2014 11th International Conference on Computer Graphics, Imaging and Visualization

Analyzing and Segmenting Finger Gestures in Meaningful Phases

Christos Mousas Paul Newbury
Dept. of Informatics
University of Sussex
Brighton BN1 9QJ, UK
{c.mousas, p.newbury}@sussex.ac.uk

Christos-Nikolaos Anagnostopoulos
Dept. of Cultural Technology and Communication
University of the Aegean
Mytilene 81100, Greece
canag@ct.aegean.gr

Abstract

This paper presents a method to segment the motion of a character's fingers in an efficient way, so that each segmented motion will retain its required meaning. To determine the meaning of a gesture, the motion sequences are analyzed, and the correlation between the finger and the wrist motion is determined. This correlation is quite important, especially in cases that require automatic synthesis of finger animation. The proposed meaningful phases of the finger gestures are evaluated against a standard segmentation method. This involves, firstly, computing the cost of transition between two motion segments and, secondly, using a previously proposed distance metric for automatic finger motion retrieval. As the results show, incorrect calculations, as well as the transition cost, are reduced by the proposed motion segmentation technique.

keywords: *finger motion, motion segmentation, motion analysis*

1 Introduction

The need for realistic animated characters requires detailed motions, where not only the main body, but also the face and the fingers, should be able to animate naturally. In general, the motion of the character's fingers is responsible for enhancing the realism of the synthesized motion and enhancing the required meaning of the character's posture, as various perceptual studies [1][2][3] have shown in the past.

In this paper, we address two basic factors that a motion segment should be able to provide. Firstly, the segmented motion sequences should be easily recognizable. This means that the system should be able to recognize a motion segment, thereby reducing incorrect estimation. Secondly, it should be possible to synthesize the motion segments efficiently. In this case, there should be a smooth transition process for each pair of motion segments such as gained by eliminating undesired jerky effects, during the motion synthesis process. In this case, it should be noted that a gesture of a character and, consequently, of humans consists of five phases. They are: the preparation, where the arm moves away from its resting position; the pre-stoke hold; the stroke; the post-stoke hold; and the retraction, where the arm moves back into its resting position as Kendon [1] and McNeil [4] mentioned. Therefore, it is necessary that there be an ability to explore the meaningful part of the finger's gesture, and its association with the character's wrist, such as by providing a flexible way not only to segment the motion data, but also to recognize and synthesize it.

Based on these assumptions, a segmentation process is presented that can efficiently split the motion data into meaningful phases - the gesture phase and the non-gesture phases. To satisfy these requirements, the following steps were performed. Firstly, by using a collection of finger gestures that were provided by [5], all motion sequences that belong to a gesture type were synchronized by using a dynamic time warping function. Then, we computed the angular velocity of the character's fingers, which indicates when the fingers of the character start to move. Then, the mean average of the angular velocity of the fingers at each time step of each motion sequence of that gesture type is calculated, in order to retrieve a generalized representation of the gesture. The same procedure is repeated for the motion of the wrist. The results of the calculations are plotted on separate graphs for each gesture type. Then, by examining those graphs, one can easily determine the time periods into which a motion sequence should be segmented.

The rest of the paper is organized in the following manner. In Section 2, work that is related both to motion segmentation and to finger motion synthesis techniques is presented. In Section 3, the methodology of the proposed solution is explained. In Section 4 the results obtained from the evaluation processes of the proposed segmentation process are provided and discussed. Finally, conclusions are drawn in Section 5.

978-1-4799-5720-0/14 $31.00 © 2014 IEEE
DOI 10.1109/CGiV.2014.15

2 Related Work

During the past various segmentation methodologies have been proposed, especially for the locomotion of the characters, although less attention has been given to segmentation techniques to split the motion data of the fingers.

The most common way to segment hand motion, which was used by Jörg et al. [6] in their finger motion synthesis, as well as by Levine et al. [7], is the methodology proposed by Fod et al. [8]. More specifically, the motion segmentation process evolves while the velocity of the wrist joints crosses a threshold that is close to zero. This results in two phases - the motion phase that contains sequences with high speed, and the hold phase that contains low speed motions. In a second step, the motions are segmented again, with each new motion being at least 0.33 seconds in length, but no more than 2.0 seconds.

