Automatic hand-over animation for free-hand motions from low resolution input

Chris Kang[1], Nkenge Wheatland[1], Michael Neff[2], and Victor Zordan[1]

[1] University of California, Riverside
[2] University of California, Davis

Abstract. Hand-over animation is the process by which hand animation is added to existing full-body motion. This paper proposes a technique for automatically synthesizing full-resolution, high quality free-hand motion based on the capture of a specific, select small number of markers. Starting from a large full-resolution hand motion corpus, our technique extracts a finite pose database and selects the marker sets offline, based on user-defined inputs. For synthesis, capture sequences that include this marker set drive a reconstruction process that results in a full-resolution of the hand through the aid of the pose database. This effort addresses two distinct issues, first how to objectively select which is the best marker set based on a fixed number of desired markers and, second, how to perform reconstruction from this data set automatically. Findings on both of these fronts are reported in this paper.

Keywords: Character animation; motion capture; hand animation

1 Introduction

Hand-over is a term used in the animation industry to refer to the process of adding hand animation to pre-existing full-body motion. Hand motion is a critical part of many animations in which a full-body character is present. However, animation of the hand can be difficult especially where realism is important. While high-quality, full-body motion capture is a popular means for animating realistic characters, hand animation is most often not recorded at the same time as the motion of the full-body for a number of reasons. Foremost, hands are relatively small in contrast to the whole body and thus, unless a specialize capture device is used, sheer size complicates the simultaneous capture of both full-body and hand data. Further, hand recordings can be plagued with occlusion, which makes a clean capture of a full dataset for the hand difficult, especially in large capture areas. As a result, in production, hand-over animation is applied to full-body capture sequences through a manual post process.

As an alternative to capturing the full-hand motion at the same time as the body or adding it afterwards manually, we propose an automatic hand-over technique that captures a select set of markers at the time of the full-body capture and use this to automatically guide a data-driven model for hand-over animation. To this end, we propose a simple process that accomplishes hand-over in an efficient and straightforward manner. By using a pre-recorded database to animate the hand, the quality of the motion can be controlled and can look as good as a full-resolution capture. However, because we only use a small number of markers at the time of capture, recording and clean-up are much less troublesome than a full-resolution capture.

We focus on free-hand motions, those that do not include manipulation of the hand within the environment. Animations of free hands are prevalent in gesture, communication, and many other activities. In contrast manipulation tasks are more constrained, which affords a unique set of pros and cons. For example, techniques such as [1] exploit contact constraints to construct hand motion when recorded data for the hand is not available. In contrast, free-hand motion must derive its shape from other sources to remain natural. Our approach takes advantage of a rich database to produce natural free-hand movement.

1.1 Overview

As a preprocessing step, we record a repository of motion data which has the characteristics of the motion of the hand that we wish to animate. This is recorded in a small capture area with a tight camera set-up. We construct a rich pose database from the recorded full-resolution hand marker corpus. Then, from this database, the best representative group of markers are found based on the target number of markers. At the time of capture, this found marker set is recorded for the actor's hand (along with full-body motion.) The select hand marker data informs a reconstruction process to produce a full-resolution version for the data of the hand. Then, the synthetic full resolution marker set is fit to a skeleton rig for the hand and added to the full body. The result is an automatically generated hand animation that is synchronized and semantically meaningful based on the full-body motion and appears perceptually similar to capturing the full-resolution marker set for the hand.

1.2 Related Work

There has been a great deal of previous work in the domain of hand animation. Most of which has investigated the problems of grasp shape, interaction, and control. Although there are specialized techniques for capturing hand data as [2], the hand-over animation problem has been largely ignored save a select set of exceptions [3, 4, 1]. Much of the previous work in graphics and robotics has been on developing controllers (as [5–7]). Because designed hand controllers can lead to unnatural hand movement, researchers have looked into incorporating human examples [8] as well as the perception of various hand models [9, 10]. Another approach seen in previous research has been to address realism through anatomical modelling of the hand [11–13], but this work does not address the problem of generating motion for hand-over animation.

In comparison to the technique we propose, research on grasping is the closest to our own (in computer graphics) because it includes the selection of appropriate poses, for example [14–17]. Similar work appears in robotics as well [18]. Unlike this work, we are interested in more continuous sequences of expressive poses for free-hand animation while grasping techniques typically focus on isolated poses that lead to holds for different objects. Other related work focuses on specialized systems for interaction such as [19, 20, 1] to position the hand in a proper method for manipulation. But only a select group of efforts have focused on the problem of automatic hand-over for free-hand movement [3, 4]. In contrast to these efforts, we present a hand-pose approach as opposed to the motion-warping techniques presented previously. Our motivation for this approach is to maintain high quality poses at all points in the sequence, exploiting the low-dimensionality present in hand animation [21].

