Title
Skeletal parameter estimation from optical motion capture data

Journal
2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2

ISSN
1063-6919

Authors
Kirk, AG
O'Brien, JF
Forsyth, DA

Publication Date
2005

DOI
10.1109/cvpr.2005.327

From the proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2005.

Skeletal Parameter Estimation from Optical Motion Capture Data

Adam G. Kirk James F. O'Brien David A. Forsyth
University of California, Berkeley

Abstract

In this paper we present an algorithm for automatically estimating a subject's skeletal structure from optical motion capture data. Our algorithm consists of a series of steps that cluster markers into segment groups, determine the topological connectivity between these groups, and locate the positions of their connecting joints. Our problem formulation makes use of fundamental distance constraints that must hold for markers attached to an articulated structure, and we solve the resulting systems using a combination of spectral clustering and nonlinear optimization. We have tested our algorithms using data from both passive and active optical motion capture devices. Our results show that the system works reliably even with as few as one or two markers on each segment. For data recorded from human subjects, the system determines the correct topology and qualitatively accurate structure. Tests with a mechanical calibration linkage demonstrate errors for inferred segment lengths on average of only two percent. We discuss applications of our methods for commercial human figure animation, and for identifying human or animal subjects based on their motion independent of marker placement or feature selection.

1 Introduction

The term *motion capture* broadly refers to any of several techniques for recording the movements of a human, animal, or other subject, and then using the recorded data for animating synthetic characters. Motion capture techniques have found widespread commercial use in the movie, television, and video game industries, as well as in other areas ranging from biomechanics studies to performance art.

Currently, the most commonly used motion capture techniques employ optical methods to record a subject's motion. A set of markers are attached to the subject and then observed by a number of cameras. The capture system infers the time-varying location in space of each marker by triangulation based on the projection of the marker onto each camera's image plane. High-end systems typically employ a redundant array of many cameras to minimize marker loss due to occlusions and to provide accuracy over large capture volumes. To facilitate the task of segmenting markers from background, the systems use one of two methods for increasing marker contrast. The more common *passive systems* work with markers made of a strongly retro-reflective material and an illumination source co-located with each camera. *Active systems*, such as the one shown in the left-hand side of Figure 1, use LED markers that pulse in sync with the cameras' digital shutters. In both types of system, the cameras are fitted with optical filters tuned to the wavelength of the illumination source or LEDs. One significant difference between the systems is that the active markers can communicate unique identifications by modulating their pulses, whereas passive systems must infer marker identity from continuous observation. In practice, both types of optical system can reliably generate accurate position data for the markers with only occasional gaps caused by occlusions or the subject exiting the capture region.

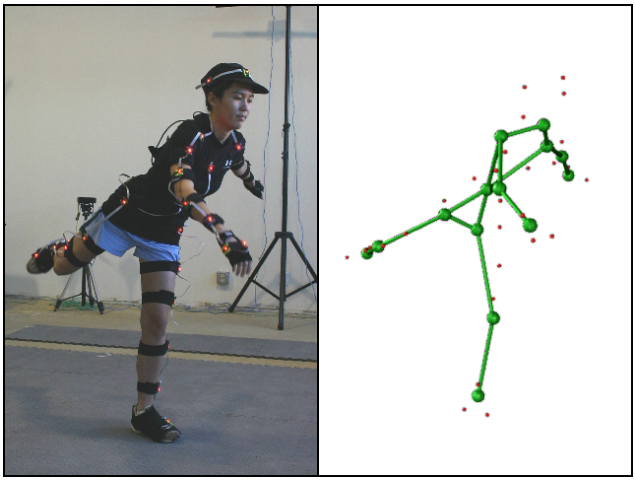

Figure 1. Automatic skeletal reconstruction for a human subject captured with active markers: The image on the left shows a photograph of the subject during the capture session. The image on the right shows the reported marker positions corresponding to when the picture was taken along with the kinematic skeleton automatically constructed by our system.

Motion capture systems output a set of point locations over time. This format does not explicitly capture the overall structure of the subject being captured. Fitting an articulated model, or skeleton, to the data uses the skeleton's structure to simplify the representation of the motion. This type of model facilitates recognizing the nature of a motion being performed as well as the identity of the subject. Pa-

Contact e-mail: *[akirk,job,daf]@eecs.berkeley.edu*

rameterizing the data with such a model also makes motion editing and visualization more convenient.

