

# Efficient Hand-Over Motion Reconstruction

Christos Mousas     Paul Newbury

Dept. of Informatics  
University of Sussex  
Brighton BN1 9QJ  
United Kingdom

{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

In this paper, a hand-over motion reconstruction methodology is presented. First to be examined in the proposed approach is the way in which the system computes the optimal markerset for a given dataset of hand motion sequences. In a second step before the motion reconstruction process and given the reduced number of markers, the system estimates the remaining markers by computing a simple distance metric. Having the complete number of markers, including both the input and the computed markers, the system reconstructs the motion of the character's fingers. The reconstruction process is formulated in a maximum a posteriori framework, which is responsible for approximating a valid pose of the character's hand, in which the mixture of factor analysis (MFA) clustering techniques was used for the prior learning process. The results show that high quality motions of the character's hand can be reconstructed with the methodology presented.

## Keywords

hand-over animation, motion reconstruction, markerset strategy, mixture of factor analysers, MAP

## 1 INTRODUCTION

Synthesizing the character's hand motion is quite a complex and time-consuming process because the human hand is highly articulated and has many degrees of freedom. This is especially true while the animator deals with the well-known keyframing techniques. Therefore, a variety of solutions for capturing the hand motion were produced in past years. Thus, by capturing the required motion data, it is possible to assemble the necessary hand motion sequences over the character's full-body motion. The result of this is an enhancement of realism of the final motion since the motion of the character's hand appears more natural. Moreover, the meaning of the synthesized postures can be clearly understood as various perceptual studies [Ken04][JHO10][JHO10] indicate.

However, automatic methods for estimating the hand motion while the number of markers during the motion capture process is reduced are important. Thus, during recent years, methodologies of adding hand motion data to an animated character have been proposed. Those methods are generally called hand-over animation. In

hand-over animation, the motion required of the character's hand is transferred automatically, while keeping the required meaning.

Conversely, motion capture systems are able to capture hand motion. However, it is always desirable to be able to capture high-quality motion sequences using as few markers as possible. Thus, a reduction in the actual number of markers to use during the motion capture process should be examined, as has been for full-body motion control in several previous works [CH05][IWZ\*09][LWC\*11]. By reducing the number of the captured markers and using motion reconstruction methods, the system should be able to reconstruct the hand motions. Therefore, it is necessary to find an efficient markerset that will result in high-quality hand-over motion reconstruction.

Based on this requirement, a hand-over motion reconstruction methodology that is able to estimate the most important markers, given a dataset of motion sequences, is presented in this paper. This methodology has been developed by assigning the optimal markerset search process to the reconstruction error produced when any of the markers is omitted. Thus, in a self-evaluation process, the system ranks the markers and removes those that produce fewer errors, thereby providing the optimal solution. However, the missing markers are important for the motion reconstruction process. For that reason, a simple distance metric is introduced that can estimate the positions of the missing markers, given the input and reference data. Having estimated the missing markers, the system

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

reconstructs the motion of the character's hand by assigning this process to a maximum a posteriori (MAP) framework, where a parametric statistical motion model, the mixture of factor analysis (MFA), is used for the prior learning process.

Based on three different datasets - those of gestures, conversation, and the American Sign Language (ASL) - the optimal markersets that are computed with the proposed methodology are presented. Those markersets consist of three, five, and six markers. Moreover, we evaluate the accuracy of the motion reconstruction process. Specifically, we evaluate the reconstruction error between the optimal marker set of a given dataset and the marker set that was computed for any other dataset. Moreover, we evaluate the proposed methodology by using different markersets as those proposed in previous methodologies. The results of this evaluation process show that high quality motion sequences can be obtained while using a reduced number of markers.

The rest of the paper is divided into the following sections: Section 2 presents work related to hand motion capture and hand motion synthesis is presented. Section 3 supplies an overview of the proposed methodology. Section 4 offers an automatic method to search for optimal markersets. Section 5 describes the process of estimating the missing markers. Section 6 presents the motion reconstruction process. Section 7 gives the results obtained by evaluating the reconstruction error of different marker set strategies. Finally, Section 8 draws conclusions and describes potential future work.

## 2 RELATED WORK

The related research on finger animation can be separated into those methodologies and techniques that are responsible for capturing the motions of the fingers and those techniques that are responsible for synthesizing the motion. Capturing the finger motion is highly challenging, since the human hand contains many joints and is highly articulated.

During past years, many different techniques to capture human fingers have been proposed. The most common method is with the use of data-gloves and, hence, some of the most popular techniques proposed by Wang and Popović [WP09]. In this case, by using a color glove and a simple camera, it is possible to execute a valid, virtual hand pose by analysing the color information of the glove's specified colors. On the other hand, finger motion capture has attracted the industrial community. Hence, solutions to capture human hand motion have been developed and include the CyberGlove System [Cyb] the Measurand [Mea], and the Leap [Lea]. The disadvantages of these solutions are their lower accuracy and the drift [KZK04]. Taking advantage of the optical motion capture devices, such as that proposed by Zhao et al. [ZCX12] and Oikonomidis et al. [OKA11]

managed to capture the human's hand motion by using the Microsoft's Kinect motion capture device. Moreover, by using computer vision algorithms, such as the solution proposed by Athitsos and Sclaroff [AS03], one can capture the motion of the human hand.