2 Full-resolution Analysis

A stand-alone goal of this work is to objectively determine which is the best set of m markers to use for the hand, given that m is the size of the small number of representative hand markers to be used. Choosing a small number for m alleviates issues related simultaneous hand/full-body capture. We set our sights on determining the best set of m markers from the total M markers used for the full resolution hand. Rather than selecting this set by hand as others have [10], we determine which marker set is best based on specific criteria.

Along with discerning the best marker sets for each value of m, we also wish to find the smallest sufficient full-resolution pose database from which to produce motion efficiently. Our hypothesis is that by using such a pose database and the best m-marker set, high-quality full-resolution data can be constructed from new test signals.

2.1 Selecting marker sets

To frame the problem of selecting the m-marker set, we will assume that we start from a given database of full-resolution markers and that some subset of the full M-markers will be employed for the production of the hand motion in the absence of the full marker set. To prepare the data, we automatically align each frame of the dataset such that the lower arm (wrist) is placed at the origin and at a common "zero" rotation. We employ three markers to do this, one on the lower arm and two at the sides of each wrist. We assume these markers will be recorded as part of the full body capture as is done in standard capture set-ups. This alignment step removes all wrist and arm movement and isolates the hand movement.

For comparing frames, we define an error metric as the sum of Euclidean distances (ED) between a given set of markers. We test every permutation of m marker configurations by computing the nearest neighbor error from the pose database to each frame of a test sequence of motion. We rank these trials to find the best m-marker set based on the average error for each permutation. Note, this test sequence is not a part of the original corpus, likewise for all subsequent query motions, etc. in the paper. By rank ordering the error associated with each m-marker combination, we find the best marker set for the given inputs. We summarize the algorithm in Figure 1.

```
marker_set_search( m, p, database ){
  for (i = every permutation of m markers) {
    for (k = every frame from test sequence){
        i' = extract_marker_set(i,k)
        for (j = 1 to p poses){
            j' = extract_marker_set(i,j)
            j_err = ED( i', j' )
            if (j_err < best_j_err) then best_j = j
        }
        i_err += ED(i', extract_marker_set(i,best_j)
    }
    if (i_err < best_i_err) then best_i = i
  }
  return(best_i)
}
```

Fig. 1.

2.2 Database construction

One assumption built into our approach is that we start from a database that encodes the full, rich expression of the hand that we expect to see in the final animation. While a very large database affords this assumption, clearly the size of the database is at odds with the efficiency and utility of the algorithm.

To produce a database of representative poses, we apply a selection process which picks unique poses from a large source database. To this end, we employ clustering on the large “raw” set of motion capture frames from a full-resolution capture of the hand and whittle the large corpus (of over 10,000 samples in our case) to a small but sufficient database of a select number of p poses. This clustering operation removes redundancy from the raw data. In practice, we found that *k-means* clustering [22] worked sufficiently for the purposes of splitting the data into like p groups and we then take a representative pose from each cluster. We use this pose in the final dataset. The algorithm computes distances between poses via ED of the full M-marker set and clusters into $p \in \{50, 100, 500, 1000\}$ groups. In the results, we find a p-pose dataset of 500 to work sufficiently, although for more rich data captures, a larger number of poses may be desirable.

3 Hand reconstruction

We produce the hand-over animation from low-resolution marker sequences through a straightforward reconstruction process. Starting from a sequence of low resolution data, the process finds the nearest example in the pose database using the wrist-aligned ED error metric. The best pose replaces the original data as a complete substitution for the low-resolution marker set for that frame. We opt to replace the full-hand to maintain pose fidelity.

Since the synthesized full-resolution marker data is computed pose-by-pose, it will be discontinuous over time. We perform a filtering pass on the marker data to smooth the individual marker trajectories and make them continuous over time. We employ a cone-filter for this process. We took some care to choose a width of the kernel to preserve features and found that a 15-sample width was acceptable for our 120 hz sampling rate. Finally, we perform the mapping of the filtered data employing the procedure described by [23]. Note, our motivation for this choice is to create synthetic virtual marker dataset and separate this process from the fitting approach used as this method is likely specific to existing pipelines. Sample results appear in Figure 2.