This paper describes an automatic method for inferring a kinematic model of the recorded subject directly from either passive or active optical motion capture data. Our method works solely from the marker trajectories and does not require any user intervention, nor does it require that the subject assume any particular pose. Our method does assume that the subject's kinematics can be well approximated by an articulated skeleton, but makes no other assumptions about the topology or structure of that skeleton. Instead the method automatically infers an appropriate skeletal topology and structure from the motion of the markers.

We rely on the key fact that, in an articulated structure, points fixed relative to one of the rigid segments in the structure will maintain a constant distance from other points on that same segment and from the centers of the joints connecting the segment to the rest of the articulated structure. Provided that the subject exercises each joint, and that one of the segments attached by the joint have at least two markers and the other at least one marker, these constant-distance requirements supply enough constraints to uniquely fix the location of the joint center.

Unlike prior approaches, our method does not operate by first finding transformations between the coordinate frames of each skeletal segment. Because there is no need to estimate those transformations, our method works with as few as one or two markers on each segment. Additionally, our algorithms do not inherit the instability of estimating rotations from a small number of closely grouped points.

The output produced by the method consists of an assignment of each marker to one of the segments in the kinematic model, the topological connectivity between the segments, the locations of the rotational joints connecting the segments, and the locations of each marker in its segment's reference frame. An example of the skeleton constructed for a human subject appears on the right in Figure 1.

This information can serve a number of practical uses. The most obvious application is motion capture for animation where it can replace the current, largely manual, calibration procedures used with most systems. Because the subject does not need to assume any special pose for our process to work, it can also be used in situations where systems that require special poses would be impractical.

In addition to animation uses, it applies to segmentation and to recognition as well. Point reconstructions could come from motion capture equipment, but they could also come from structure-from-motion procedures. Several papers have demonstrated methods for *segmenting* motions of distinct rigid bodies from image observations, and then reconstructing the points on the bodies separately (see, for example, [2,4,14,15]). Structure from motion produces unstructured point clouds in 3D, which are often then formed into meshes which produce 3D models (*e.g.* [5]). Such models are useful for rendering, but can be difficult to match for recognition purposes. For example, it would be extremely difficult to match clouds of points lying on the surface of a person to a model of the individual, because the samples may be taken at different points on the example and on the model. Given points that are derived from an appropriate structure (a kinematic tree), our method can be used to: (a) determine which such points form a kinematic tree, where the rigid bodies are connected by rotational joints; (b) obtain a skeleton representation of the tree, suitable for matching to object descriptions. As evidenced by the data plotted in Figure 5, this structure should considerably simplify object matching.

We have tested our method using data from a variety of optical motion capture sources, both active and passive. For human subjects, the results show that a qualitatively good skeleton can be constructed reliably without imposing substantial constraints on the recorded motion. Inferred skeletons for a single subject have low variance with regard to geometry and topology. Quantitative results obtained for a mechanical linkage show that the estimated skeletal parameters accurately reflect those of the recorded subject.

2 Previous Work

The task of determining an appropriate articulated skeleton for some recorded motion data is a specific instance of the general problem of fitting a generative model to observed data. Numerous researchers in several different fields have studied variations of this problem, and as discussed in Section 3 we have borrowed several well studied techniques to address our particular goal.

Several computer graphics researchers have examined this specific problem of skeletal parameter estimation for various types of motion capture data. O'Brien and his colleagues [9] devised an algorithm to estimate skeletons from magnetic motion capture data. Because magnetic systems include both position and orientation information, those authors are able to set up a linear system with the joint location for a given pair of bodies as the unknowns and where each frame of motion contributes a set of constraints. The residual of the least-squares solution to that system can be used to determine whether of not two bodies are in fact connected by a joint, which in turn allows them to infer both an appropriate skeleton topology and joint locations. Our method follows their general approach, however because optical systems cannot measure marker orientation, we end up with substantially different, non-linear system constraints that require different solution methods.