Conversely, many different approaches for synthesizing finger motion sequences by simplifying the rules of the motion synthesis process have been proposed. Jörg et al. [JHS12] proposed a method that is based on the ability to synthesize a character's hand motion by assigning a weighted variable to the wrist's position and orientation for use as control parameters of the motion synthesis process. In Majkowska et al. [MZF06] finger and body motion are captured separately in a preprocessing stage. Then, during the composition process, those motion sequences are combined, while using spatial and temporal alignment methodologies to determine their correlation to the motion. In general, this technique has the advantage of motion transplantation techniques [vBE12], which were previously examined to identify a method that combines the motions of different body parts to form a new motion sequence.

On the other hand, the solution proposed by Ye and Liu [YL12] generates the motion of the finger based on the wrist's movements and the handled object's specified motion constraints. Specified manipulation strategies are assigned to the fingers while the character's wrists predict that a specified action will occur. Similarly, finger motion synthesis in which the virtual characters call out to interact with musical instruments has been proposed. One of the most recent approaches is one that Zhu et al. proposed [ZRH\*13] It involves assigning specified action to the fingers in conjunction with an ability to execute information from a midi input. On predefined parameters of piano performance, the system generates a valid motion sequence of the virtual character, in which the fingers play an active role in the motion synthesis process. Similarly, the solution proposed by El Koura and Singh [ES03] generates finger motion for specific tasks such as musical instruments.

Finally, physics-based approaches have been used to generate finger motion, especially for tasks of manipulation, such as the solutions proposed by Liu [Liu08][Liu09], Pollard and Zordan [PZ05] and Andrews and Kry [AK13]. Other works, by using sensors to measure the parameters of forces [KP06], attempt to generate correct finger motion. Neff and Seidel [NS06] used the ability to synthesize human hand motion by using relaxed hand shapes, derived from physics-based parameters that were retrieved from video recordings. Another interesting solution for animating detailed and anatomically correct hand and finger motion was proposed by Tsang et al. [TSF05].

The presented hand-over motion reconstruction process was inspired by Wheatland et al. [WJZ13]. How-

ever, evaluating the proposed markerset strategy with [WJZ13] our approach can generate higher quality finger hand motion sequences for different hand motion databases. Moreover, it should be mentioned that the motion reconstruction process follows the established strategy of using existing motion sequences for the prior learning process. Although, instead of performing a separate post-filtering step as in most previous work, such as that in Chai and Hodgins [CH05], we integrate animation prior directly into the tracking optimization using an MAP estimation [MNA14a], as has been previously examined in full-body motion reconstruction [MNA14b]. This is similar in spirit to those of Wei and Chai [WC11] and Liu et al. [LHC11] who use a static pose prior for interactive design of full-body motion sequences. Finally, it should be mentioned that the mixture of factor analysis motion clustering technique was chosen since it scales well with the size of the data set that was used.

### 3 OVERVIEW

The proposed methodology is presented in four parts: the process of searching for the optimal markerset for a given dataset, the missing marker estimation process, prior motion modelling, and the motion reconstruction process. A short explanation of each is presented below. Finally, the pipeline of the proposed methodology is illustrated in Figure 1.

**Searching Optimal Markerset:** One of the basic advances in the proposed methodology is an automatic process that searches the optimal markerset for a given dataset of motion sequences. More specifically, given the required number of markers and the collection of motion data, the system is able to estimate the optimal markerset that enables one to reconstruct the character's hand motion with the least possible error.

**Estimating Missing Markers:** In this step and having a collection of motion data, a distance metric was implemented and evaluated that can estimate the position of any missing marker by combining the position of the markers at the present time step  $t$ , in conjunction with the previously synthesised hand pose at time step  $t - 1$ .

**Prior Motion Modeling:** This process uses the existing motion sequences related to directions and the motion of hand gestures while conversing that were provided by [HMD] and [ASL]. Based on those sequences, the system is responsible for learning an MFA. The MFA model is responsible for measuring the naturalness of the human motion poses and is used to compel the synthesised poses to remain within a natural-looking space.

**Motion Reconstruction:** During the application runtime, the system is able to reconstruct a valid pose of the character's hands. Based on the inputs and

the computer markers, the system automatically measures the global location and orientation of each marker  $\{m_1, \dots, m_n\}$ . Based on that information, the motion reconstruction process is formulated in an MAP framework, which combines the prior term that is enclosed by MFA with the likelihood term that is defined by the markerset.

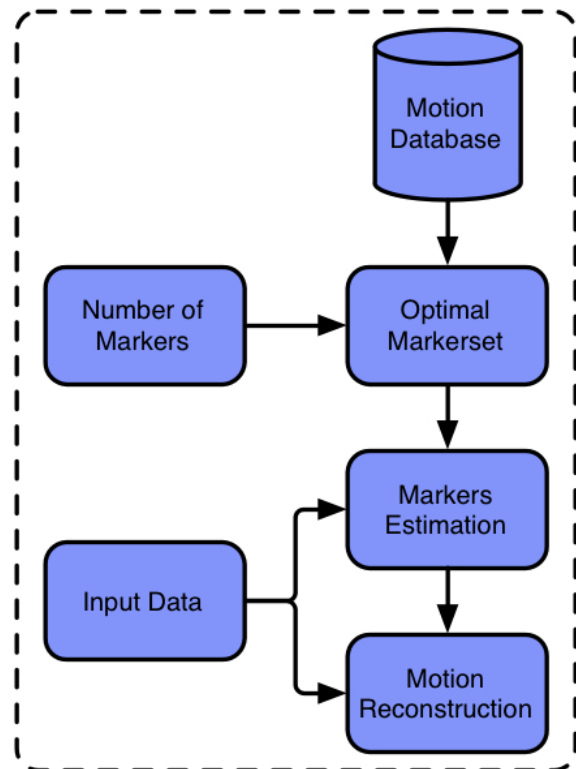


Figure 1: The pipeline of the proposed methodology.

