The Title of Your Paper Goes Here

Abstract

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Radiosity;

Keywords: character animation, motion capture, hand motion, dimensionality reduction, PCA

1 Introduction

Producing quality whole-body motion involves the movement of the hand in relation to the rest of the body. Using a motion capture system, it is difficult to record the full body of a moving person while also capturing the hand and all of its detail because the whole-body and hand appear at largely different scales. While it is possible to record a high-resolution capture of the hand through a comprehensive set of markers (typically 13-20 markers), this is best conducted in a small capture region, isolating the motion of the hand. However, in a larger, full-body capture region the complete set of markers becomes difficult to discern and so this approach is usually abandoned in lieu of the capture of a smaller set of markers (2-6 markers) coupled with a process for reconstructing the full hand animation. In this paper, we provide a robust technique for the latter that both automatically selects a sparse marker set to record and subsequently produces joint trajectories for a full skeleton from the sparse marker set.

Specifically, our technique employs a combination of Principle Component Analysis (PCA) [**?**] to construct a low-dimensional representation of the hand data along with a linearly weighted regression (LWR) model to aid in the reconstruction step. Starting from a reference database that is recorded using a full-resolution marker set, we both determine the best sparse marker set based on the PCA representation from this data as well as use it in the synthesis step with LWR. We experimented the size of the marker set to record, specifically reduced marker sets of six and three markers and compare our technique with different approaches proposed for selecting the marker, including manual selection as in [Hoyet et al. 2012] and the use of an representative cluster-based search approach [Kang et al. 2012]. In contrast, our technique computes the marker set directly.

For reconstruction, our method employs a smaller number of markers to generate full-hand motion. Based on a test query, we use LWR to build a local model of the full-resolution PCA-version... The results from our reconstructions show that much of the finger specificity needed when communicating with sign language is preserved. We show the power of our technique using American Sign Language (ASL) as a primary testbed along with a few other more generic hand motions, including gesturing and finger-counting.

2 Related work

The detailed and subtle motions of fingers are hard to capture. Many approaches have been suggested, each of them having their own advantages and disadvantages. Optical motion capture systems, while being very accurate, require substantial post-processing due to occlusions and mislabelings. CyberGloves [CyberGlove Systems 2013] require regular calibrations and do not provide the accuracy needed for our purpose [Kahlesz et al. 2004]. For image-based systems or systems using depth-data, the hand needs to be in a confined space and the body can not captured synchronously [Wang and Popović 2009; Zhao et al. 2012].

Our work focuses on facilitating capturing hand and finger motions together with body motions in a motion capture system. To reach this goal, we investigate the most effective way to capture accurate hand motions using the smallest possible number of markers and an effective reconstruction method. Previous research analyzed finger motions to find correlations between different degrees of freedom. Rijpkema and Girard [1991] found that the relationship between the flexion of the distal and the proximal interphalangeal joint (DIP and PIP, respectively) is approximately linear with $DIP = 2/3 * PIP$. Jörg and O'Sullivan [2009] reduce the 50 degrees of freedom of both hands to 15 by eliminating irrelevant and redundant information. These approaches show that finger motions are highly redundant. We take advantage of the correlations between different degrees of freedom of the hand to optimize the capturing and animation of hand motions.

Principal component analysis (PCA), as a standard technique to analyze and reduce high-dimensional data, has also been used to study finger motions. In Braido and Zhang's study [2004], the two first principal components of a PCA accounted for over 98% of all variance in the joint angles. However, their motion database did not take into account the thumb and involved only two types of tasks - cylinder grasping and voluntary flexion of individual fingers - which were repeated by different participants. Santello et al. [1998] studied a variety of 57 grasp poses and found that over 80% of the measured 15 degrees of freedom could be described by their first two principal components. Ciocarlie... However, all those studies apply to grasps that do not require specific motions from individual fingers. We present a method, which is optimized for American sign language (ASL), which exhibits an impressive dexterity and variety of finger motions. We hypothesize that there is less redundancy in typical finger motions of ASL than in standard grasping motions.