Other researchers have worked on skeleton fitting techniques for use with optical motion capture data. Silaghi and colleagues [12] describe a partially automatic method for inferring skeletons from motion. They solve for joint positions by finding the center of rotation in the inboard frame for markers on the outboard segment of each joint. This process requires at least three markers on each segment in order to estimate reference frames for each of the inboard

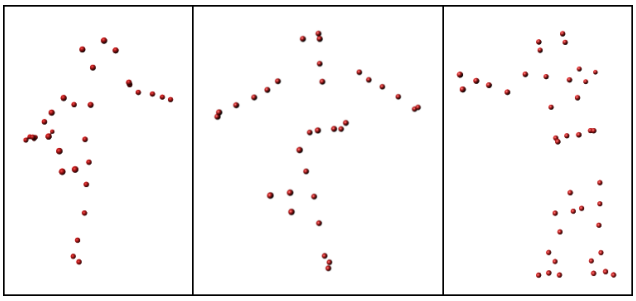

Figure 2. These images show marker configurations typical of the data from an optical motion capture system. Each image corresponds to one frame for one of our human subjects.

bodies. Their system requires substantial user interaction and also suffers from errors introduced by unreliable segment transformation estimates.

The method of Ringer and Lasenby [10], like ours, works with distance constraints although they still rely on rotation estimates. They assume that the skeletal topology is known beforehand and use heuristics to test multiple possible marker assignments.

Similar problems have also been studied in the biomechanics and robotics literatures. A few specific examples of methods for inferring information about a human subject's skeletal anatomy from the motions of bone or skin mounted markers can be found in [1, 13, 16]. Karan and Vukobratović have published a survey of calibration by parameter estimation for robotic devices [7].

3 Methods

Our skeleton inference procedure contains three stages. During the first stage, our method groups together markers that roughly move as a single rigid body. We refer to these groups as marker groups. The second stage determines the topology and joint positions of the skeleton by solving for the location of a rotational joint between all pairs of marker groups and selecting only the joints with low fit residuals. This fit process produces a "noisy" skeleton, meaning that the joints connected to a single segment may translate with respect to each other. To correct, we have an optional third stage that performs a least squares fit to find the single optimal length of each segment and offsets for the markers.

The input obtained from an optical motion capture system consists of the three-dimensional position for each marker over time. (Figure 2 shows typical examples.) The marker positions are given at discrete points in time called *frames*. Not all marker positions are reported for every frame: occlusions and other factors can cause the motion capture system to lose track of a marker for some time period, creating gaps. The data for a given marker typically contains large errors just before the system loses track of that marker and for a short period after the marker is rediscovered. To eliminate problems with "ghost" markers and erroneous position data during those periods, the first few frames are trimmed off the beginning and end of each marker's data segment, and any marker with a maximum number of consecutive frames less than one half second is ignored. For active systems, a marker's identity is consistent across gaps, but a passive system cannot tell a previously unseen marker from the reappearance of an old one. As a result, passive system will very frequently generate multiple marker identities for a single physical marker. For now, we assume that marker identities are constant, but we later discuss how to merge different marker identities in Section 4.

3.1 Marker Segmentation

The first step of our method clusters markers into groups that represent body segments. These are the rigid components of the resulting skeleton. For instance, given an appropriate input motion of a human arm, our algorithm will segment the data into two sets, one representing the upper arm and the other the lower arm. This grouping occurs because the motion of upper arm markers can be well approximated with a rigid body transformation, whereas the motion of a set of markers spanning both the upper and lower arm cannot be expressed as a rigid body transformation. In an ideal rigid body, the points on the body do not move with respect to each other over time, and in particular the standard deviation of distances between points on a rigid body over time is zero. Therefore, to determine marker groups, our method clusters based on the standard deviation in distance over time between all pairs of markers.

Using all frames to compute the standard deviation in distance between two markers can be expensive and the computation will be unduly influenced by the sporadic errors in optical motion data where a marker's position may jump several inches for short periods. To address the speed issue, our method calculates this quantity only over a jittered uniform sampling of frames. In our implementation, these samples are selected over all possible frames at intervals of one half second, plus or minus a few thirtieths of a second. This jitter ensures that any periodic errors do not affect the segmentation. In particular, let us define a cost matrix, A, such that element A_{ij} is the standard deviation in distance between markers i and j for a particular sampling of frames. This dataset is segmented into n groups using spectral clustering [8], where n is input by the user[1].

We make the process resistant to sporadic jump errors in the marker positions using a variation of the RANSAC procedure [3]. Rather than using one sampling of frames, we find marker groups by clustering multiple times using several different samplings. From among these multiple clusterings, our algorithm selects the one which minimizes the sum standard deviation of distances over all marker pairs in each group, for all clusterings. To avoid penalizing large

[1] Alternatively, the system's eigen-gap provides a reasonable way to determine an appropriate number of groups.

marker groups, the standard deviation within a group is normalized by the number of markers in the group.