### 4 SEARCHING THE OPTIMAL MARKERSET

Searching for the optimal markerset is quite a complex process, since various combinations of markers' positions can be retrieved. For that reason, we designed an automatic methodology to solve this problem. More specifically, given the initial markerset where the motion data were captured, and the required number of markers, the system is able to find the optimal markerset for each motion database.

The proposed methodology works as follows. First, considering the initial markerset, it is necessary to find the markers that least influence the motion reconstruction process. At each iteration, the marker that has the lowest score is removed until the number of markers that should be used for the motion reconstruction process is reached. It should be noted that, in the dataset that we used, the motion data was captured with  $m_{set} = 13$  markers, as is illustrated in Figure 2.

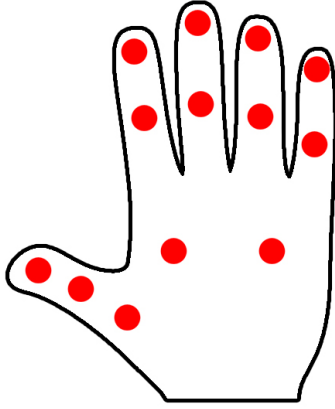


Figure 2: The markerset that was used for the finger motion capture process.

First it is necessary to estimate the influence of each marker given a collection of motion data. In this case, it is first assumed that a marker that moves less, has less influence on the motion reconstruction process. Thus, to find the marker that least influences the motion reconstruction process, the local velocity of each marker is computed for each of the motion sequences contained in the database. In addition, this process sorts the markers in a row, starting from the one of least influence to the one of greatest influence. It should be mentioned that this sorting process in this case is quite useful since it enables the search for the least influential marker to begin at a specific starting point, rather than searching randomly every marker.

On the other hand, the aforementioned assumption should be validated. A verification method is employed that evaluates the influence that the marker has during the reconstruction process. In this case, the verification process is assigned to any of the  $N$  number of markers that the user requested as the number of markers to use for the reconstruction process. The system uses all of the hand postures contained in the database for a self-evaluation. Thus, for all of the postures, it is necessary to calculate the joint's angle distance between the posture that contains the marker, which is the one contained in the database, and the posture produced by a forward kinematics function as it will be produced during the reconstruction process (see Section 6). Based on the validation process, the marker of least influence from the  $N$  number of test markers is removed. This validation concludes with the  $N$  markers that influence most the motion reconstruction process. Algorithm 1 represents the searching process for markers that influence least the motion reconstruction process. Finally, based on this search process we conclude with three different markersets, each of which consists of three, five, and six markers, for each of the different datasets that were used. The resulting markersets are illustrated in Figure 3.

---

**Algorithm 1:** The search process for retrieval of the optimal markerset.

---

**Input:** The target motion data  $M$ , and the number of markers  $N$  defined by the user.

**Output:** The optimal markerset  $Markers$ .

---

```

1 while  $M$  do
2   foreach marker  $m_i$  in  $Markers$  do
3     foreach motion  $m_j$  in  $Motions$  do
4        $v_i = \sum v_j^i$ ;
5        $Markers.sort(v_i)$ ;
6   for  $i=1$  to  $N$  do
7     foreach posture  $p_i$  in  $Postures$  do
8        $d_i = \|p_i^{ref} - p_i^{test}\|^2$ ;
9        $Markers.sort(d_i)$ ;
10      if  $i=N$  then
11        if  $Markers.size()=N$  then
12          break;
13        else
14           $Markers.remove(0)$ ;
15 return  $Markers$ ;
```

---

## 5 ESTIMATING MISSING MARKERS

This section presents the estimation process of any missing marker. The approach is based on the ability to combine the example posture of the hand that is most closed contained in the database for the current time step  $t$ , by combining the knowledge of the previous posture at  $t-1$  that was calculated during the motion reconstruction process.

More specifically, consider an input markerset  $M$  that contains an  $N$  number of markers  $m_i$ , which are represented as  $M = \{m_1, \dots, m_n\}$ . In addition, consider that, for the same markerset as in the previous time step  $t-1$ , the system had reconstructed a hand pose  $q^{t-1}$ . In this case, it is necessary to estimate the pose that is as close as possible to the input markerset taking the advantage of the previous, reconstructed pose. For that reason, we adopt the missing marker estimation process that computes the posture of the character that is closest to the one contained in the database in conjunction with the previous synthesized posture. Thus, we calculate:

$$M_{search} = \frac{1}{n} \sum_{i=1}^n (m_i - m_i^{db})^2 + w_f q^{t-1} \quad (1)$$

where  $m_i^{db}$  is the marker of any posture contained in the database and  $w_f$  is the weight influential factor assigned to the previous reconstructed poses  $q^{t-1}$ . The weight factor based on cross validation that is assigned to  $w_f = 1/3$ . The evaluation of the weight factor is presented in Section 7.