One of our goals is to determine which is the most effective set of markers for capturing hand motions. Previous work has optimized marker sets for hand motions by choosing markers sets manually and a reconstructing the motions with inverse kinematics [Hoyet et al. 2012] or with a brute-force approach that compares the error of similar poses found in a database [**?**]. In a similar manner, but for the full body, Chai and Hodgins,

Sign language: ... Hand low-res: KanWheZor12c, CioGolAll07, ChaPolXin07

Correlations between body and finger motions, in case we manage to include data on the body: MajZorFal06, JoeHodSaf12

Body low-res?: SafHodPol04 Perception?: JoeHodOSu10

3 Overview

Our overall technique is divided into two stages: 1) the computation of the sparse marker dataset; and 2) the reconstruction of the full-resultion, skeleton-driven hand animation from the sparse marker set. For our study, we collect of full-resolution motion capture data of hand motions in a small capture area. Our database consists of a total of XXX seconds of American sign language as well as XXX Part of that database is a XXX long sequence consisting of two

repetitions of the American sign language alphabet at 120fps (6570 frames) that we use as reference data for our first stage. Our actor wears 13 small (approx. 6mm) markers directly on the hands as well as three markers on the lower forearm. The lower forearm acts as the root link for our hand skeleton with the assumption that these same three markers will appear in full-body captures. To account for gross body hand motion, marker positions in the database are put into the same coordinate frame by computing the transformation of each marker relative to the root link. Our hand model consists of 18 joints.

In the first phase, we employ the reference data and perform PCA over the markers and derive a rank ordering for the markers based on their influences over the principle components. From this rank-ordered list, we select the top markers to act as our sparse marker set. For the second phase, reconstruction, we set up a locally weighted regression (LWR) model to map from the sparse marker positions to a set of derived principle components. In this case, PCA is applied to the joint angles. The LWR model is built for each test query based on the input markers for the query sample and their proximity to the analogous markers in the reference data after correcting for lower arm movement. Joint angles for the low dimensional input are then reconstructed by reversing the PCA process, from the principle components produced by the regression back to a full set of joint angles.

4 Hand Motion Dimensionality

At the core of our technique is the assumption that hand motion is relatively low-dimensional. Even though a full resolution skeleton of the hand can have several dozen degrees of freedom (DOF), many of the DOFs of the hand show correlations, so that the inherent dimensionality of the hand motions is much lower [Santello et al. 1998; Braido and Zhang 2004; Jörg and O'Sullivan 2009]. In our approach, PCA is used to exploit this low dimensionality as we assume that PCA will allow us to capture the important features of the whole-body hand motion in a small number of principle components.

To support these assumptions, we performed various tests to exploit the power of PCA for hand animation. First, we use PCA on the marker positions of our reference data. The dimensionality of this principle component representation is 39, comprised of 3 root-corrected position values for each of the 13 markers. *XXX It would be nice to have something here such as 80% of the information can be explained through the first three components.* We show that joint angles alone are not capable of producing the reconstructions we realize through PCA in Figure **??**.

PCA identifies the most important information in this database by making it the first component. Each subsequent component is less valuable than its predecessor. Using this knowledge, we sum up the values of each component over all of the samples and weigh each sum based on importance (which component). Each weighted sum correlates with a marker, and the markers are then ordered in terms of importance by their weighted sum. The chosen marker set is then used for the data we wish to reconstruct. To reconstruct a sequence of motion, we first only use data from the marker set selected in the previous step. This is the equivalent of having our performer only wear these markers in the initial capture session. These marker position are plugged into our regression model to determine the full motion of our low dimensional sequence in principle component space.