3.2 Fitting Skeletons

Given marker groups for each segment in the kinematic skeleton, our algorithm must also determine the skeleton's topology and the locations of the connecting joints. Both are determined by minimizing the same quantity, called the joint cost. A joint between two segments in an articulated skeleton should maintain a constant distance from the markers in marker groups for both segments. For a set of markers on two body segments and a joint, we define the joint cost to be the mean variance in distance between the joint and each marker.

Given two bodies, the optimal joint position minimizes the joint cost for connecting them. The optimal skeleton topology is the one which minimizes the sum of joint costs over all connected segments. Our strategy for finding the optimal skeleton is to first find the joint costs for all segment pairs by solving a nonlinear optimization that finds the optimal joint position, and then second, to find the optimal topology by solving for a minimal spanning tree. To avoid excessive computational costs we only solve the all-pairs joint optimization approximately, then once we know the skeleton topology we solve for just those joints more accurately.

As stated above, we define the cost of placing a joint between two marker groups, b_a and b_b, to be the mean variance in distance between each marker and the joint position at each frame. Because the position of the joint is not known, our method optimizes to find the position that minimizes this cost. Unfortunately, the trivial solution to this minimization is to place the joint infinitely far away, making the variance in the distance between the joint and each marker zero. To keep the joint close to the marker groups a small distance penalty is added to the cost function. If we denote the location of a marker at a given frame as m_f and the location of a joint at that frame as c_f, then the average distance between a marker and a joint is $\bar{d}(c,m) = \frac{1}{|m|}\sum_f \|c_f - m_f\|$, where $|m|$ is the number of frames in which marker m appears. The variance in distance is then $\sigma(c,m) = \frac{1}{|m|}\sum_f (\|c_f - m_f\| - \bar{d})^2$. The joint cost is then

$$Q_{ab} = \min_c \frac{1}{|b_a| + |b_b|} \sum_{m \in b_a \cup b_b} \sigma(c,m) + \alpha \bar{d}(c,m)$$

where α is a small coefficient weighting the average distance penalty and $|b_a|$ is the number of markers in group a. This formula averages only over the number of frames that each marker appears in to avoid sensitivity to dropped frames. To find the position of a joint at each frame, our algorithm optimizes for c using the nonlinear conjugate gradient method [11]. A solution will be fully determined provided a and b together contain at least three markers, neither a nor b are empty, and that the motion exercises at least two of the three rotational degrees of freedom at the joint.

To determine the topology of the skeleton, *i.e.* how the marker groups are connected, our method uses an approach similar to that in O'Brien *et al.* [9]. In this stage, our method treats marker groups (which represent body segments) as nodes in a graph, and joints are possible edges. Because the topology of the skeleton is unknown, any pair of marker groups could possibly be connected by a rotational joint. The edge weight between any two nodes is the corresponding joint cost. Therefore, marker groups that should not be connected, such as the markers on the hand and the foot, will have a high joint cost. To determine the optimal skeleton, our method computes the minimum spanning tree of this graph.

Theoretically, the marker group segmentation step (Section 3.1) could be skipped. Instead of trying to model rotational joints between pairs of marker groups, our method could try to model rotational joints connecting all possible groups of markers. However, there would be a combinatorial explosion in the number of possible joints, so marker group segmentation is used to make the problem tractable.

In the discussion above, we assumed that the positions for all possible joints were found for all frames. The positions of joints that were not included in the optimal skeleton were then discarded. Rather than waste computation, our method drastically subsamples the number of frames given to the optimization procedure when determining topology. This procedure slightly affects the residual, however in our experiments the resulting topology is the same as when the method uses all frames. Once the topology is known, we solve the full optimization to find the joint position at all frames. However, now our method only runs the optimization procedure for joints that are known to exist.

When determining topology, a small average length term was added to the optimization criteria to keep the joints close to the skeleton. This is because incorrect joints could have otherwise been assigned a low joint cost by placing them at infinity. Because there are no incorrect joints in the second pass, this small average length term can be dropped. Interestingly, we found that with input motions that do not fully exercise all degrees of freedom, leaving the length term in can improve results. This length term amounts to a prior on the joint position that states the joint is close to the mean position of the markers in b_a and b_b.