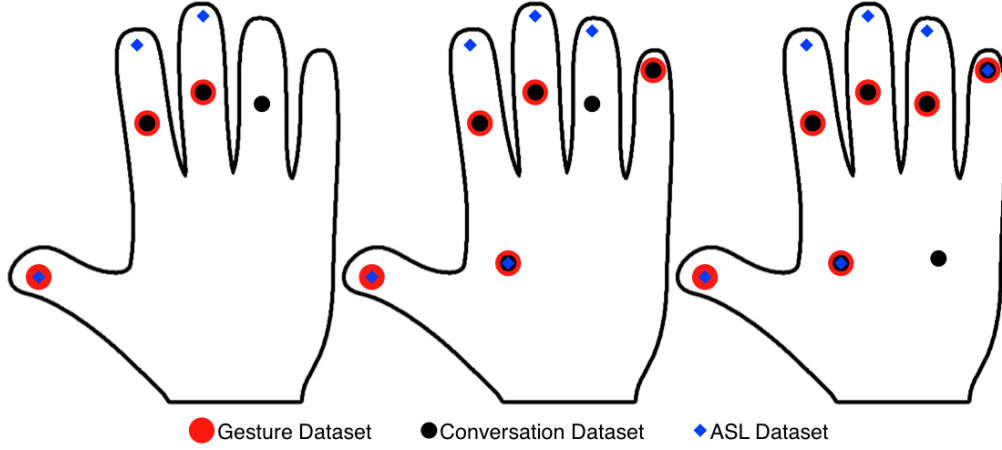


Figure 3: The proposed optimal marker search process estimates the positions of markers. The user requires three, five, and six markers for each of the three different datasets.

Based on Equation 1 the system is able to estimate which of the postures contained in the database is close to the input markerset. Since a meaningful continuity is required, it is possible to estimate the closest posture to the input one by assigning the previous reconstruction posture to the distance metric. Therefore, we conclude our estimation of the remaining markers that were not used during the motion capture process by assigning the estimated markers of the closest hand posture to the current marker set. An example of the estimation process is shown in Figure 4.

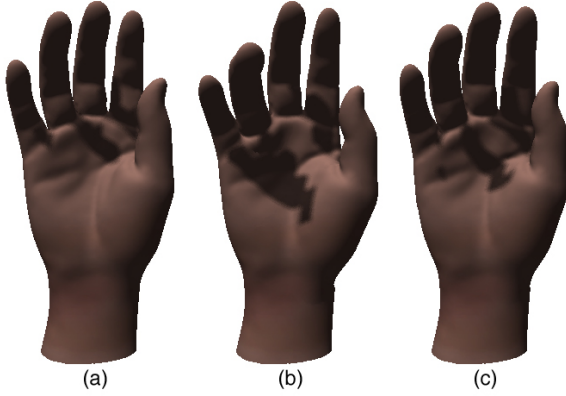


Figure 4: Given an input markerset, the system estimates the closest posture (a). Since the closest posture may provide undesired discontinuities during the motion reconstruction process, by combining the knowledge of the previous reconstructed posture (b), the system retries the new closest posture (c).

## 6 MOTION RECONSTRUCTION

The system reconstructs a pose of the user's hand  $q^*$  based on the extracted hand features,  $c_t$ , that are defined as user-specified constraints at the current  $t$ -th frame. This is achieved by combining the current control features  $c_t$  that contain both the input and the estimated

markers like those computed in Section 5. It is represented as  $c_t = \{m_1, \dots, m_n\}$ , and the constructed probabilistic model of the previous  $j$  reconstructed poses  $\tilde{Q} = [\tilde{q}_{t-1}, \dots, \tilde{q}_{t-m}]$ . The system reconstructs the current pose  $q^*$  in a constrained MAP framework. Therefore, the motion synthesis process is represented as:

$$q^* = \arg \max_{q_t} p(q_t | c_t, \tilde{Q}) \quad (2)$$

where  $p(\cdot|\cdot)$  denotes the conditional probability, and by using Bayes' rule it is obtained:

$$q^* = \arg \max_{q_t} p(c_t | q_t, \tilde{Q}) p(q_t, \tilde{Q}) \quad (3)$$

Assuming that  $q_t$  is conditionally independent of  $\tilde{Q}$  and given  $c_t$ , one obtains the following:

$$q^* \approx \arg \max_{q_t} p(c_t | q_t) p(q_t, \tilde{Q}) \quad (4)$$

In this case, by applying the negative logarithm to the posterior distribution function,  $p(q_t | c_t, \tilde{Q})$ , the constrained MAP problem is converted into the following energy minimisation problem:

$$q^* = \arg \min_{q_t} \underbrace{-\ln p(c_t | q_t)}_{E_{\text{likelihood}}} + \underbrace{-\ln p(q_t, \tilde{Q})}_{E_{\text{prior}}} \quad (5)$$

where the first term ( $E_{\text{likelihood}}$ ) measures how well the reconstructed pose  $q_t$  of the character fits the markers  $c_t$  at the  $t$ -th frame. The term ( $E_{\text{prior}}$ ) describes the prior distribution of the human motion data.

### 6.1 Prior Distribution

Each human's hand pose in the database is represented as a 25-dimensional vector  $q \in R^{25}$  in the joint angle space, from which each hand's root position and orientation (wrist position and orientation) are excluded.