Second, starting again with our reference database, we conduct a PCA over the joint angles of the hand motion. With 18 joints with 3 Euler angles each, this results in a new principle component representation of the joint angles with 54 components. We performed an analysis to judge the ability of the PCA to directly reconstruct the original database motion and found that with as few as ten components the PCA could produce a motion with small but acceptable visual artifacts. An error plot of the reconstruction error measured by the joint angle deviation from the synthesized motion and the original motion appears in Figure **??**. Similar findings are reported using a small set of components from PCA to encapsulate the motion of full-body motion [Safonova et al. 2004] and our results here support similar observations made over hand motions.

5 Sparse Marker Selection

To construct an effective sparse marker set, our method exploits the full set of 13 markers recorded in the reference database and evaluates each marker's contribution to the whole-hand motion. In constrast to the exhaustive search proposed by Kang et al. [2012], our technique computes the markers directly using PCA.

We conduct PCA with the Cartesian positions of the markers relative to the root link. With 13 markers, this leads to a PCA of motion data with 39 dimensions. In general, PCA produces a covariance matrix and the eigenvectors of this matrix create a list of components ordered from most important to least important. Each component has 39 coefficients that describe the influence of each marker on that component. By adding up the contribution of each marker to all of the components, we can rank-order the influence of the markers on the full-set of components. Furthermore, from the eigenvalues, we are given the relative importance of each principle component with respect to each other. We can use this importance as a weighting to bias the components. Thus, by summing the weighted contribution of each marker to each of the components, our marker rank ordering can account for the described bias.

In our results we highlight sparse marker sets of three and six markers, as those form the range of what can be captured and post-processed reasonably well based on our experience. Given the number of markers desired for the sparse set, we select the set simply as the top markers based on the rank-ordering. We experimented with two methods of producing this rank-ordering, one with the eigenvalues acting as a weighting bias and the second treating all of the top-N principle components as equally important and simply ignoring the remaining components. Conservatively experimenting with N to be between one fourth and three fourths of the full dimensionality, these two approaches produced similar results. However, if we selected N to be the value of the full dimensionality, we did see reduced quality solutions. In practice, we employ the eigenvalue weighted ranking for all results showcased in this paper.

A nice feature of selecting the marker set in this fashion is that the rank-ordering simply adds subsequent markers from smaller sets to produce the larger sets. Thus, the described priority ranking reveals which are the definitively *most* influential markers regardless of the size of the sparse marker set. And so, in practice, adding more markers for higher quality recordings does not require a complete change of markers, only the addition of the desired number of markers to the ones employed in the lower quality recording.

6 Reconstruction

The reconstruction process takes as input a recorded sequence from the sparse marker set. It then produces joint angle trajectories that estimate the full hand motion. To this end, we build a regression