3.3 Fitting Rigid Bodies

Due to noise in the input data, the joint positions found in Section 3.2 contain noise. Often times it is useful to find a true rigid body skeleton, both as a means of parameterizing the data and obtaining an estimate of noise in the input data. We present a method to project the joint and marker positions onto a rigid body skeleton. This is done one marker group at a time. This stage begins by collecting all the frames in which all the markers in a group appear, *i.e.* if

a marker in group b_a went missing at frame f, that frame is not used in solving for the projection. Using the method described by Horn [6], we compute the rigid-body transformations that best mutually align the marker and joint locations for that segment. If all the frames are lined up using their respective transformations, there will appear several small clouds of points representing each of the markers and joints connected to the body segment. The average position of each cloud of points is the model of the true offset of the marker or joint. These offsets are all defined with respect to the average position of the markers and joints connected to each body segment. Because the topology of the skeleton has already been determined, simply connecting the body segments based on the offsets from each segment center results in the correct rigid body skeleton.

As mentioned previously, estimating rotation matrices from groups of markers suffers from marker noise and requires at least three markers per segment. However, once a rigid body skeleton is fitted to the data, rotations can be found using inverse kinematics (IK). Since the rigid body skeleton and the offset of each marker from the segment is known, IK finds the rotations that minimize the distance from the marker position on the segment and the input data.

4 Marker Correspondence

The previous section described how to generate articulated skeletons, however it assumed known correspondence between markers. When using passive optical motion capture systems, correspondence information may be incomplete. Passive systems use markers that do not relay identities to the cameras. However, because the systems capture at fairly high rates, markers often do not move much between consecutive frames and identities can be tracked from frame to frame. If the system temporarily loses track of a marker, it will give the marker a new label, treating it as if the actor placed a new marker on their body. This means that, although the actor only wore a certain number of markers, the system often reports a significantly larger number of markers in the data. Our algorithm performs better the more we know about each marker, so it is beneficial to infer these correspondences.

Our method for determining marker correspondence over frames is based on the observation that portions of an actor's body will often pass through similar configurations at different times. This means that any particular configuration (minus the global rotation and translation) is likely to appear at some other frame in the data. The distance between any two markers in matching poses provides an estimate of the log likelihood that those markers are actually the same.

Our method begins by grouping the data into n sets, where n is the total number of markers reported by the system. In our method, the number of physical markers n' equals the maximum number of markers appearing in a single frame. This assumes preprocessing has eliminated ghost markers. In this stage we assume n is greater than n', otherwise correspondence is already known. The ith set, p_i, consists of all frames in which marker i appears. Again, each frame contains the position data plus a flag indicating existence for all markers in that frame. In some sense, set p_i consists of all known correlations between marker i and all other markers. Note that in most cases any two sets will have some markers in common. Since we know there are n' physical markers, we would like to cluster these n sets into n' clusters, each of which represents a physical marker. Of course, clustering requires a measure of distance between two sets. We define the distance between p_i and p_j, D_{ij}, to be the minimum distance between markers i and j in all pairs of poses. Again, a pose is defined to be invariant to global rotation and translation. To compute this distance our method takes each pair of frames, one in p_i and one in p_j, and determines the rotation and translation that line up the markers found in both frames. In other words

$$D_{ij} = \min_{a \in p_i, b \in p_j} \|m_{i,a} - A m_{j,b}\|$$

where A constitutes the rotation and translation that aligns frame b with frame a based on the set of markers that appear in both frames, and $m_{i,a}$ is the position of marker i at frame a. Because poses don't change much in consecutive frames, rather than compare every frame our method samples frames from each set at a constant interval. Once the samples with the minimum distance between markers i and j are found, our method searches all pairs of frames around those samples for the optimal alignment. This is done purely for efficiency, and in all of our experiments has produced the same answer as comparing all pairs of frames. Now that we have distances between all sets, spectral clustering [8] is used to cluster the virtual markers into actual markers. Assuming preprocessing has eliminated ghost markers, if marker i and marker j appear in the same frame they should not be clustered together. Therefore, if p_i and p_j have any frames in common, the clustering algorithm is biased to prevent those markers being grouped together. As described in [8], distances are converted to affinities using a Gaussian distribution. In the case of overlapping sets, the affinity is set to zero, and a repulsion matrix is used to indicate the sets should not be clustered together [17]. Note that the method will fail if all markers are lost simultaneously, for example if the actor steps out of the capture volume.