The prior distribution involves models using MFAs. In general, the MFAs describe the high-dimensional pose of the human hand's space with a probabilistic combination of different local regions, where each of those regions is modelled by an FA with a small number of latent variables. The MFA provides a probability density function  $p(q_t, \tilde{Q})$  for the entire space.

### 6.1.1 Factor Analysis

Factor analysis (FA) is a parametric statistical model that represents high-dimensional  $q \in R^D$  data using a low-dimensional vector of hidden factors  $s \in R^d$  and a multivariate Gaussian Random vector  $u \in R^D$ . Therefore, a generative model for the factor analyzers can be represented mathematically as:

$$q = \Lambda s + u \quad (6)$$

where  $\Lambda \in R^{D \times d}$  is a factor loading matrix. The covariance matrix is constrained to be in the following form:

$$\Sigma = \Psi + \Lambda \Lambda^T \quad (7)$$

Each column of  $\Lambda$  can be associated with a latent variable. In the diagonal matrix  $\Psi$ , the variance of each data coordinate is modelled separately, and an additional variance is added in the directions spanned by the columns of  $\Lambda$ . The covariance matrix is specified by the number of parameters equal to  $O(Dd)$ . To solve the mixture of factor analyzers (MFA), the expectation-maximization (EM) algorithm for mixtures of factor analyzers is used (see Reference [GH96] for an explanation of how the EM is applied). The determinant and the inverse of the covariance matrix are efficiently computed using two identities as follows:

$$|A + BC| = |A| \times |I + CA^{-1}B| \quad (8)$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)DA^{-1} \quad (9)$$

Using these identities, the inverses and determinants of  $d \times d$  and diagonal matrices, rather than the full  $D \times D$  matrices, are computed. This results in a lower performance of the system. Thus, the equations are computed as follows:

$$|\Psi + \Lambda \Lambda^T| = |\Psi| \times |I + \Lambda^T \Psi^{-1} \Lambda| \quad (10)$$

$$(\Psi + \Lambda \Lambda^T)^{-1} = \Psi^{-1} - \Psi^{-1} \Lambda (I + \Lambda^T \Psi^{-1} \Lambda)^{-1} \Lambda^T \Psi^{-1} \quad (11)$$

Considering the covariance matrix  $\Sigma$  and the mean vector  $\mu$ , the distribution of each model can be computed based on the following:

$$p(q) = N(q | \mu, \Psi + \Lambda \Lambda^T) \quad (12)$$

However, because the inputs are human hand poses that are contained in a database, the mixture model must be considered because all human hand poses form a non-linear manifold in the character configuration space. Thus, a global linear parametric model is often not sufficient to capture the non-linear structure of natural human motion. Therefore, a better solution is to use a mixture model, as presented in the following subsection.

### 6.1.2 Mixture of Factor Analyzers

The mixture models probabilistically partition the entire configuration space into multiple local regions and then model the data distribution in each local region using a weighted variable for each region. The mixture model of the previously mentioned methodologies to constrain the covariance matrix is represented mathematically as follows:

$$p(q) = \sum_{k=1}^K \pi_k N(q | \mu_k, \Psi + \Lambda_k \Lambda_k^T) \quad (13)$$

Using this formulation, the prior in Equation 5 is defined as a weighted combination of  $K$  Gaussians and is represented as:

$$p(q_t, \tilde{Q}) = \sum_{k=1}^K \pi_k N(q_t, \tilde{Q} | \mu_k, \Psi + \Lambda_k \Lambda_k^T) \quad (14)$$

where  $\pi_k$ ,  $\mu_k$  and  $\Sigma_k$  denote the weighted scalar of the  $k$ -th component of the mixture model, the mean vector value and the covariance matrix, respectively, for each of the four restriction processes of the covariance matrix. In general, the goal of the model learning process is to automatically find the model parameters  $\pi_k$ ,  $\mu_k$  and  $\Sigma_k$  for  $k = 1, \dots, K$  from the training data  $q_n$ ,  $n = 1, \dots, N$ , where  $N$  is the number of poses in the database. For each different constrained model that uses these variables, the model parameters are calculated separately using the EM algorithm. In these experiments, the number of mixture components,  $K$ , that is used is set to  $K = 50$ , and the dimension of the latent space  $d$  is set to  $d = 5$ . The values of  $K$  and  $d$  are determined empirically. However, cross-validation techniques can also be used to set the appropriate values for these parameters.

## 6.2 Likelihood Distribution

The likelihood distribution is responsible for estimating how well the locations of the corresponding joint in the reconstructed hand pose fit the input parameters obtained from the user-defined constraints. Thus, by assuming a Gaussian noise with a standard deviation of  $\sigma_d$ , the likelihood function is computed as:

$$E_{likelihood} = -\ln p(c_t | q_t) \propto \frac{\|f(q_t; s) - c_t\|^2}{2\sigma_d^2} \quad (15)$$

where the vector  $q_t$  represents the reconstructed pose of the character's hand at each time step,  $s$  denotes the hand's skeletal size, and  $c_t$  is the observed data obtained from the user-specified constraints. The function  $f$  is the forward kinematics function that calculates the global coordinates value of the current pose  $q_t$ .

