Automatic Hand-Over Animation using Principle Component Analysis

Abstract

This paper introduces a method for producing high quality hand motion using a small number of markers. The proposed "hand-over" animation technique constructs joint angle trajectories for the hand with the help of a full-resolution reference database. Utilizing principle component analysis (PCA) applied to the database, the system automatically determines the sparse marker set to record. Further, to produce the hand animation, PCA is used along with a locally weighted regression model to reconstruct joint angles. The resulting animation is a full-resolution hand which reflects the original motion without the need for capturing the full-resolution marker set. Comparing the technique to other methods reveals improvement over the state of the art in terms of the marker set selection. In addition, the results highlight the ability to generalize the motion synthesized, both by extending the use of a single reference database to new motions, and from distinct reference datasets, over a variety of freehand motions.

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. However, using a motion capture system, it can be 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 often only possible 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 "hand-over" process for reconstructing the full hand animation [Kang et al. 2012]. In this paper, we propose a robust technique to accomplish the latter that both automatically selects the marker set to record, and subsequently produces joint trajectories for a full skeleton from the marker set.

Our technique employs a combination of principle component analysis (PCA) [Bishop 1995] to construct a low-dimensional representation of the hand data along with linearly weighted regression (LWR) to aid in the reconstruction. Starting from a reference database that is recorded using a full-resolution marker set, we first determine the best *sparse* marker set to record based on a PCA representation of this data. We experiment with different test sizes for the marker set to record, specifically reduced marker sets of six and three markers, and we compare our selection method with different ones proposed for selecting the markers, including manual selection, following [Hoyet et al. 2012], and a method that uses representative cluster-based search for selection [Kang et al. 2012]. In contrast, the technique in this paper computes the marker set directly from the PCA, and our findings show that this marker set is superior to the other methods of selection for the reconstruction techniques we tested. For reconstruction, our proposed method employs a second PCA in a synthesis step combined with LWR. Starting from a test query that records only the sparse marker set, we use LWR to build a locally sensitive model between the markers and the principle components. These are then converted back to joint angles to complete the reconstruction of the recorded hand.

We show the power of our technique using American Sign Language (ASL) as our primary testbed. ASL is an important and interesting freehand application of hand motion. Further, it includes a richly diverse set of configuration poses for the hands. We show that we can construct new (unseen) ASL signs with high-visual quality using a relatively simple, generic ASL database. Generalization of the database reveals that we can use our technique to capture other motions, such as counting, even though they are not closely related to the original motion. However, this database reveals itself to be too specialized for subtler freehand motion, such as figurative gestures. Instead, when a more closely related gesture reference database appears, novel gesture animation quality improves drastically.

Our effort holds close similarities to previous work, especially the full-body motion control of [**?**]. In contrast, our main contributions include the distinct exploration of rich hand data, such as ASL, as well as our method for determining the best reduced marker set to take advantage of the power of dimensionality reduction realized by PCA. Further, our approach is far simpler and lends itself to ease-of-use and re-implementation. Our approach also has notable advantages over other related papers for hand-over animation, such as [Hoyet et al. 2012; Kang et al. 2012] in that we compute the best reduced marker set directly, rather than selecting it manually, or through brute-force search. Compared to all of these techniques, our technique is both simple to implement and fast to compute, striking a valuable compromise which is likely to lead to greater adoption for commercial use.

2 Related work

The detailed and subtle motions of the hands are hard to capture. Several approaches for recording have been suggested, each of them having advantages and disadvantages. In particular, optical motion capture systems, while being very accurate, can require substantial post-processing to handle occlusions and mislabelings. While Cyber Gloves [2013] can deal with captures in larger spaces, they also require regular calibrations and do not provide high enough accuracy for many applications [Kahlesz et al. 2004]. For other camera-based or range-scan type systems, the hand needs to be in a confined space and the body can not captured synchronously [Wang and Popović 2009; Zhao et al. 2012]. The lack of robust recording solutions has lead to a practice of hand-over animation in industry, foregoing detail capture in lieu of often laborious post-processing.