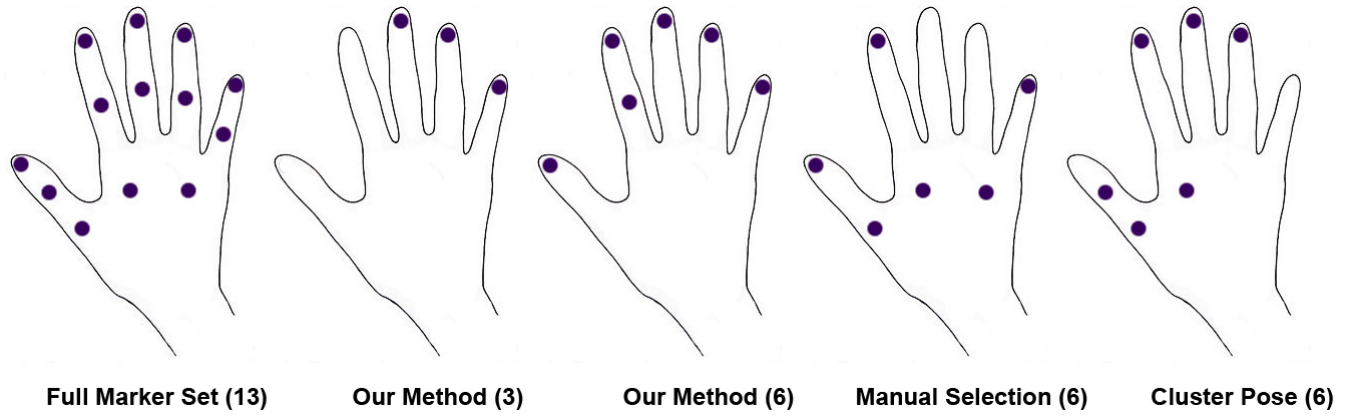

Figure 1: *The tested marker sets.*

model to construct joint angle measurements for a full motion sequence. Specifically, our locally weighted regression (LWR) model maps marker positions in the recorded sequence to principle components. Subsequently, the principle components are converted into joint angles using the covariance matrix from the PCA to produce the final motion.

The LWR model is built for each individual sample, or query, taken from the recorded sequence. Through this process, each instance in the database is weighted and this weighting is used to bias the model. The weighting is computed simply as the inverse of the Euclidean distance from the (root-link corrected) marker positions between the query and the samples in the database. The result is a regression that places importance on the reference samples that are close to the test query while also down-weighting the the influence of reference samples which are more distant from the query. For further discussion on this topic, we refer interested readers to similar efforts with whole-body work [].

At run-time, we introduce an input query sequence recorded from the sparse marker set. The input data is put through the regression modeling step to predict the principle components. To ensure smoothness, the trajectories of the principle components are filtered before they are converted into joint angles. In our results, we use a cone filter with a size of seventeen, with our sample rate for the motion recordings set at 120 hz.

We also experimented with filtering the joint angles to produce smoothness, but found more visually appealing results when we filtered the principle components. Our assumption for this finding is that the principle components combine to produce more "crisp" poses even when they are filtered while the joint angle filtering dilutes the unique features of individual poses over time. Further study of this phenomena is likely to produce interesting findings.

7 Results

Our primary database is used to reconstruct American Sign Language. The database is composed simply of two contiunous runs of the signs of the letters for the complete alphabet, signed by the same actor. We test our method on various sequences that include words in sign language, most of which are not included in the database.

For our sparse marker set, we choose to use six markers as our baseline in order to compare our technique to existing solutions. Using the method described in Section 5 to determine marker importance, the markers chosen are all of the fingertips and one on the lower part of the index finger. We also choose a marker set of three. The markers chosen are the fingertips of the middle, ring, and pinky fingers.

Sequences to be recontructed are initially recorded with a full marker set. Markers not selected for the sparse marker set are left out of the regression process.

Our method uses regression to predict principle components for a sequence of motion. In Figure 2, we compare the top three predicted components of a sequence of baby sign language to the components of the original sequence with a full marker set. The predicted components of six markers and three markers are shown. Though there are differences, the motion of each component closely follows that of the ground truth for both marker sets. This can be seen clearly witha another reconstruced sequence in the video.

We also compare a regression mapping marker positions to principle components to a regression mapping marker positions directly to joint angles. Figure 3 shows that the average joint angle error for mapping direclty to joint angles produces a much higher average joint angle error. The reconstructed hand with six markers fails to reach every distinct pose in the original animation.

We compare our marker set of six to the markers sets derived from the Manual Selection Method presented by Hoyet et al. (2011) and the Cluster Pose Error Method presented by Kang et al.(2012). Using the regression method, our marker set produces a smaller average joint angle error per frame for all of our current sign language tests. The Manual Selection Method's marker set consistently has the largest joint angle error. Figure 4 shows these differences, again using the clip of baby sign language as an example. The three distinct poses reached in the baby sign language clip are also shown using the different marker sets in Figure 5. Our marker set is consistently close to the original pose where as the two other marker fail at achieving at least one pose.

8 Discussion

When performing the regression we map marker positions to a certain number of components. We experimented with a smaller number of components would produce a better reconstruction of the joint angle data. We found that mapping to the full amount of components (54) produces the smallest average error, we can map up to 35 components with very little degradation from a full component set.