## 7 EVALUATION AND RESULTS

This section presents the results obtained from the evaluation process of the proposed hand-over motion reconstruction process. More specifically, we first evaluated the marker estimation process in the following subsection as it was presented in Section 5. Moreover, we evaluated the motion reconstruction process, as presented in Section 6, while using the optimal markersets that were computed in Section 4. Since different optimal markersets are computed for each dataset, we evaluate the reconstruction error of a markerset computed for a given dataset, while using the markersets that were computed from any other dataset. In this case, it should be noted that, for the evaluation process, several databases with hand motion sequences were used that can be found in [HMD] and [ASL]. Specifically, three different databases are used for the testing process. It consists of motion sequences related to conversations, gestures, and the ASL.

### 7.1 Evaluating Marker Estimation

In this subsection, the distance metric that was used to estimate the position of the missing markers is evaluated. For the evaluation process we assigned different weighted values to the  $t - 1$  reconstructed pose, which is represented as  $q^{t-1}$ . Thus, to calculate the influence of the weight factor  $w_f$  we used the leave-one-out cross validation process. More specifically, the quality of the hand motion reconstruction is evaluated by omitting one hand posture of the motion capture data from each database as the testing posture. Then, for each of the five different markersets, as well as for different values of  $w_f$  the reconstruction error was evaluated by computing the difference in angles between each reference hand posture and the one reconstructed with the proposed solution. The results of the evaluation process are illustrated in Figure 5. Thus, as those results indicate, the optimal weight that minimizes the reconstruction error is approximated by  $w_f = 1/3$ .

### 7.2 Evaluating Markersets

In this section, it is presented results that held out while evaluating the optimal markersets that were computed in Section 4. As the method is able to estimate an optimal markerset for a given dataset of motion sequences, it is necessary to show the efficiency of the proposed methodology. For that reason, the reconstruction error was computed while using the optimal markerset for

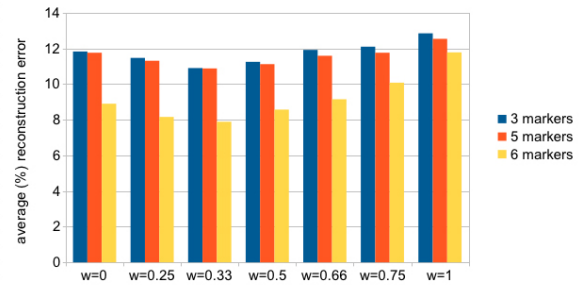


Figure 5: Results obtained while searching the optimal weighted value  $w_f$  for each markerset.

each dataset. Moreover, the reconstruction error was evaluated against the markersets of those that resulted from any of the other datasets. The results for each of the markersets obtained from this evaluation process are summarized in Table 1.

To show the efficiency of the proposed methodology, the reconstruction error was evaluated by using different markersets that were resulted from previous solutions. More specifically, the methodology is evaluated against the markersets proposed by Wheatland et al. [WJZ13] which consist of three and six markers respectively, the markersets that resulted from manual selection based on perceptual studies by Hoyet et al. [HRM12], and the markerset that resulted from the cluster pose error method as proposed by Kang et al. [KWN\*12]. Each of those markerset strategies is illustrated in Figure 6, and the results of the evaluation process are shown in Table 2.

As the results show, the proposed methodology provides an optimal markerset since the reconstruction error is minimal in comparison to any other previously proposed markerset. Moreover, the great advantage of the proposed methodology is the ability to reconstruct efficiently the required motion sequences while using five, instead of six, markers. More specifically, the reconstruction error that was obtained from the markerset proposed by Wheatland et al. [WJZ13] (see Figure 6(d)), which uses six markers, is greater than obtained with the proposed markerset (see Figure 3), which uses five markers. Therefore, the advantage of the proposed methodology is its ability to minimize the reconstruction error while using even fewer markers. Thus, the proposed methodology provides a better solution for reconstructing the character's hand motion. Finally, examples of hand postures that have been reconstructed from different marker sets based on the proposed methodology are illustrated in Figure 7.

## 8 CONCLUSIONS AND FUTURE WORK

One of the key issues in computer animation and motion capture research is the ability to capture high-

<b>DB \ Markerset</b>	<b>Conversation</b>	<b>Gesture</b>	<b>ASL</b>
<b>3 Markers</b>			
<b>Conversation</b>	9.15%	11.07%	10.96%
<b>Gesture</b>	10.22%	10.72%	11.28%
<b>ASL</b>	11.63%	13.26%	10.24%
<b>5 Markers</b>			
<b>Conversation</b>	8.73%	10.21%	10.09%
<b>Gesture</b>	9.36%	8.47%	10.31%
<b>ASL</b>	10.11%	11.89%	8.64%
<b>6 Markers</b>			
<b>Conversation</b>	7.05%	8.13%	8.45%
<b>Gesture</b>	8.12%	6.01%	8.92%
<b>ASL</b>	8.83%	9.29%	6.24%

Table 1: The reconstruction error while using both the optimal markerset for a given dataset, and the markerset that resulted from the other dataset.

