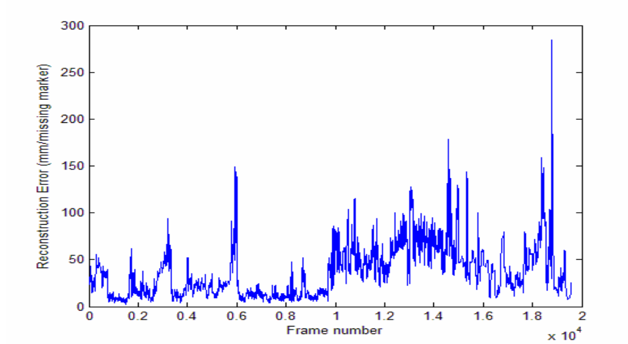
We used a testing set to validate our method. The testing set contained 28 sequences with 19,553 frames from 18 subjects. Among them, there were 9 walking sequences, 6 running, 5 golfing, 2 cartwheel, 2 Salsa dancing, 1 walking on uneven terrain, 1 running jump, 1 soccer kick, and 1 climbing three steps. None of the testing sequences was included in the training set. Four testing sequences, namely, 1 walking, 1 soccer kicking, 1 running and 1 golfing are from 4 new subjects who never appeared in the training set.

We assessed the performance of our method with different number of markers missing (i.e., 5, 8 10, 15 and 20) in each frame of the testing sequences. In each experiment, for every testing motion sequence, we randomly chose a fixed number of markers to be missing for a period of 1 second (120 frames). For example, in the first experiment we randomly chose markers 1, 15, 21, 32, 38 to be missing for the 1st second and 2, 7, 12, 29, 40 to be missing for the 2nd second.

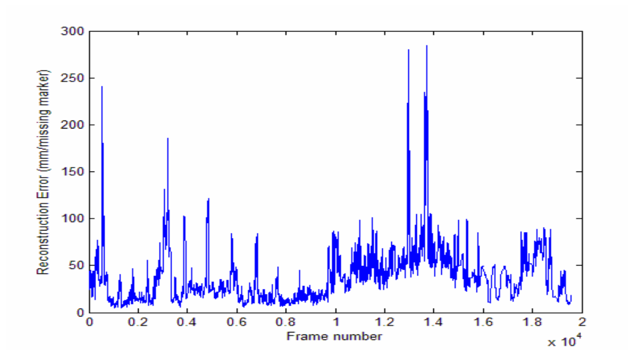
We also compared our method to spline interpolation in two scenarios where there were always 8 markers missing in the middle of a sequence with missing marker set being changed for every second. However, in the first scenario full marker positions were known at the two ends; while in the other scenario 8 randomly selected markers were also missing for a period of time either at one end or two ends of a sequence.



(a) 8 markers missing



(b) 15 markers missing



(c) 20 markers missing

Figure 3. Reconstruction errors.

4.2. Results

Figure 3 shows the frame-by-frame root mean squared errors (mm per missing marker) of each experimental result. It appeared that the reconstruction errors were minimal when there were 8 markers missing. Although the errors increased with increasing number of missing markers, the magnitude remained acceptable. Even when the number of missing markers reached 20, i.e., 50% of the total markers, the reconstructed motions as shown in the accompanying video were still plausible. We tested our method against various types of motion sequences that were not included in the training set. The results showed that our method was robust enough to produce reasonably good estimation to these heterogeneous motions. We also observed that the estimation results may appear unsatisfactory when many important markers were missing together, e.g., when all the markers on one leg or arm are missing.

We show in figure 4 the reconstruction results for each coordinate of two missing marker positions in a short segment of a motion sequence, using both the spline interpolation method and our method. When a marker was missing in the middle of the sequence, shown in the top panel of figure 4, our method recovered sufficient details missed by spline interpolation. In another example where a marker was also missing at the very beginning of the sequence, as shown in the last two panels in figure 4, spline extrapolation totally failed due to the fact that there was only support on one side of the missing frames. However, our method was still able to estimate the missing marker reasonably. The accompanying video also showed some examples where our method clearly outperformed the spline interpolation in full-body motion reconstruction.

