Data Driven Gesture Model Acquisition using Minimum Description Length

Michael Walter † Alexandra Psarrou † and Shaogang Gong ‡
† Harrow School of Computer Science, University of Westminster,
Harrow HA1 3TP, U.K. zeoec,psarroa@wmin.ac.uk
‡ Dept. of Computer Science, Queen Mary and Westfield College,
London E1 4NS, U.K. sgg@dcs.qmw.ac.uk

Abstract

An approach is presented to automatically segment and label a continuous observation sequence of hand gestures for a complete unsupervised model acquisition. The method is based on the assumption that gestures can be viewed as repetitive sequences of atomic components, similar to phonemes in speech, starting and ending in a rest position and governed by a high level structure controlling the temporal sequence. It is shown that the generating processes for the atomic components and derived gesture models can be described by a mixture of Gaussian in their respective component and gesture space. Mixture components modelling atomic components and gestures respectively are determined using a standard EM approach, while the determination of the number of mixture components and therefore the number of atomic components and gestures is based on an information criterion, the Minimum Description Length (MDL).

1 Introduction

Natural gestures are expressive body motions with underlying spatial and in particular temporal structure. The temporal structure of gestures can be modelled as stochastic processes under which salient phases are modelled as states and prior knowledge on both state distributions and observation covariances is learned from training examples [3, 13, 16, 8, 17]. However, the collection of training examples as well as the determination of states requires temporal segmentation and alignment of gestures. This task is ill-conditioned due to measurement noise, non-linear temporal scaling based on variations in speed and most notably human variation in performing of gesture. As a result the segmentation in gesture recognition typically involves manual intervention and hand labelling of image sequences.

This paper presents a method to automatically segment and cluster continuous observation sequences of natural gestures for a complete unsupervised acquisition of gesture models, using only contextual information derived from the observation sequence itself. Our work is motivated by recent work in the field of Natural Gestures that identified two basic gesture types. Gestures based on two movement phases, away from a rest position into gesture space and back to the rest position and gestures based on three movement phases, away from the rest position into gesture space (preparation), followed by a small movement with hold (stroke) and back to the rest position (retraction) [9]. Our approach makes the assumption that gestures can be viewed as a recurrent sequence of atomic components, similar to phonemes in speech, starting and ending in rest positions and

BMVC 2001 doi:10.5244/C.15.69

governed by a high level structure controlling the temporal sequence. The extraction of gestures consequently involves in a first step the segmentation of the complete observation sequence and the extraction of atomic components and in a second step the identification of rest positions and the sequence of enclosed components. We show that atomic components and derived gestures once normalised and projected into their respective component and gesture space form clusters. Both distributions can be described by mixtures of Gaussian, where each mixture component models a different atomic component or gesture respectively. Consequently, the determination of atomic components and gesture models requires the determination of an optimal number of mixture components, known as the problem of model order selection, and the estimation of the model parameters.

Maximum likelihood methods such as k-means [6] or Expectation-Maximisation (EM) [5] provide effective tools for the determination of mixture components. However, the resulting mixture models depend on the *a priori* knowledge of the number of mixtures. The model order can be determined using constructive algorithms that employ cross validation techniques for model training [10]. However the disadvantage of such methods is that they require a validation set, which is often not available. Alternative approaches to determine the number of clusters are based on information criteria, such as A Information Criterion (AIC) [1], Bayesian Information Criterion (BIC) [14] and Minimum Description Length (MDL) [11]. In the following sections we show how MDL can be used to automatically segment the components within gesture space into clusters that correspond to atomic gestures, without any *a priori* knowledge on the number of atomic components present. We extract high level knowledge on gestures based on the assumption that gestures can be described as sequence of atomic components, starting and ending in a rest position. Rest positions as well as the number of gesture models present in an observation sequence are identified using MDL.

The rest of this paper is organised as follows. Section 2 describes the temporal segmentation of gestures and their partitioning into atomic components. Section 3 describes the component space representation. Section 4 shows how to use MDL for the automatic clustering of atomic components in component space. Section 5 shows how to derive high level knowledge on gesture models. Experiments on data driven clustering for gesture model acquisition are described in Section 6 and the paper concludes in Section 7.