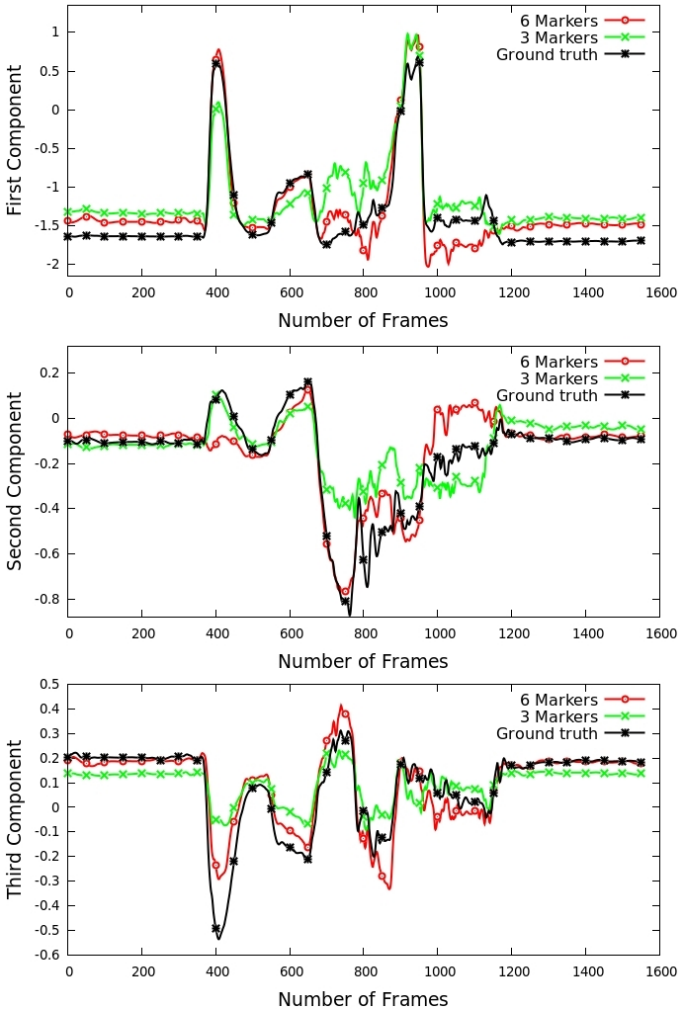

Figure 2: *Comparison of the components of a reconstructed clip of baby sign language when using 6 markers and 3 markers. Ground Truth is the original clip recorded with 13 markers.*

We also tested using our locally-weighted regression model to map marker positions directly to joint angles represented as Euler angles. The average joint angle error per frame was very high. This can easily be seen in the animations produced by the reconstructed joint angles. The hand does not reach the majority of the poses in the motion. From this we see that there is a clear benefit to using and producing principle components to reconstruct the joint angles of the hand over mapping directly to joint angles.

We also experimented with three has a larger average error than the marker set of six, but still appears to produce reasonable results. We can see this when looking at the top principle components of the reconstructed motion and comparing it to the top principle components of the original motion. For one sequence of motion, (in Figure BLAH), the top three components for both the original motion and the reconstructed motion with 3 markers appear to follow very similar patterns. Although there is information lost in the reconstruction, the general pattern of motion is the same.

Lastly, we attempt to reconstruct motions that are not sign language using our alphabet database. The motions include counting and general gesticulations. While the general poses in the sequences appear to be reached, the accuracy of the joint angles is visibly not as good the sign language reconstructions. It may be necessary to have a different database to properly reconstruct these motions.