5 Results

We tested our method on multiple human datasets, however lack of ground truth measurements[2] means that evaluating performance on this data is difficult. To provide a

[2]We can measure limb lengths by hand, but accurately determining joint locations under skin is difficult and error prone. Our tests do show that the computed limb lengths agree with crude hand measurements to within two inches, which is consistent with our ability to correctly measure the joint locations.

Figure 3. Reconstructed skeleton overlaid on video frame. This figure allows rough visual assessment showing that the correct topology has been recovered and that joint locations are approximately correct.

quantitative analysis of our method's performance, we constructed a three-link chain of aluminum rods connected by universal joints, pictured in Figure 4. We captured this model using a PhaseSpace active capture system and were able to reconstruct the length of the middle rod to a mean length of 34.31 centimeters (0.21 centimeters too long) with standard deviation 0.65 centimeters. The average error per trial was 0.58 centimeters with a standard deviation of 0.32 centimeters. These values are within the accuracy limits of the motion capture system. See Table 1 for complete results.

We reconstructed human skeletons using data sets gathered from both a PhaseSpace active motion capture system and from a Vicon passive motion capture system. The results are visually plausible, as seen in Figure 1 and Figure 3. In these experiments we had several subjects, both male and female. Between thirty to forty markers were placed on the subjects in various configurations. These markers were positioned on the body such that they were able to capture the rigid motions of a particular body segment. In particular, markers on the back could not be placed on the shoulder blades, due to the translational motion in the shoulder joint. Each subject then performed a calibration motion in which they exercise each joint through the full range of motion. This type of motion produced the best reconstructions. We also reconstructed skeletons from more typical motions, such as walking or dancing. Often times we were unable to segment out the hands in walking motion, due to the lack of

Trial	Result (cm)	Error (cm)	Error (%)
1	34.29	0.19	0.6
2	33.35	-0.75	2.2
3	33.72	-0.38	1.1
4	33.60	-0.50	1.5
5	34.49	0.39	1.2
6	35.39	1.29	3.8
7	34.64	0.54	1.6
8	34.80	0.40	1.2
9	34.54	0.44	1.3

Table 1. Reconstructed length for middle segment in aluminum rod linkage, actual length 34.1 cm

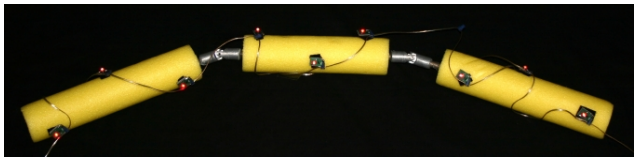

Figure 4. Aluminum rod linkage connected with universal joints, tracked using nine active markers

movement in the wrist degrees of freedom. In these cases, we found that we could append a calibration motion of the same actor, allowing us to reconstruct the position of the wrists in the walking motion.

The accompanying video shows motion clips from two different subjects. The first clip was obtained on the PhaseSpace active system, the second on the Vicon passive system. In both cases, we ran computations on the raw marker data. For the data from the active system, the skeletal reconstruction appears to correspond well to video of the subject despite noise in the form of jitter and occasional jumps in the marker positions. Lighting conditions with the passive system precluded recording a comparison video, but visual inspection shows that the reconstructed skeleton has the proper topology and appropriate proportions. That data is slightly cleaner so that the motion of the skeleton exhibits less jitter, except near the end when the motion capture system completely loses track of the subjects lower-left arm. However, that missing data does not prevent our system from reconstructing the skeleton, which allows the inverse kinematics routine to do a plausible job of filling in the missing data.

To illustrate the ability of our method to identify subjects, we compared segment lengths from the reconstructed skeletons of four different subjects. The motions used as input to this experiment were either calibration or dance motions, both of which exercise all major joints of the body. The reconstructed skeletons all had the same topology, namely that shown in Figure 1 and Figure 3, which allows a direct comparison of geometry. In Figure 5 the reconstructed lengths of the back (hips to neck) and upper arm (shoulder to elbow) segments are compared for several subjects. Note that the samples for each subject do not overlap

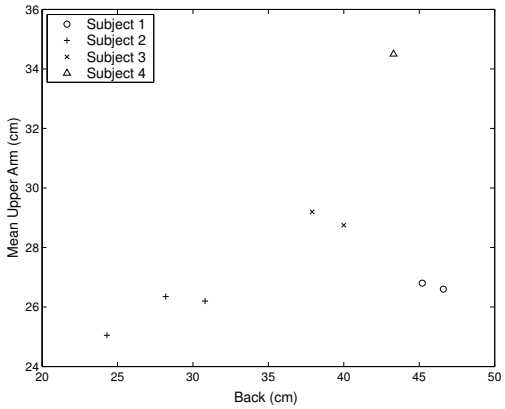

Figure 5. Plot of reconstructed back length versus mean reconstructed upper arm length for different subjects: This example illustrates the ability of our method to recognize subjects based on reconstructed skeletons. Similar plots result using other body parts.

and tend to group together. A subject's proportions could possibly be used to help distinguish between individuals.