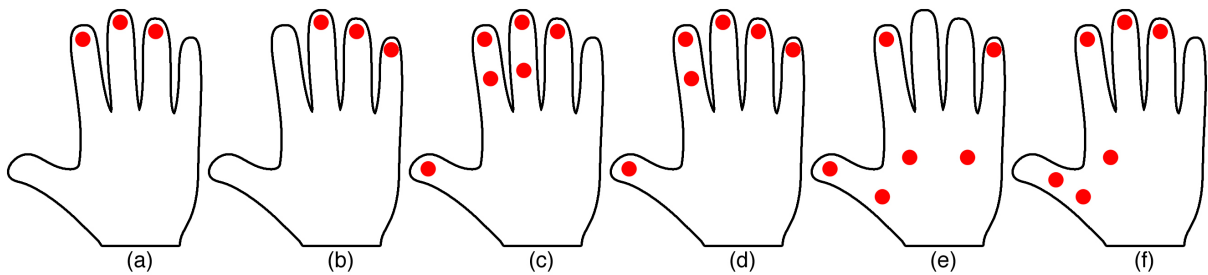


Figure 6: The markersets that were proposed in previous solutions. (a), (b), (c), and (d) are the markersets that Wheatland et al. proposed [WJZ13]. Specifically, (a) and (c) represent the markersets while using the gesture dataset, whereas (b), and (d) represent the markersets while using the ASL dataset. The manual selection markerset (e), which is based on perceptual studies, resulted from the research conducted by Hoyet et al. [HRM12]. Finally, (f) represents the marker sets that was proposed by Knag et al. [KWN\*12] it is retrieved by use of the cluster pose error method.

<b>DB \ Markerset</b>	<b>Conversation</b>	<b>Gesture</b>	<b>ASL</b>
<b>(a)</b>	10.22%	12.37%	13.12%
<b>(b)</b>	9.96%	12.07%	12.38%
<b>(c)</b>	9.41%	10.59 %	10.96%
<b>(d)</b>	8.54%	10.03 %	9.98 %
<b>(e)</b>	11.03%	11.87 %	12.16%
<b>(f)</b>	9.23%	10.49 %	10.74%

Table 2: The reconstruction error that was computed for each dataset that was examined while using the markersets proposed in previous methodologies. (a) to (d), indicates that the markerset strategies were those illustrated in Figure 6.

quality motion sequences using fewer markers. Hence, over the past several years, various solutions for reconstructing human motion based on fewer markers have been proposed. However, because a realistic representation of animated sequences requires detailed motions, methodologies to approximate valid, human motions for specified body-parts should be examined.

In this paper, a methodology for reconstruction of valid poses of a character's hand was examined. Specifically, the advantage of the proposed methodology is first the automatic estimation of the most active markers based on implementation of a simple algorithm. Sec-

ond, based on existing motion data, the system using the knowledge of the previous,  $t - 1$ , reconstructed pose of the character's hand, estimates the position of the remaining markers based on a simple distance metric. Finally, the proposed methodology reconstructs the hand poses by assigning the motion estimation process to a maximum a posteriori framework, which ensures a smooth transition between hand poses.

Based on this approach, the so-called hand-over animation technique can reconstruct high-quality hand motion sequences. Therefore, the proposed approach can be quite beneficial, especially in cases where fewer





Figure 7: Given a reference input posture (a), the system reconstructs the character's hand motion while it uses three (b), five (c), or six (d) markers.

markers are used in the motion capture process. In addition, the benefit of such a technique is based on minimizing the time required to synthesize the desired motion of the character's hand, because such a method provides the desired result automatically.

Finally, we assume that the search process to find the optimal markersets can be beneficial in various cases. For example, it will be beneficial in computing the optimal markerset for the full-body reconstruction process. Thus, in our future work we would like to examine the possibility of reducing the actual number of markers that are used for reconstructing full-body motion sequences, by computing the optimal markersets for different actions that can be reconstructed.

## 9 REFERENCES

- [AK13] Andrews, S., Kry, P. G. Goal Directed Multi-Finger Manipulation: Control Policies and Analysis. *Computers & Graphics*, 37(7), pp. 830-839, 2013.
- [AS03] Athitsos, V., Sclaroff, S. Estimating 3D Hand Pose from a Cluttered Image. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 432-439, 2003.
- [CH05] Chai, J., Hodgins, J. K. Performance Animation from Low-Dimensional Control Signals. *ACM Transactions on Graphics*, 24(3), pp. 686-696, 2005.
- [ASL] CUNY ASL Motion-Capture Corpus, from <http://latlab.cs.qc.cuny.edu/corpus/>, accessed 27/03/2014.
- [Cyb] Cyberglove Systems, from <http://www.cyberglovesystems.com/products/cyberglove-ii/overview>, accessed 27/03/2014
- [ES03] ElKoura, G., Singh, K. Handrix: Animating the Human Hand. In *Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 110-119, 2003.
- [GH96] Ghahramani, Z., Hinton, G. E. The EM Algorithm for Mixtures of Factor Analyzers. Technical Report CRG-TR-96-1, University of Toronto, 1996.
- [HMD] Hand Motion Database, Clemson University, from <http://people.cs.clemson.edu/~sjoerg/fms.html>, accessed 27/03/2014.
- [HRM12] Hoyet, L., Ryall, K., McDonnell, R., O'Sullivan, C. Sleight of Hand: Perception of Finger Motion from Reduced Marker Sets. In *Proc. of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pp. 79-86, 2012.
- [IWZ\*09] Ishigaki, S., White, T., Zordan, V. B., Liu, C. K. Performance-Based Control Interface for Character Animation. *Transactions on Graphics*, 28(3), Article No. 61, 2009.
- [JHO10] Jörg, S., Hodgins, J., O'Sullivan, C. The Perception of Fnger Motions. In *Proc. of ACM Symposium on Applied Perception in Graphics and Visualization*, pp. 129-133, 2010.
- [JHS12] Jörg, S., Hodgins, J., Safonova, A. Data-Driven Finger Motion Synthesis for Gesturing Characters. *ACM Transactions on Graphics* 31(6), Article No. 189, 2012
- [KZK04] Kahlesz, F., Zachmann, G., Klein, R. Visualfidelity Dataglove Calibration. In *Proc. of Computer Graphics International*, pp. 403-410, 2004.
- [KWN\*12] Kang, C., Wheatland, N., Neff, M., Zordan, V. Automatic Hand-Over Animation for Free-Hand Motions from Low Resolution Input. In *Proc. of Motion in Games*, pp. 244-253, 2012.
- [Ken04] Kendon, A. *Gesture - Visible Action as Utterance*. Cambridge University Press, Cambridge, UK, 2004.
- [KP06] Kry, P. G., Pai, D. K. Interaction Capture and Synthesis. *ACM Transactions on Graphics*, 25(3),