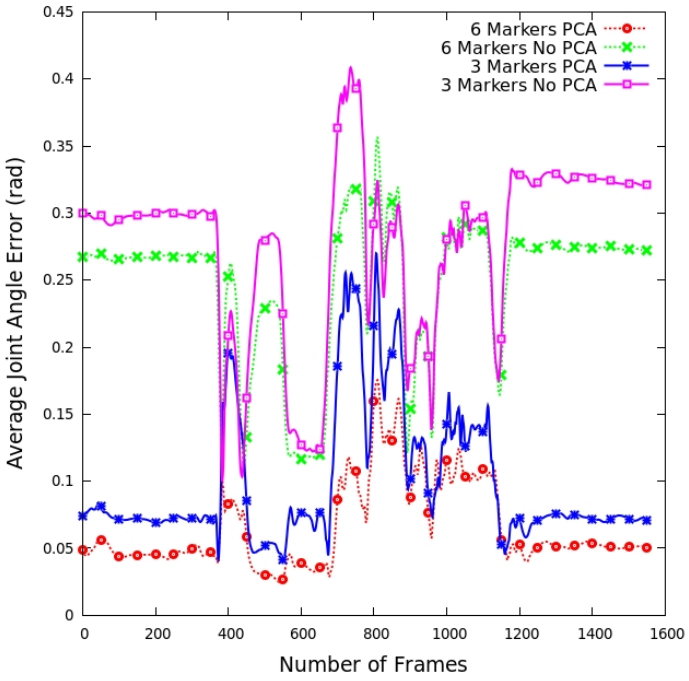

Figure 3: *Comparison of two regression methods: regression to principle components and regression to joint angles.*

9 Conclusion

In this work, we present a method to capture subtle hand motions with a sparse marker set consisting of three to six markers. Our method first specifies an appropriate set of markers using principle component analysis to exploit the redundancies and irrelevancies present in hand motion data. It then reconstructs the full hand motion based on the sparse marker set with a locally weighted regression mapping marker positions to principle components.

We show that our technique can reconstruct complex finger motions based on only three markers and outperforms methods presented by Hoyet et. al [2012] and Kang et. al [2012] in recent years. Our findings also indicate that using a regression mapping marker positions to principle components leads to better results for reconstruction of the full hand motion than using a regression mapping marker positions directly to joint angles.

The main limitation of our work is that the selection of the markers to capture is not readily applicable to other types of hand motions and the first step of our method – computing an efficient sparse set of markers based on a database of hand motion – has to be performed for every type of hand motions. Future work will explore what sparse marker sets would be most valuable for other types of hand motions such as grasping motions or gestures accompanying speech and thus investigate how far our approach generalizes to different types of hand motion databases.