6 Summary

This paper described an automatic method for estimating skeletal parameters from noisy point data. Our method is able to determine the overall topology of the motion capture subject, the length of each segment in the skeleton, the assignment of markers to segments in the skeleton, and the relative location of each marker with respect to the segment to which it is assigned. From this information our method is able to reconstruct orientation over time for each segment. Estimating this quantity directly from markers is an unstable process; by fitting a skeleton to the data our method provides a much more stable means of finding orientation.

Acknowledgments

We thank the other members of the Berkeley Graphics Group for their helpful criticism and comments. We especially thank Leslie Ikemoto and our other motion capture subjects. This work was supported by NSF CCR-02-04377, California MICRO 03-067 and 04-066, and by contributions from Pixar Animation Studios, Intel Corporation, Sony Computer Entertainment America, Apple Computer, Alias, the Hellman Family Fund, and the Alfred P. Sloan Foundation. Motion capture data was donated by PhaseSpace, the CMU motion capture database (*mocap.cs.cmu.edu*), and the Sony Computer Entertainment America motion capture department. The CMU motion capture database was created with funding from NSF EIA-0196217.

References

[1] J. H. Challis. A procedure for determining rigid body transformation parameters. *Journal of Biomechanics*, 28(6):733–737, 1995.

[2] J. Costeira and T. Kanade. A multibody factorization method for independently moving-objects. *Int. J. Computer Vision*, 29(3):159–179, 1998.

[3] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[4] C. Gear. Multibody grouping from motion images. *Int. J. Computer Vision*, 29(2):133–150, 1998.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[6] B. K. P. Horn. Closed form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4:629–642, April 1987.

[7] B. Karan and M. Vukobratović. Calibration and accuracy of manipulation robot models – an overview. *Mechanism and Machine Theory*, 29(3):479–500, 1992.

[8] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, Cambridge, MA, 2002. MIT Press.

[9] J. F. O'Brien, R. E. Bodenheimer, G. J. Brostow, and J. K. Hodgins. Automatic joint parameter estimation from magnetic motion capture data. In *Proceedings of Graphics Interface 2000*, pages 53–60, May 2000.

[10] M. Ringer and J. Lasenby. A procedure for automatically estimating model parameters in optical motion capture. In *BMVC02*, page Motion and Tracking, 2002.

[11] J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain (edition 1 1/4). 1994.

[12] M.-C. Silaghi, R. Plankers, R. Boulic, P. Fua, and D. Thalmann. Local and global skeleton fitting techniques for optical motion capture. In N. M. Thalmann and D. Thalmann, editors, *Modelling and Motion Capture Techniques for Virtual Environments*, volume 1537 of *Lecture Notes in Artificial Intelligence*, pages 26–40, 1998.

[13] J. J. Spiegelman and S. L.-Y. Woo. A rigid-body method for finding centers of rotation and angular displacements of planar joint motion. *Journal of Biomechanics*, 20(7):715–721, 1987.

[14] P. H. S. Torr. Geometric motion segmentation and model selection. In J. Lasenby, A. Zisserman, R. Cipolla, and H. Longuet-Higgins, editors, *Philosophical Transactions of the Royal Society A*, pages 1321–1340. Roy Soc, 1998.

[15] P. H. S. Torr and A. Zisserman. Concerning bayesian motion segmentation, model averaging, matching and the trifocal tensor. In H. Burkharddt and B. Neumann, editors, *ECCV98 Vol 1*, pages 511–528. Springer, 1998.

[16] F. E. Veldpaus, H. J. Woltring, and L. J. M. G. Dortmans. A least-squares algorithm for the equiform transformation from spatial marker co-ordinates. *Journal of Biomechanics*, 21(1):45–54, 1988.

[17] S. X. Yu and J. Shi. Understanding popout through repulsion. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.