Table 1. Running time (ms/frame) of estimation procedure when various numbers of markers are missing.

#of missing markers	Global PCA estimation	Local linear model classification	Local model estimation	Total time
5	0.52	15.36	1.90	17.78
8	0.50	14.98	1.94	17.42
10	0.50	15.32	1.95	17.77
15	0.46	15.25	1.93	17.64
20	0.42	15.61	1.95	17.98

One advantage of our method is that the estimation procedure can run very fast after off-line motion modeling with the training set. In Table 1, we show the distribution of the time spent at each key step. It only takes in average less than 18 milliseconds to estimate the missing marker positions per frame. That is over 50 frames per second, well above the typical interactive frame rate (i.e., 30 frames/second). It also appears that the total estimation time per frame remains about the same in spite of the increasing number of missing markers. We ran our experiments in Matlab V7 on a Dell Inspiron Laptop, with 1.4GHz CPU and 512M physical memory. A more powerful computer and more efficient code implementation may push the performance much higher.

5. Discussion and Future Work

We presented a piecewise linear modeling approach to estimating missing markers in human motion capture data and reconstructing plausible human motions. We learned the local linear models from a training set without prior knowledge of the human skeleton. We exploited the correlations among mocap markers to infer the missing marker positions from the positions of the known markers. The motion reconstruction process was efficient with no need to search in a database, or to estimate/calibrate a skeleton model. The experimental results demonstrated that our method can quickly generate plausible human motions on a *frame-by-frame* basis without any manual intervention.

Our method complements the interpolation-based methods in that it consistently produces reasonable estimation of missing markers even when the missing gap is so long such that the interpolation methods become ineffective or inapplicable. It also achieves better estimation than the spline interpolation methods when the frames at either ends of a sequence have missing markers. On the other hand, this non-linear cluster-based modeling method has limitations similar to other data-driven modeling approaches. It assumes that the training set is both representative of and adequately samples the data space. Moreover, its ability to extrapolate from the training data is more limited than its interpolation capabilities. The notion, however, of an interpolation system that is limited by its underlying model is not unique to data-driven methods. Kinematic models are also limited by the accuracy with which they accurately represent linkages and their motion ranges.

We limited out model to only marker positions and ignored velocities and accelerations. Using more information could improve our model, however, in our approach, adding more data may also increases the dimensionality of the problem, which implies the need for even more data, and we are already undersampled. This increases the likelihood of overtraining our model, thereby limiting its ability to generalize as discussed earlier. One of the strengths of our models is that it is very simple and fast. There are few parameters to be tweaked during modeling. Incorporating the velocity and even acceleration may make the model too complicated and slow down the estimation. Another reason that prevents us from using the acceleration and velocity is the concern of the accumulation of errors. Computing the acceleration and velocity requires the knowledge of the positions of the previous frames. However, some markers in the previous frames may have been missing and have to be estimated as well. So these frames may not be accurate enough to be used to estimate the current frame. We are concerned that this may in fact affect the estimation of the current frame due to the accumulation of errors. In our opinion, only the available marker positions of the current frame have the most accurate information since they are actually measured. So they should play more important roles in estimating the other marker positions.

There may be a normalization issue with the use of marker data. However, the performance of our method was not sensitive to the size variation in the subjects as shown in the experiments. Nevertheless, we attempt to conduct more experiments to verify this finding. In reality, a complete training data may not be available. So the EM algorithm-based PCA [17] may be potentially useful to handle the training set with slight to moderate missing. Further study is needed in this respect.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. The data used in this project was obtained from CMU's motion capture database (mocap.cs.cmu.edu), which was created with support from NSF EIA-0196217.

References

[1] J. Barbič, A. Safonova, J. Pan, C. Faloutsos, J. Hodgins, and N. Pollard, "Segmenting motion capture data into distinct behaviors," *Graphics Interface*, 2004, pp. 185–194.