Algorithms have been proposed to generate hand motion automatically based only on the motion of the body [Jörg et al. 2012] or also on contacts with objects [Ye and Liu 2012]. A different approach suggests to capture the body and hand motions separately and to combine them afterwards [Majkowska et al. 2006]. However, none of those techniques ensures that the resulting hand motion is going to be the same than the one performed during the initial body capture.

Our aim focuses on facilitating the quality capture of hand motions, together with full-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 suggest a corresponding, specialized hand-over technique to reconstruct the full hand from the markers. Other researchers have analyzed finger motions and found strong correlations between different degrees of freedom. Rijpkema and Girard [1991] report 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] show how to reduce the degrees of freedom of the hands by eliminating irrelevant and redundant information. These approaches reveal that finger motion is highly redundant. And we take advantage of correlation between different degrees of freedom of the hand to optimize the capturing and construction of quality hand animation.

Principal component analysis, as a standard technique to analyze and reduce high-dimensional data, has also been used to study hand movement. Braido and Zhang [2004], show that for the hand 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. Santello et al. [1998] studied a variety of grasp poses and found that over 80% of the measured degrees of freedom could be described by their first two principal components. Chang et al. [2007] used supervised feature selection to find a set of five optical markers that could classify grasps with a 92% prediction accuracy. However, these studies, applied to grasps, do not require specific motions from individual fingers. In contrast, we present a method applied to American sign language (ASL), which exhibits impressive dexterity and variety of finger motions [Courty and Gibet 2010]. 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 studied best marker sets for hand motions, for example, by testing and comparing marker sets chosen manually (with reconstruction done using inverse kinematics) [Hoyet et al. 2012], or through a brute-force approach, that compares the error of similar poses found in a database [Kang et al. 2012]. Chai and Hodgins [2005] studied full-body motion with similar goals to the ones described in this paper, but once again use a manually selected marker set.

3 Overview

Our overall technique is divided into two stages: 1) the computation of the sparse marker set; and 2) the reconstruction of the full-resolution, 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. The actor wears 13 small (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. For the results in this paper, we construct two such databases, one for sign language and the other freehand gesture data.

In the first phase, we employ the reference data and perform PCA over the markers to 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 and their proximity to the analogous markers in the reference data after correcting for the lower arm (root) movement. Joint angles for the low dimensional input are then reconstructed by reversing the PCA process, from the principle components compued by the regression to a full set of joint angles.

4 Hand Motion Dimensionality and PCA

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 study the power of PCA for capturing the desired reduced dimensionality of hand motion. In Figure 1 below we show that PCA is indeed capable of reducing the dimensionality of the joint angle motion from the database, revealing low average errors for simple reconstruction with reduced numbers of components. This figure shows erros applied to our ASL database, which represents a diverse expression of poses for the hand. We see that PCA shows significant reduction in reconstruction error after around 10 components. While this is larger than reported findings for finger motion, the rich full hand gestures of ASL are still well-represented with a relatively small number of components. 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.

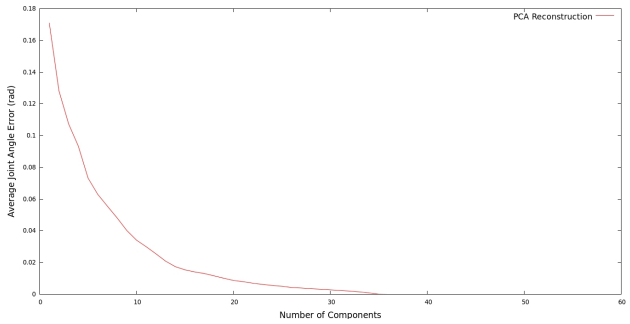

Figure 1: *Dimensionality reduction for sign language database. PCA is capable of using as few as ten components with relatively small average errors.*