Finger animation techniques could be quite beneficial to the animation community. The techniques that are known as hand over animation [9][10]or hand-over motion reconstructoin[14], which are based on statistical analysis and synthesis of motion data [15], are processes in which the finger motion captured data is added over the character's hand motion on pre-existing, full-body motion captured data. Recently, various data-driven methodologies have been proposed for the automatic assembly of finger motion to a virtual character. More specifically, among the first methodologies proposed was that by Jin and Hahn [11]. In this approach, the pose space distance from the character's motion capture data is analyzed and a stepwise searching algorithm is used to find the key poses of the synthesized hand motion. Jörg et al. [6] proposed one of the most interesting solutions for finger motion synthesis for gesturing characters. That approach is based on the ability to synthesize a character's finger motions by assigning weight variables to the wrist's position and orientation, which are used as input control parameters of the motion synthesis process. Finally, another solution for synthesizing a character's finger motion was proposed by Majkowska et al. [12]. In the latter's methodology, finger motion and body motion are captured separately in a pre-processing stage.

3 Methodology

In the following subsections a way to analyze the motion data in two stages is presented. Firstly, we adopt a time warping methodology, as was proposed by Kovar and Gleicher [13] to synchronize the motion sequences that belong to each gesture type separately. Next, the motion of the fingers is analysed and then the motion of the character's wrist. Having analyzed the motion data, a simple correlation between the pieces of information is found, which will result in a segmentation technique that is able to restrict the cutting of the motion into meaningful phases. It should be noted that 64 gestures (8 gesture types x 8 repetitions), such as those provided by [5], were used for the motion analysis process. Each gesture type is treated separately based on the motion sequences that belong to it. Finally, the ability to automatically segment the motion data based on the examined parameter is presented.

3.1 Motion Analysis

Having synchronized the motion sequences to estimate the time period in which a gesture begins to appear, it is necessary to compute the necessary features of motion that can provide this ability. In this case, the calculation of the kinetic energy that the finger motions produce is a possible solution. However, by assuming that the mass of the character's hand is a constant variable, we conclude by computing the angular velocity of each of the degrees of freedom of the character's fingers, according to:

$$V(t) = \frac{1}{N \times D} \sum_{n=1}^{N} \sum_{d=1}^{D} v_{n,d}(t) \quad (1)$$

where $v_{n,d}$ is the angular velocity of the $d-th$ degree of freedom of the character's hand skeletal model for $d = 1, ..., D$, where in our case $D = 25$. Moreover, n denotes the $n-th$ motion sequence that is used for the gesture type analysis process, where in our case $n = 8$. $t = 1, ..., T$ denotes the total number of frames of each motion sequence. Finally, $V(t)$ is the angular velocity calculated for each frame for each degree of freedom of each motion that belongs to a gesture type of the character's hand. Therefore, for each of the motion sequences of a gesture type, the generalized representation of the gesture's motion is determined.

In the same manner as used previously to analyze the motion of the character's wrist, the velocity of each motion of a specified gesture is calculated. Then, for every motion of the character that belongs to a specified gesture, the mean value of the velocity at each time step for retrieving the generalized representation of the motion data is calculated. The resulting graph that illustrates the representation of the angular velocity parameter for the motion of the character's fingers and the velocity of the character's wrist appears in Figure 1.

3.2 Motion Segmentation

The process of correlation of the motion of the finger and the motion of the wrist is important. Although the ability to automatically synthesize the most valid motion of the character's finger into a new motion sequence that does not