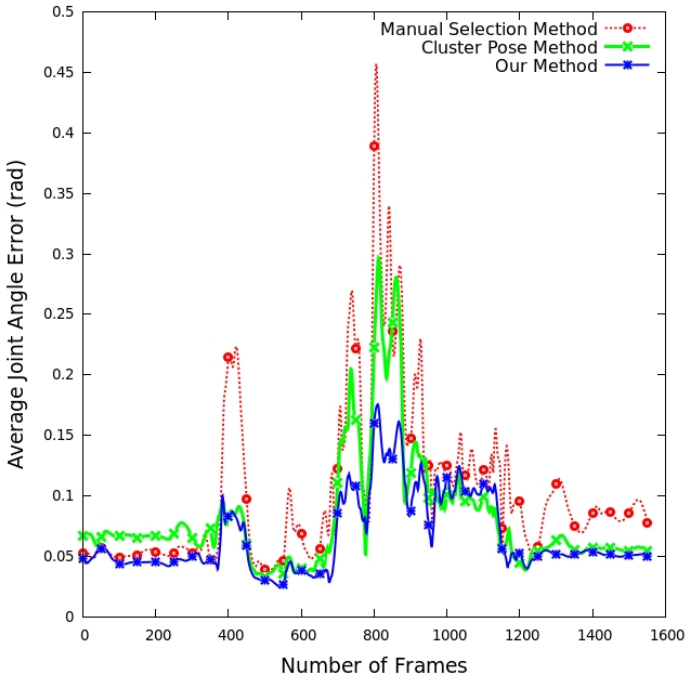

Figure 4: *Comparison of three marker set selection methods that use 6 markers.*

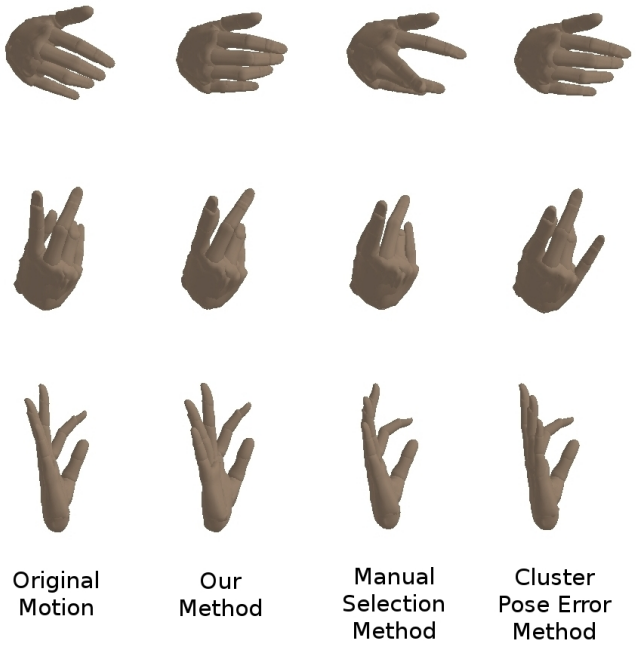

Figure 5: *The three distinct poses of the baby sign language clip reconstructed with the three different marker sets. They are compared to the original poses.*

Acknowledgements

References

BRAIDO, P., AND ZHANG, X. 2004. Quantitative analysis of finger motion coordination in hand manipulative and gestic acts. *Human Movement Science 22*, 6, 661–678.

CYBERGLOVE SYSTEMS, 2013. http://www.cyberglovesystems.com/products/cyberglove-iii/overview.

HOYET, L., RYALL, K., MCDONNELL, R., AND O'SULLIVAN, C. 2012. Sleight of hand: perception of finger motion from reduced marker sets. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '12, 79–86.

JÖRG, S., AND O'SULLIVAN, C. 2009. Exploring the dimensionality of finger motion. In *Proceedings of the 9th Eurographics Ireland Workshop (EGIE 2009)*, 1–11.

KAHLESZ, F., ZACHMANN, G., AND KLEIN, R. 2004. 'visual-fidelity' dataglove calibration. In *Proceedings of the Computer Graphics International*, IEEE Computer Society, Washington, DC, USA, CGI '04, 403–410.

KANG, C., WHEATLAND, N., NEFF, M., AND ZORDAN, V. 2012. Automatic hand-over animation for free-hand motions from low resolution input. In *Motion in Games*. Springer Berlin Heidelberg, 244–253.

RIJPKEMA, H., AND GIRARD, M. 1991. Computer animation of knowledge-based human grasping. In *Proceedings of the 18th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, SIGGRAPH '91, 339–348.

SAFONOVA, A., HODGINS, J. K., AND POLLARD, N. S. 2004. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. In *ACM Transactions on Graphics*, 514–521.

SANTELLO, M., FLANDERS, M., AND SOECHTING, J. F. 1998. Postural hand synergies for tool use. *The Journal of Neuroscience 18*, 23, 10105–10115.

WANG, R. Y., AND POPOVIĆ, J. 2009. Real-time hand-tracking with a color glove. *ACM Transactions on Graphics 28*, 3, 1–8.

ZHAO, W., CHAI, J., AND XU, Y.-Q. 2012. Combining marker-based mocap and RGB-D camera for acquiring high-fidelity hand motion data. In *Proceedings of the 11th ACM SIGGRAPH / Eurographics conference on Computer Animation*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, EUROSCA'12, 33–42.