[2] C. Bregler, and S. Omohundro, "Nonlinear image interpolation using manifold learning", In *Advances in Neural Information Processing Systems*, 1995, pp. 43–50.

[3] L. Breiman, "Random forest", Machine Learning, 2001, 45 (1), pp. 5–32.

[4] J. Chai, and J. Hodgins, "Performance animation from low-dimensional control signals", *ACM Trans. Graph.* 2005, 24(3), pp. 686–696.

[5] K. Dorfmuller-Ulhaas. Robust optical user motion tracking using a Kalman filter. In 10th ACM Symposium on Virtual Reality Software and Technology, 2003.

[6] K. Fukunaga, and D. Olsen, "An algorithm for finding intrinsic dimensionality of data", In *IEEE Transactions on Computers*, 1971, C-20: pp. 176–183.

[7] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popovic, "Style-based inverse kinematics", In *Proceedings of SIGGRAPH,* 2004, pp. 522–531.

[8] S. Guo and J. Roberg, "A high-level mechanism for human locomotion based on parametric frame space interpolation", *Computer Animation and Simulation*, 1996, pp. 95–107.

[9] L. Herda, P. Fua, R. Plankers, R. Boulic, and D. Thalmann, "Skeleton-Based Motion Capture for Robust Reconstruction of Human Motion", *Computer Animation* 2000, pp. 77–85.

[10] G. Hinton, M. Revow, and P. Dayan, "Recognizing handwriting digits using mixtures of linear models", In advances in Neural Information Processing Systems 7, 1995, pp. 1015–1022.

[11] A. Hornung, S. Sar-Dessai, " Self-Calibrating Optical Motion Tracking for Articulated Bodies", In *Proceedings of the 2005 IEEE Conference on Virtual Reality*, 2005, pp. 75–82.

[12] N. D. Lawrence, "Gaussian process latent variable models for visualization of high dimensional data", In *Advances in Neural Information Processing Systems 16*, 2004, pp. 329–336.

[13] G. Liu, J. Zhang, W. Wang, and L. Mcmillan, "Human Motion Estimation from a Reduced Marker Set", In Proceedings of *ACM SIGGRAPH Symposium on Interactive 3D Graphics* (I3D), 2006, pp. 35–42.

[14] Peak Performance Inc. http://www.peakperform.com.

[15] W. M. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes: The art of scientific computing*, New York, NY: Cambridge University Press, 1986.

[16] C. Rose, M. Cohen, and B. Bodenheimer, "Verbs and adverbs: Multidimensional motion interpolation", *IEEE Computer Graphics and Applications*, 18(5):32–40, 1998.

[17] S. Roweis, "EM algorithms for PCA and SPCA", In *Proceedings of Neural Information Processing Systems*, 1998, 10, pp. 626–632. The MIT Press.

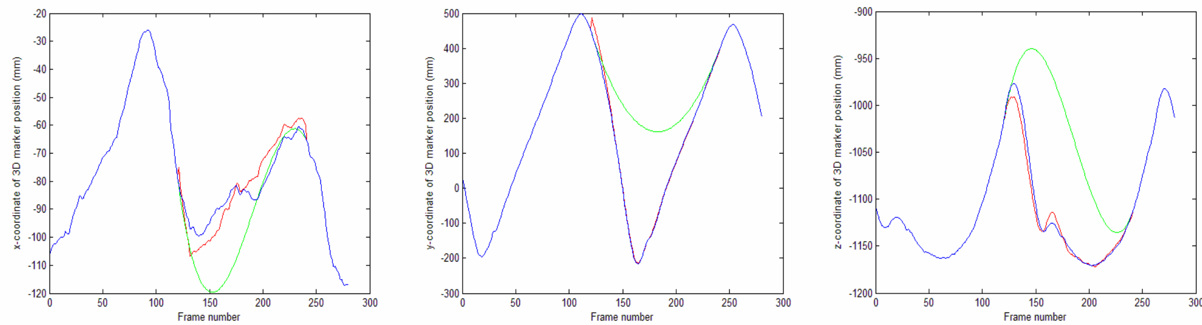
[18] A. Safonova, J. Hodgins and N. P. Pollard, "Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces," In *Proceedings of SIGGRAPH,* 2004, pp. 514–521.