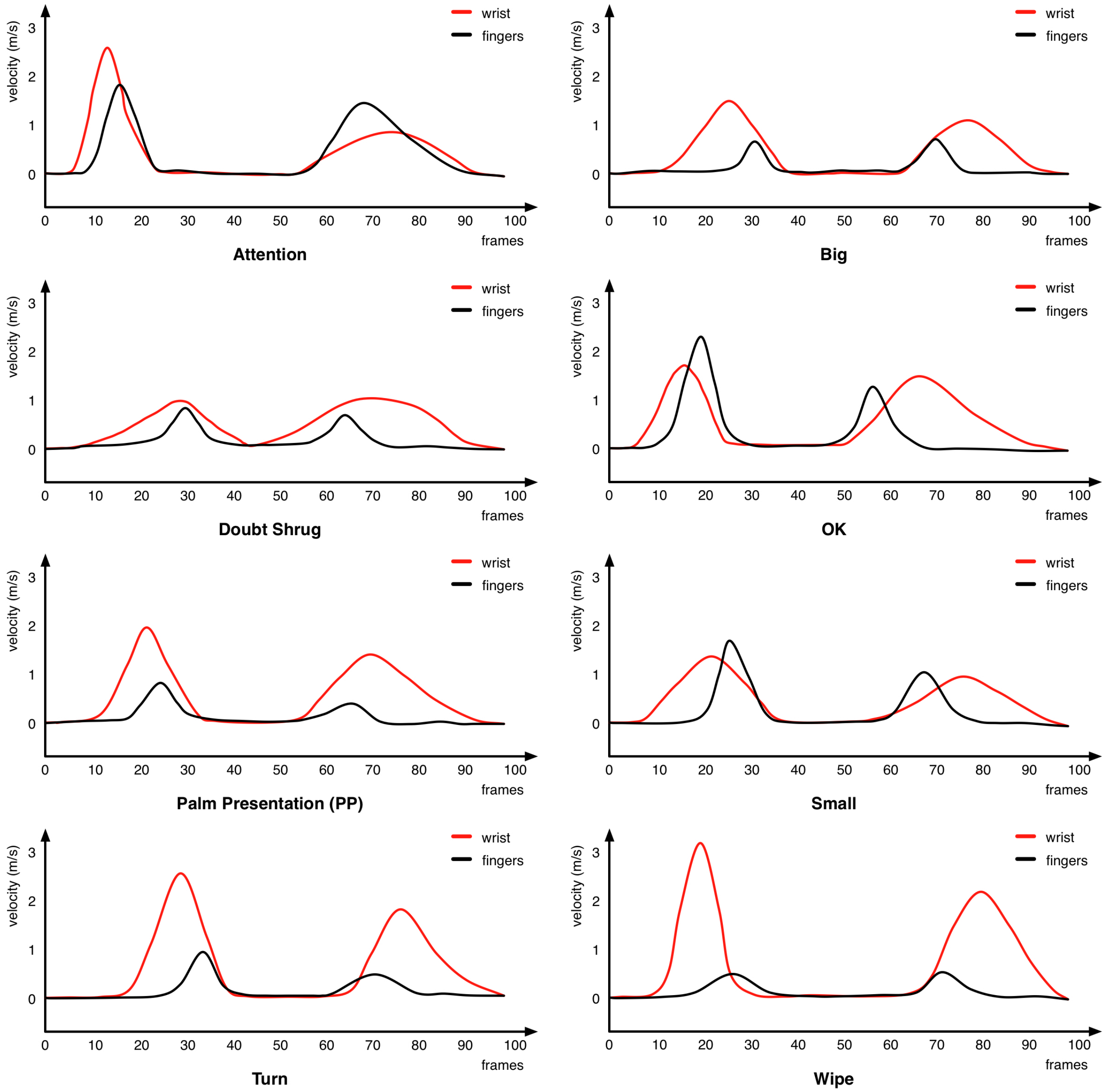

Figure 1: The generalized representation of the angular velocity parameter of the character's fingers and wrist for each of the different gesture types that was examined.

contain the motion of the fingers is required, a method to segment the input motion into the required phases should be investigated. In addition, since the input motion is segmented according to the wrist's parameters, this correlation is necessary for the system to efficiently synthesize the new motion, while keeping the necessary temporal parameters. Therefore, undesired effects such as incorrect synchronization should be eliminated.

In this case, in examining the graphs illustrated in Figure 1, one can assume that the most valid way to segment the motion of the character's fingers is while the character's wrist crosses its maximum velocity. This is because the character's fingers begin to react in order to perform a gesture during the period in which the wrist is located between the pre-stroke hold and the stroke phase of a gesture. This in-between phase (phase 2, see Figure 2) can be characterized by where the meaning of the gesture appears in the first frame of the phase, and disappears in the last frame of the phase. Therefore, the required meaningful phase of a motion segment is segmented efficiently by using the proposed segmentation method that relies on the analysis of the motion capture data.