Next, to compare the power of PCA for our particular application we experimented with two reconstruction methods with and without PCA. The details of the reconstruction appear in Section 6, however, we include the plot in Figure 2 here to support that PCA is very effective in producing higher quality hand motion. In the figure, we clearly see the benefit of employing PCA as a go-between from markers to joint angles. When we attempt to reconstruct without it (i.e. markers to joint angles directly) the error remains large even as the number of markers employed to inform the hand-over

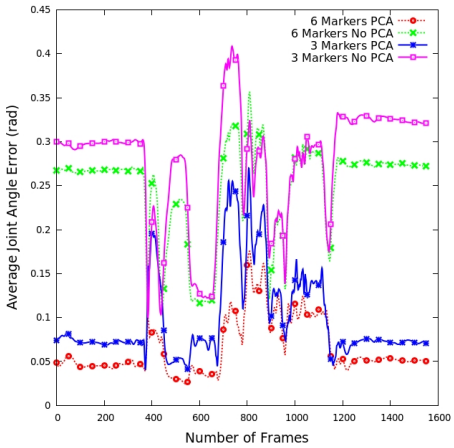

Figure 2: *Sign language sample motion with and without PCA employed. Note the error for six markers without PCA is larger than that of three markers with it.*

process is doubled.

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 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 also 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 easily 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 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 of the sparse marker set. It produces joint angle trajectories that estimate the full hand motion. To this end, we build a regression 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 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 distant from the query.

At run-time, we introduce an input 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" motion 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 reveal some interesting findings.

7 Results

Our primary database is used to reconstruct American Sign Language. The database is composed of two continuous runs of the letters of the alphabet, signed by the same actor. We test the database on various sequences that include "word" signs (e.g. boy or girl), 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. When we choose a marker set of three, the markers chosen are the fingertips of the middle, ring, and pinky fingers (Figure **??**).

Our method uses regression to predict principle components for a sequence of motion. In Figure 4, we compare the top three predicted components of a sequence of sign language to the components of the original sequence with a full marker set. Plots of the components of six markers and three markers are shown. Though there are differences, the motion of each component closely follows

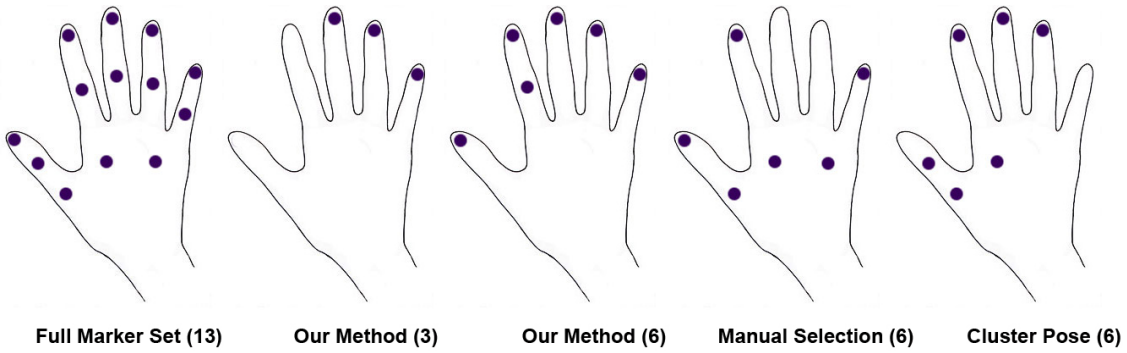

Figure 3: *The tested marker sets.*

that of the ground truth for both marker sets. This can also be seen in a reconstruced animation in the accompanying video.