- p. 872-880, 2006.
- [Lea] Leap Motion, from <https://www.leapmotion.com/>, accessed 27/03/2014.
- [Liu08] Liu, C. K. Synthesis of Interactive Hand Manipulation. In Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 163-171, 2008.
- [Liu09] Liu, C. K. Dextrous Manipulation from a Grasping Pose. ACM Transactions on Graphics, 28(3), Article No. 59, 2009.
- [LHC11] Liu, H., He, F., Cai, X., Chen, X., Chen, Z. Performance-Based Control Interfaces Using Mixture of Factor Analyzers. The Visual Computer, 27(6-8), pp. 595-603, 2011.
- [LWC\*11] Liu, H., Wei, X., Chai, J., Ha, I., Rhee, T. Realtime Human Motion Control with a Small Number of Inertial Sensors. In Proc. of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 133-140, 2011.
- [MZF06] Majkowska, A., Zordan, V. B., Faloutsos, P. Automatic Splicing for Hand and Body Animations. In Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 309-316, 2006.
- [Mea] Measurehand, from [www.shapehand.com](http://www.shapehand.com), accessed 27/03/2014.
- [MNA14a] Mousas, C., Newbury, P., Anagnostopoulos, C.-N. Evaluating the Covariance Matrix Constraints for Data-Driven Statistical Human Motion Reconstruction. In Proc. of 30th Spring Conference on Computer Graphics, 2014.
- [MNA14b] Mousas, C., Newbury, P., Anagnostopoulos, C.-N. Data-Driven Motion Reconstruction Using Local Regression Models. In Proc. of 10th International Conference on Artificial Intelligence Applications and Innovations, 2014.
- [NS06] Neff, M., Seidel, H.-P. Modeling Relaxed Hand Shape for Character Animation. In Proc. of Articulated Motion and Deformable Objects, 262-270, 2006.
- [OKA11] Oikonomidis, I., Kyriazis, N., Argyros, A. Efficient Model-Based 3D Tracking of Hand Articulations Using Kinect. In Proc. of British Machine Vision Conference, pp. 1-11, 2011.
- [PZ05] Pollard, N. S., Zordan, V. B. Physically Based Grasping Control from Example. In Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 311-318, 2005.
- [SDR\*11] Samadani, A.-A., Dehart, B. J., Robinson, K., Kulić, D., Kubican, E., Gorbet, R. A Study of Human Performance in Recognizing Expressive Hand Movements. In Proc. of IEEE International Symposium on Robot and Human Interactive Communication, pp. 93-100, 2011.
- [TSF05] Tsang, W., Singh, K., Fiume, E. Helping Hand: An Anatomically Accurate Inverse Dynamics Solution for Unconstrained Hand Motion. In Proc. of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 319-328, 2005.
- [vBE12] van Basten, B., Egges, A. Motion Transplantation Techniques: A Survey. IEEE Computer Graphics and Applications 32(3), pp. 16-23, 2012.
- [WP09] Wang, R. Y., Popović, J. Real-Time Hand-Tracking with a Color Glove. ACM Transactions on Graphics, 28(3), Article No. 63, 2009.
- [WC11] Wei, X. K., Chai, J. Intuitive Interactive Human-Character Posing with Millions of Example Poses. IEEE Computer Graphics and Applications, 31(4), pp. 78-88, 2011.
- [WJZ13] Wheatland, N., Jörg, S., Zordan, V. Automatic Hand-Over Animation Using Principle Component Analysis. In Proc. of ACM SIGGRAPH Motion in Games, pp. 175-180, 2013.
- [YL12] Ye, Y., Liu, C. K. Synthesis of Detailed Hand Manipulations Using Contact Sampling. ACM Transactions on Graphics 31(4), Article No. 41, 2012.
- [ZCX12] Zhao, W., Chai, J., Xu, Y. Q. Combining Marker-Based Mocap and RGB-D Camera for Acquiring High-Fidelity Hand Motion Data. In Proc. of ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 33-42, 2012.
- [ZRH\*13] Zhu, Y., Ramakrishnan, A. S., Hamann, B., Neff, M. A System for Automatic Animation of Piano Performances. Computer Animation and Virtual Worlds, 24(5), pp. 445-457, 2013.