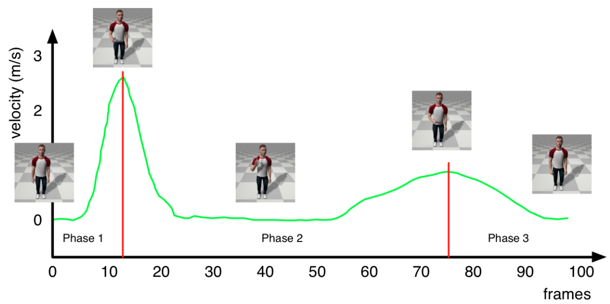

Figure 2: An example of a graph that plots the velocity of the character's wrist. The phases are split according to the proposed method. Phases 1 and 3 are the preparation and the retraction respectively, whereas phase 2 is the meaningful part of the gesture.

In this case, it should be noted that to automatically segment the motion data into the desired phases, as inspired by For et al. [8] we create segment boundaries when a shift from fast motion to slow motion is detected. Therefore, for each frame of motion data, we compute $z = \|v_{wrist}\|^2$, where v_{wrist} denotes the square of the velocity of the character's wrist. Therefore, based on the proposed segmentation method, we conclude by computing various finger gestures (see Figure 3) more efficiently as the evaluations that are presented in the following section.

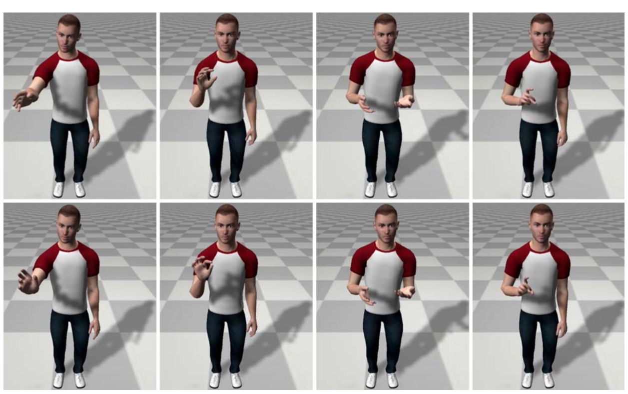

Figure 3: Given an input motion sequence (upper row), the system estimates the most valid gesture of the fingers (lower row), while using the proposed segmentation technique.

4 Evaluations and Results

In this section, the results obtained from the evaluation process are presented. More specifically, to show the efficiency of segmenting the motion data in the proposed phases, two different evaluation studies were conducted. Firstly, by using the motion estimation process that was proposed by Jörg et al. [6], we evaluated the accuracy of the system by correctly estimating the most appropriate gestures. Secondly, since it should be possible to efficiently synthesize each segmented motion, the transition cost for every pair of motions that corresponded to those resulting from the segmentation process was evaluated. Finally, it should be mentioned that, after the proposed segmentation process, by using the 64 gestures (8 repetitions x 8 gesture types) provided by [5], we end with a total of 192 segments. More specifically, 64 motion segments contain the meaningful phase of a gesture, while the remaining 128 segments contain non-gesture phases.

4.1 Evaluating Motion Retrieval

In this case, the results obtained by evaluating the accuracy of the segmentation method are presented. As mentioned, a motion segment must be easily recognizable by the system. Therefore, while the system searches for the most appropriate motion segments, it should be able to minimize any incorrect estimation.

Based on this assumption, we evaluated the proposed segmentation method by assigning the finger motion retrieval to the implementation as proposed by Jörg et al. [6]. The evaluation conducted by using the leave-one-out cross validation process. More specifically, each of the gestures contained in those datasets was left out and the estimation rate of the gesture type was computed. The results obtained from the evaluation process are illustrated in a confusion matrix in Figure 4. In comparing the estimation rate of the proposed methodology, which leads an average of 84%, with the estimation rate of Jörg et al. [6], which lead an average of 80%, the followings should be mentioned. When splitting the motion data into longer segments, the meaning of the gesture is retained. Moreover, the parameters of the distance metric that are compared, which in this case are the wrist position and orientation, result in a discrete separation of the different motions. Therefore, the proposed segmentation method is not only responsible for keeping the required meaning of a single segment, but also for reducing the wrong computation rate. Nevertheless based on the results obtained from the cross validation process, there are issues related to the motion estimation of similar motions. An example is the attention and the small gesture, which appeared in both of those quite similar gesture phases. Therefore, based on those results, we conclude that, even if a better segmentation method enhances the motion estimation process, it is necessary to take into account that the ability to investigate additional parameters of a distance metric is necessary, in order to minimize incorrect estimations.