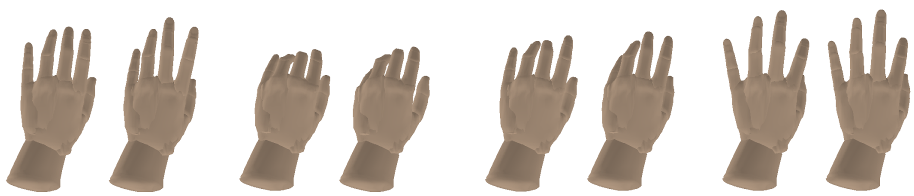

Fig. 2. A selection of frames contrasting the output of our system (right - each frame) with the original data (withheld from the database for testing.)

4 Results

We employ a Vicon 12-camera system for capture. To construct our corpus, the cameras are brought in to allow good coverage of a small capture space of approximately a one-meter cube. Within this space, using a sixteen-marker hand configuration, we recorded a wide range of free-hand movements. From this dataset we produced the pose database. We assess the choice of the number of poses p and the number of markers m based on their average ED marker error for all markers. Table 1 summarizes our findings with the average error per marker for various database sizes.

To assess the quality of the poses found in contrast to the marker selection, we include the full marker set which subsequently finds the best pose in the database using the full-resolution marker set and computes average ED for that pose. This measure provides a baseline and lower bound for the error for the pose database. As we can see from the table, the number m converges to this lower bound quickly and we conclude a marker set of 6 markers is sufficient, and the sets of 2 and 4 are even surprisingly good. From Table 1 we also see that the smaller pose databases are not as good as the larger ones. We conclude

p / m	2	4	6	Full
50	1.2 ($\sigma = .38$)	1.1 (.31)	1.1 (.32)	1.0 (.27)
100	1.1 (.32)	1.3 (.40)	.88 (.21)	.86 (.19)
500	1.0 (.57)	1.0 (.54)	.68 (.23)	.65 (.21)
1000	.88 (.27)	.80 (.35)	.65 (.23)	.62 (.20)

Table 1. Average error and standard deviation (cm) based on number of poses and number of markers chosen.

that 500-1000 appears to be sufficient based on the input corpus to obtain high-quality results with our approach. Figure 3 shows the results of the marker sets for best m of two, four, and six markers. Based on these results, we conclude a marker set of three is a compromise between the number of markers and the error realized. For the remaining results we adopt the best marker set of three markers with the 1000-pose database as the basic result and use it for further comparison.

To assess the quality of the best marker set found, we compare our results against the ground truth as well as the heuristically chosen marker set proposed in [10]. In Figure 4, we compare results on the 1000-pose database with the original data mapped for our test sequence. We treat the original as ground truth and showcase our found best-six marker set against the six-marker set suggested by Hoyet and colleagues. The plot also shows our marker set of best three which out performs their manually selected marker set in most cases. Note, while Hoyet and his colleagues produced the remaining animation heuristically for their results, we employ our reconstruction technique on their marker set as a control. (See the corresponding video at http://graphics.cs.ucr.edu/projects/handover/ for visual assessment of quality differences.)

One observation we draw empirically from working with the system is that the technique appears to act well with many markers sets. That is, the top n marker sets all appear to have comparable quality for a given pose database. As such, we consider a second criteria by which to judge the marker sets. Namely, considering our original goal of recording the low-resolution marker set in the full-body capture space, it seems important to keep the markers as far away from each other on the hand as possible, to avoid occlusion and marker swapping. So, we conduct an additional experiment that adapts the set selection by modifying the objective function to maximize the average distance between the markers on the hand and we apply this new search to the top-ranked markers sets based on our previous findings. Specifically, the previous top fifty marker sets are ranked based on the spread (ED) of the included markers on the hand, taken from a representative set of reconstructed poses. The top "distance" marker set for six markers is compared to the "best error" marker set and Hoyet's marker set in Table 2. As shown, the found marker set improve the distance measurement substantially with minimal reduction in quality.

In the video, we include two "stress tests." In the first, we use data for a different hand, showing the original motion compared to the best three and

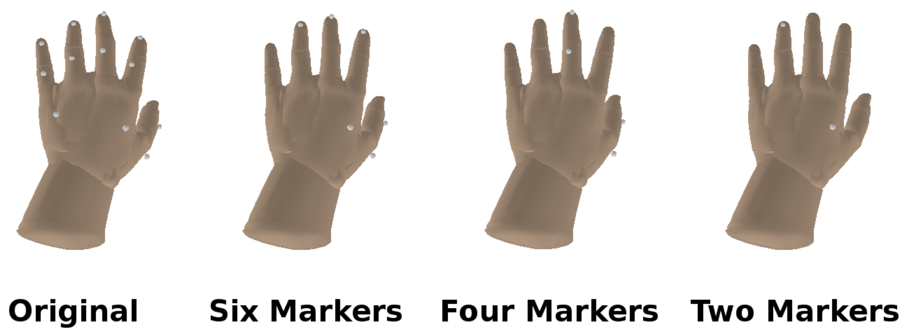

Fig. 3. Placement of best marker sets selected by our algorithm compared to the original marker set (left). The video compares the performance of each, we showcase here their placement on the hand. The markers shown indicate the position of the marker set relative to the hand.