[19] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis", *Journal of the Royal Statistical Society* Series B, Vol. 61(3), 1999, pp. 611–622.

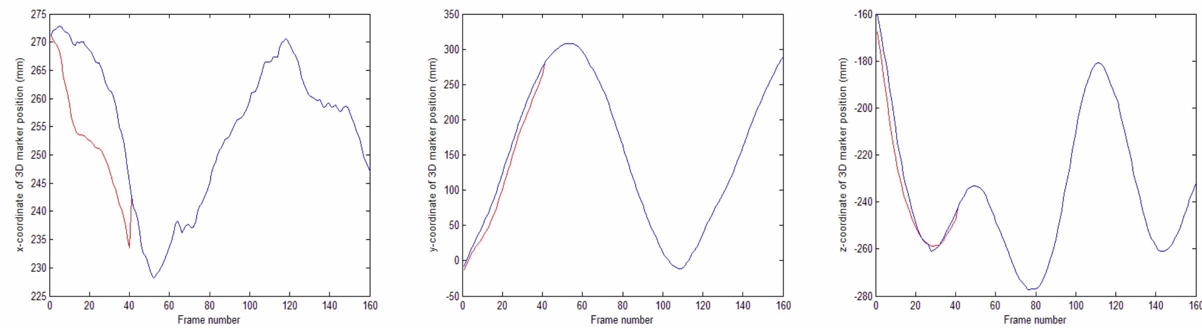
[20] Vicon Motion System Ltd, http://www.vicon.com/.

[21] G. Welch, G. Bishop, L. Vicci, S. Brumback, K. Keller, and D. Colucci, "The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments", In *Symposium on Virtual Reality Software and Technology*, 1999, pp. 1–10.

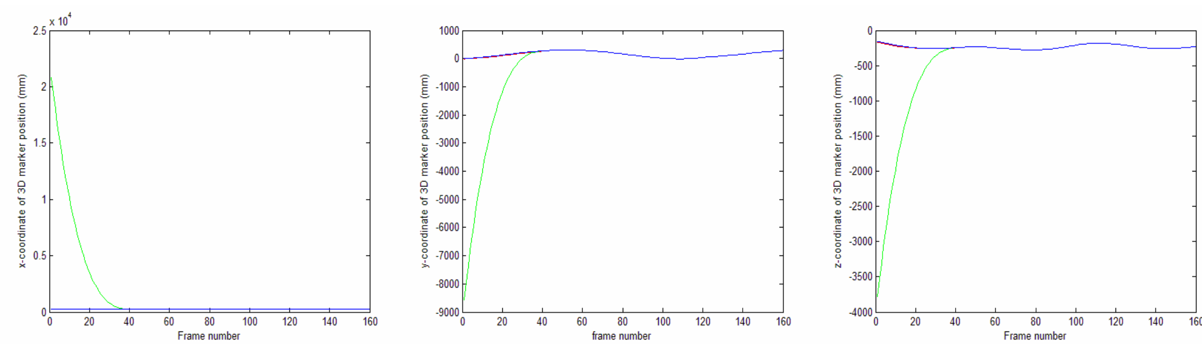
[22] *D. J. Wiley and J. K. Hahn,* "Interpolation synthesis of articulated figure motion", *IEEE Computer Graphics and Application*, 17(6), 1997, pp. 39–45.



(a) 3D positions of marker on the right ankle



(b) 3D positions of marker on the left arm (Original motion and our estimation result only)



(c) 3D positions of marker on the left arm

Figure 4. Comparison of estimating results (original motion in blue, spline interpolation results in green and our estimation results in red). The top panel corresponds to the marker on the right ankle; the middle and bottom panels correspond to the marker on left arm. The middle panel only shows the original marker positions and our estimations; while the bottom panel shows the original marker positions, spline interpolations and our estimations respectively in a larger scale.