gesture type of the closest segments (columns); segment taken out of the database (rows)

	att.	big	d-s	OK	PP	small	turn	wipe
att.	68	2	2	12	2	10	2	2
big	3	83	1	1	3	3	6	1
d-s	2	0	87	0	6	2	1	2
OK	9	0	1	83	1	2	3	1
PP	0	1	2	1	95	1	0	0
small	12	4	2	2	0	78	0	2
turn	1	7	0	1	1	0	89	1
wipe	7	0	1	4	0	1	1	86

Figure 4: The class confusion matrix resulting from the leave one out cross validation process. The numbers represent the percentages that a gesture found in the closest segments.

4.2 Evaluating the Transition Cost

The second feature that a segmentation process should possess is the ability to be synthesized efficiently. This means that each pair of corresponding motions should be able to transit smoothly during the motion synthesis process. Thus, a segmentation process should be able to provide the ability to split the motion data into parts that minimize the transition cost.

For that reason, the transition cost of every pair of segmented data based on the proposed methodology, as well as on the methodology used, in [6] was computed. This is the transition cost of each pair of motion sequences, such as from every segment that belongs to phase 1 with any segment that belongs to phase 2, and from every segment that belongs to phase 2 with any segment that belongs to phase 3. Additionally, since each of the meaningful phases (phase 2) will appear after any other segment that belongs to the same phase, an additional calculation is made to compute the transition cost. Thus, we computed the transition cost between the last frame, t_{last}, of an origin segment A and the first frame, t_1, of a target segment B by the following equation:

$$c_T = \frac{1}{D}\sum_{d=1}^{D}(A_d(t_{last}) - B_d(t_1))^2 \qquad (2)$$

where $d = 1, ..., D$ is the total number of degrees of freedom of the character's hand skeleton. Having computed the transition cost function for each pair of phases, the mean average of each pair of phases is then calculated. The results of this evaluation process are summarized in Figure 5. The following should be noted when examining these results. Firstly, the proposed segmentation method minimizes the transition cost between two motion segments comparing to [6], where on average the transition cost has been reduced by 25% while the transition process evolves both from phase 1 to phase 2 and from phase 1 to phase 3. Moreover, the interesting results arise when evaluating the gesture-to-gesture transition, such as while a motion segment from phase 2 is required to transit to a motion segment that belongs to phase 2. In this case, the proposed segmentation methodology achieves a much better transition since the transition cost function is minimized by average 33% of [6]. Therefore, we assume that the proposed method not only ensures a higher estimation rate, but also a smoother transition that will make the motion synthesis process more natural looking.

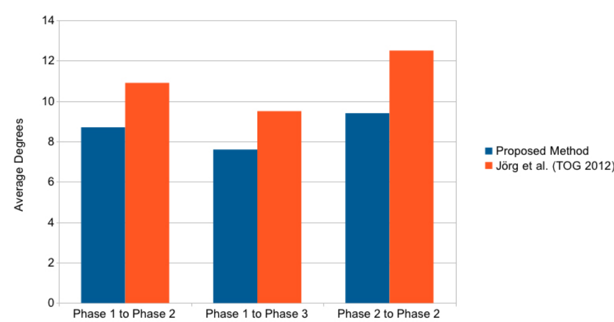

Figure 5: The results obtained from the transition cost evaluation between the proposed segmentation method, and that used in [6].

5 Conclusion

The automatic synthesis of finger motion requires an efficient methodology for segmentation of the motion of the fingers. This methodology should be able to provide the ability to retain the required meaning of the segmented motion, as well as being easily recognized and efficiently synthesized.