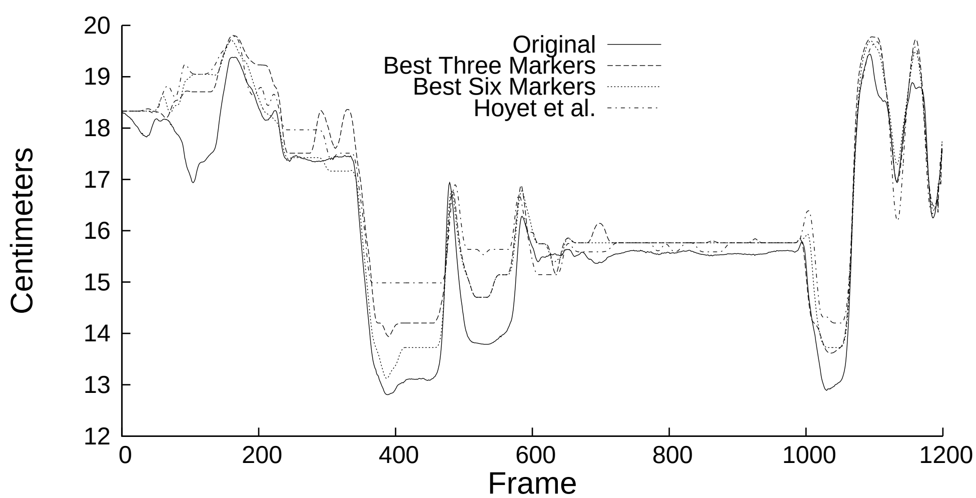

Fig. 4. Position of representative marker (base of index finger) for various marker sets in contrast with the original data. The Hoyet et al's set is the manually selected set of six markers suggested for hand capture based on their findings [10].

	Best Dist	Best Err	Hoyet et al.
Marker Error	.68	.65	.67
Marker Distance	8.4	6.0	5.3

Table 2. Comparison of three marker sets with 6 markers each for average marker error in resconstruction and average distance between markers (both in cm). The latter is a measure of how spread out the markers are on the hand.

the full marker set both reconstructed by our approach. The full marker set acts as a baseline from which to judge the pose database and to contrast the quality difference derived from the reduction of markers. The accompanying video reveals little perceptual difference between the full marker set and the best three marker set using our method but some discrepany between the full and the original, indicating the pose database may be the weakest link of the proposed approach. In the second stress test, we use our free-hand database on a grab action. We see that the animation reconstructed is highly dissimilar to the input motion. This is due to the fact that the poses simply do not appear in the set of poses in the database. We conclude from this ‘failure’ that the approach is best when a representative set of motions are used in constructing the database. We suggest that, in practice, good pre-made databases for specific sets of actions (e.g. free-hand gestures) might be used universally across subjects and capture sessions.

5 Conclusions

We propose a straightforward approach for automatically producing free-hand hand-over animation. Our technique is simple and we show its results in comparison to recent findings.

There are several limitations of the current system. First, we are making the assumption that the motion of the hand is free of obstacles, as we are not investigating the behaviors of grasping and manipulation and there are many challenging aspects of this class of behaviors that have been addressed in previous publications. Second, we assume we have examples of all the movements we might like to employ in the pose database. Third, we expect that the pose dataset is fairly densely sampled and therefore the resulting motion will not be overly jumpy. This assumption is likely to be the weakest of the paper, as the results do reveal periodic jumping, especially as the result dithers between two poses and the hand appears to pulse in an undesired manner between the two solutions. In practice, we did not follow any special procedures to make the result smooth. Since the capture data itself coming from the actor is continuous, we assume its temporal consistency will lead to the appearance of smooth final motion. The weakness of this assumption should be addressed in future work. However, as is, the quality of our automatic technique should be considered in comparison to other techniques (as [3]) and we anticipate that there will be need for post-clean up, but our results provide a much better starting point than complete animation synthesis.