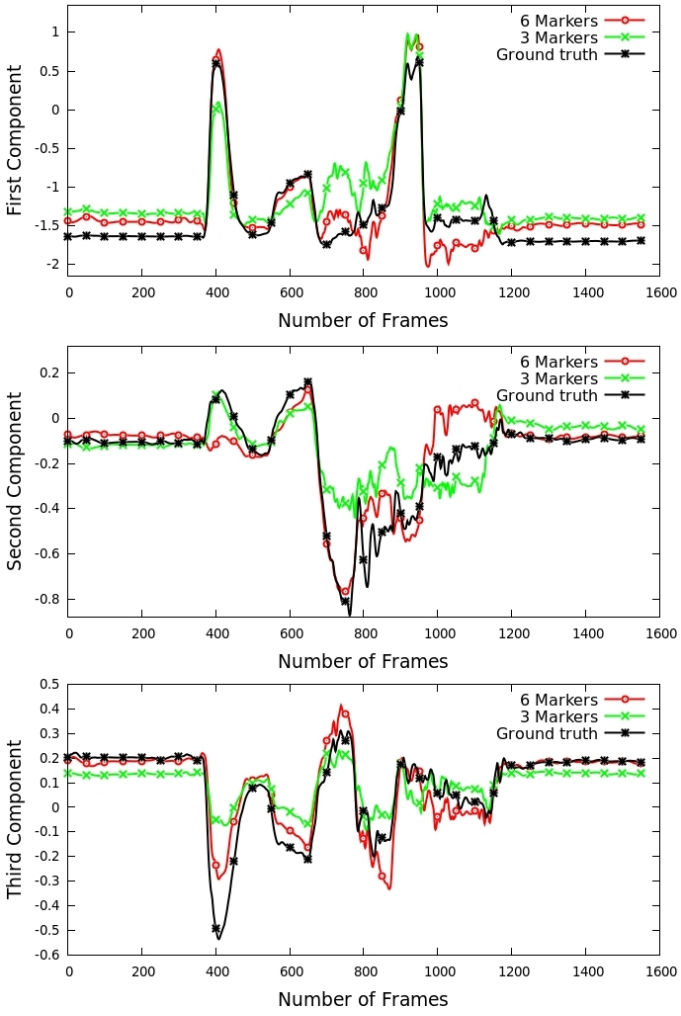

Figure 4: *Comparison of the components of a reconstructed clip of baby sign language when using 6 markers and 3 markers. Ground Truth is the original clip recorded with 13 markers.*

We compare our marker set of six to markers sets derived from the manually selected set, proposed by Hoyet et al. (2011) and the cluster pose error method proposed 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. Figure 5 shows these differences, again using the previous sign language clip as an example. The three distinct poses reached in the sign language clip are also shown using the different marker sets in Figure 6. Our marker set is consistently close to the original pose where as the two other marker sets fail at achieving at least one pose.

To test the robustness of the database, we attempt to reconstruct motions that are not sign language. The motions we test include counting and general gesticulations. Our sparse marker set of six successfully reconstructs counting the numbers 1 through 5, but the marker set of three fails to reconstruct the number 5. For the gesture based motion, many of the general poses in the sequence appear to be reached, but the accuracy of the joint angles is visibly not as good the sign language reconstructions. We then test to see if the use of another database can improve the gesture reconstruction. Our method is performed using a gesture database. The selected sparse marker sets of six and three are different than our previous sets. Using the gesture database results in high quality gesture reconstructions for both marker sets of six and three.

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.