Based on those assumptions, a motion segmentation method that retains both of the aforementioned factors has been presented in this paper. To achieve an accurate segmentation method that will count parameters not only by the character's fingers, but also by the character's wrist, an analysis of existing finger motions conducted. This analysis is responsible for showing the possible similarities between the motion of the fingers and the motion of its wrist and, therefore, associating those motions. In conclusion, an efficient segmentation method is based on the ability to segment finger gestures while the motion of the wrist crosses its maximum velocity, instead of the minimum velocity that has been proposed in previous solutions.

Based on the evaluations that conducted, the efficiency of the proposed methodology can be summarized as follows. Firstly, by splitting the motion data with the proposed methodology, the required meaning is kept within the gesture phase of the motion. Moreover, as the results have shown, incorrect estimations are decreased while using the motion synthesis process as proposed by Jörg et al. [6]. Therefore, we can assume that a different segmentation method can work as an optimization of that solution. Finally, an additional evaluation that conducted has shown that the proposed segmentation method improves the synthesis process, especially the motion transition process, since on average it minimizes the transition cost between every two motion segments. Therefore, it is possible to synthesize a more natural transition while the underside jerkiness effect is either minimized or removed entirely.

References

[1] A. Kendon, "Gesture Visible Action as Utterance", *Cambridge University Press*, Cambridge, UK, 2004.

[2] S. Jörg, J. Hodgins, and C. O'Sullivan, "The Perception of Finger Motions", In Proc. of *ACM Symposium on Applied Perception in Graphics and Visualization*, pp. 129-133, 2010

[3] A.-A. Samadani, B. J. Dehart, K. Robinson, D. Kulić, E. Kubican, and R. Gorbet, "A Study of Human Performance in Recognizing Expressive Hand Movements", *IEEE International Symposium on Robot and Human Interactive Communication*, 93-100, 2011

[4] D. McNeill, "Hand and Mind: What Gestures Reveal About Thought", *University of Chicago Press*, USA, 1992

[5] Data-Driven Finger Motion Synthesis for Gesturing Characters (project page), from `http://people.cs.clemson.edu/~sjoerg/fms.html`, accessed 24/02/2014

[6] S, Jörg, J. Hodgins, and A. Safonova, "Data-Driven Finger Motion Synthesis for Gesturing Characters", *ACM Transactions on Graphics* vol. 31, no. 6, Article No. 189, 2012

[7] S. Levine, C. Theobalt, and V. Koltun, "Real-Time Prosody-Driven Synthesis of Body Language", *ACM Transactions on Graphics*, vol. 28, no. 5, Article No. 172, 2009

[8] A. Fod, M. J. Matarić, and O. C. Jenkins, "Automated Derivation of Primitives for Movement Classification", *Autonomous Robots*, vol. 12, no. 1, pp.39-54, 2002

[9] C. Kang, N. Wheatland, M. Neff, and V. Zordan, "Automatic Hand-Over Animation for Free-Hand Motions from Low Resolution Input", In Proc. of *Motion in Games*, pp. 244-253, 2012

[10] N. Wheatland, S. Jörg, and V. Zordan, "Automatic Hand-Over Animation using Principle Component Analysis", In Proc. of *Motion in Games*, pp. 175-180, 2013

[11] G. Jin, and J. Hahn, "Adding Hand Motion to the Motion Capture Based Character Animation", In Proc. of *Advances in Visual Computing*, pp. 17-24, 2005

[12] A. Majkowska, V. Zordan, and P. Faloutsos, "Automatic Splicing for Hand and Body Animations", In Proc. of *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 309-316, 2006

[13] L. Kovar, M. Gleicher, "Flexible Automatic Motion Blending with Registration Curves", In Proc. of *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 214-224, 2003

[14] C. Mousas, P. Newbury, C.-N. Anagnostopoulos, "Efficient Hand-Over Motion Reconstruction", In Proc. of *22nd International Conference on Computer Graphics, Visualization and Computer Vision*, 2014

[15] C. Mousas, P. Newbury, C.-N. Anagnostopoulos, ' Evaluating the Covariance Matrix Constraints for Data-Driven Statistical Human Motion Reconstruction", In Proc. of *30th Spring Conference on Computer Graphics*, 2014