The contributions of this work are both practical and fundamental. First, we outline a straightforward procedure by which hand animation can be generated by first producing a representative pose database and then accessing that database to create high-resolution motion for new sequences of data using a simple input marker set. We provide insight about how large the pose database needs to be and note that this database need not be recreated for every capture or every actor depending on quality requirements of the application. We address fundamental questions about how many and which markers should be recorded when a small set of markers is desired for recording of the hand. We provide our results on a comprehensive search of the best marker sets for various numbers of markers and compare them to a complete (full) and hand-selected marker set. As an added benefit, our method is straightforward and requires no special tools or expertise.

Given our discussions with several motion-capture industry specialists, this work appears to be addressing an important concern related to full-body capture and hand-over animation. Our technique is a step up both in quality and ease of use in contrast to existing processes in production today which are often labor intensive and subject to animator biases (which can reduce consistency and thereby quality.) As such, the strength of this work is truly in its simplicity and we anticipate it and follow-up work will be an asset to professionals that include hand-over in their animation pipeline.

Acknowledgments. The authors wish to acknowledge Karen Liu and the GVU Center for use of and help in capturing the hand animation database.

References

1. Ye, Y., Liu, C.: Synthesis of detailed hand manipulations using contact sampling. In: ACM Transactions on Graphics (TOG). Volume 31., SIGGRAPH (2012)
2. Wang, R., Popović, J.: Real-time hand-tracking with a color glove. In: ACM Transactions on Graphics (TOG). Volume 28., ACM (2009) 63
3. Jin, G., Hahn, J.: Adding hand motion to the motion capture based character animation. Advances in Visual Computing (2005) 17–24
4. Majkowska, A., Zordan, V., Faloutsos, P.: Automatic splicing for hand and body animations. In: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2006) 309–316
5. Iberall, T.: Human prehension and dextrous robot hands. International Journal of Robotics Research **16** (1997)
6. Mas Sanso, R., Thalmann, D.: A hand control and automatic grasping system for synthetic actors. In: Eurographics '94. (1994)
7. Bekey, G.A., Liu, H., Tomovic, R., Karplus, W.J.: Knowledge-based control of grasping in robot hands using heuristics from human motor skills. IEEE Transactions on Robotics and Automation **9** (1993) 709–722

8. Pollard, N., Zordan, V.: Physically based grasping control from example. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2005) 311–318
9. Jörg, S., Hodgins, J., O'Sullivan, C.: The perception of finger motions. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, ACM (2010) 129–133
10. Hoyet, L., Ryall, K., McDonnell, R., O'Sullivan, C.: 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, New York, NY, USA, ACM (2012) 79–86
11. Albrecht, I., Haber, J., Seidel, H.P.: Construction and animation of anatomically based human hand models. In: ACM Symposium on Computer Animation. (2003)
12. Tsang, W., Singh, K., Fiume, E.: Helping hand: an anatomically accurate inverse dynamics solution for unconstrained hand motion. In: Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2005) 319–328
13. Sueda, S., Kaufman, A., Pai, D.: Musculotendon simulation for hand animation. In: ACM Transactions on Graphics (TOG). Volume 27., ACM (2008) 83
14. Aydin, Y., Nakajima, M.: Database guided computer animation of human grasping using forward and inverse kinematics. Computers and Graphics **23** (1999) 145–154
15. Huang, Z., Boulic, R., Magnenat Thalmann, N., Thalmann, D.: A multi-sensor approach for grasping and 3D interaction. In: Proc. Computer Graphics International. (1995)
16. Rijpkema, H., Girard, M.: Computer animation of knowledge-based human grasping. In: Computer Graphics (SIGGRAPH 91 Proceedings). Volume 25. (1991) 339–348
17. Chang, L., Pollard, N., Mitchell, T., Xing, E.: Feature selection for grasp recognition from optical markers. In: Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, IEEE (2007) 2944–2950
18. Bicchi, A.: Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity. IEEE Transactions on Robotics and Automation **16** (2000) 652–662
19. Kim, J., Cordier, F., Magnenat Thalmann, N.: Neural network-based violinist's hand animation. In: Proc. Computer Graphics International. (2000)
20. ElKoura, G., Singh, K.: Handrix: Animating the human hand. In: ACM Symposium on Computer Animation. (2003)
21. Jörg, S., O'Sullivan, C.: Exploring the dimensionality of finger motion. In: Proceedings of the EGIE 2009, Eurographics Ireland Workshop (2009) 1–11
22. Frahling, G., Sohler, C.: A fast k-means implementation using coresets. In: Proceedings of the twenty-second annual symposium on Computational geometry, ACM (2006) 135–143
23. Zordan, V., Van Der Horst, N.: Mapping optical motion capture data to skeletal motion using a physical model. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, ACM (2003) 245–250