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

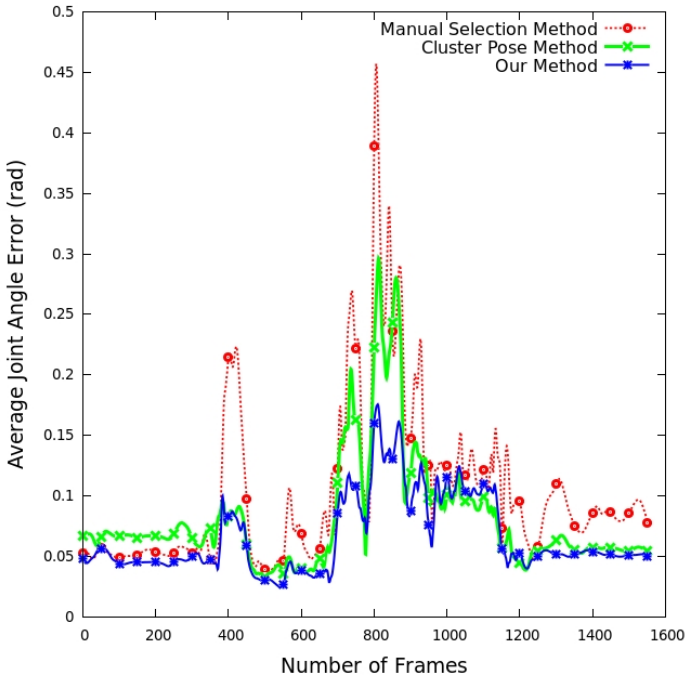

Figure 5: *Comparison of three marker set selection methods that use 6 markers.*

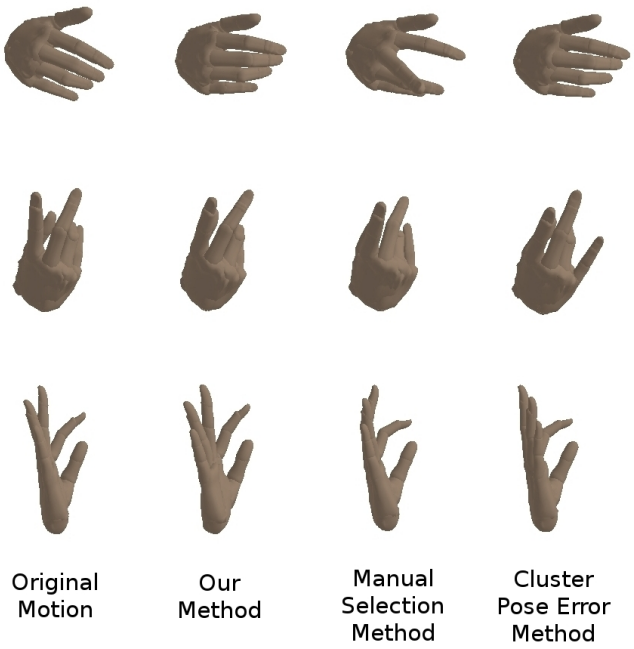

Figure 6: *The three distinct poses of the baby sign language clip reconstructed with the three different marker sets. They are compared to the original poses.*

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.

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.

References

BISHOP, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA.

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.

CHAI, J., AND HODGINS, J. K. 2005. Performance animation from low-dimensional control signals. *ACM Transactions on Graphics 24*, 686–696.

CHANG, L. Y., POLLARD, N., MITCHELL, T., AND XING, E. P. 2007. Feature selection for grasp recognition from optical markers. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2944–2950.

COURTY, N., AND GIBET, S. 2010. Why is the creation of a virtual signer challenging computer animation? In *Proceedings of the Third international conference on Motion in games*, Springer-Verlag, Berlin, Heidelberg, MIG'10, 290–300.

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.

JÖRG, S., HODGINS, J. K., AND SAFONOVA, A. 2012. Data-driven finger motion synthesis for gesturing characters. *ACM Transactions on Graphics (SIGGRAPH Asia) 31*, 6 (November), 189:1–189:7.

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.

MAJKOWSKA, A., ZORDAN, V. B., AND FALOUTSOS, P. 2006. Automatic splicing for hand and body animations. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, SCA '06, 309–316.

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.

YE, Y., AND LIU, C. K. 2012. Synthesis of detailed hand manipulations using contact sampling. *ACM Transactions on Graphics 31*, 4 (July), 41:1–41:10.

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.