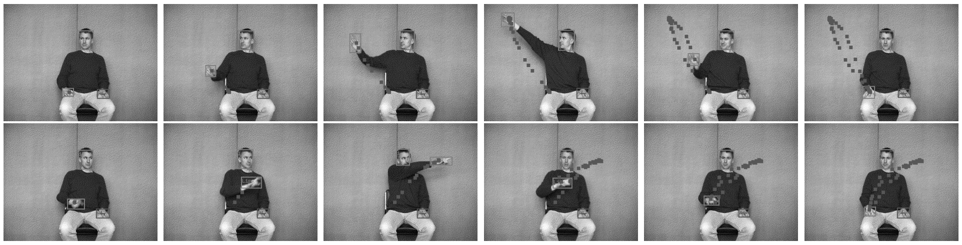

Figure 1: Deictic gestures: *"pointing left"* (top row), *"pointing right"* (bottom row)

2 Temporal Segmentation

Temporal segmentation partitions a continuous observation sequence into plausible atomic components. Our approach is motivated by recent work in the field of Natural Gestures [9] that has identified five basic hand gesture types, iconic, metaphoric, cohesive, deictic and beat gestures. All gestures have their temporal signature in common. Gestures

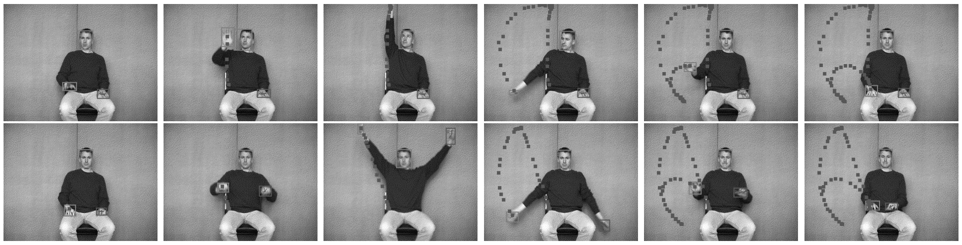

Figure 2: Metaphoric gestures: *"he bent a tree"* (top row), *"there was a big explosion"* (bottom row)

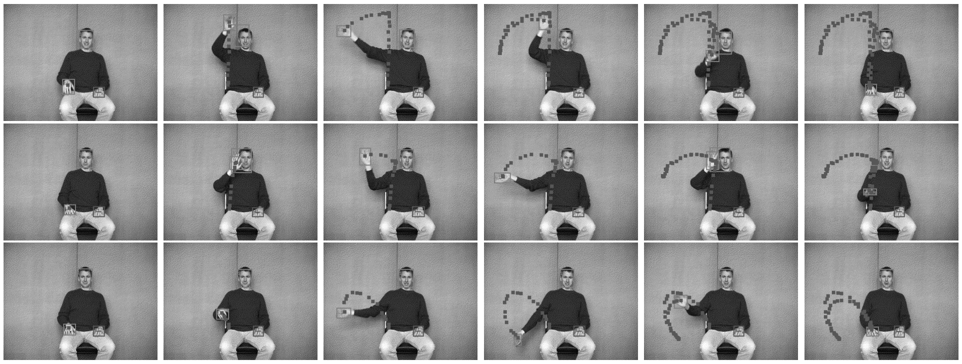

Figure 3: Communicative gestures: *"waving high"* (top row), *"waving low"* (middle row) and *"please sit down"* (bottom row)

are typically embedded by the hands being in a rest position and can be divided into either bi-phasic or tri-phasic gestures. Beat and deictic gestures are examples for bi-phasic gestures. They have just two movement phases, away from the rest position into gesture space and back again, while iconic metaphoric and cohesive gestures have three, preparation, stroke and retraction. They are executed by transitioning from a rest position into gesture space (preparation), this is followed by a small movement with hold (stroke) and a movement back to the rest position (retraction). Examples are shown in Figure 1, Figure 2 and Figure 3.

The complete gesture sequence, is recorded as a continuous sequence of $2D$ vertices, containing the x and y positions of a person's moving hand in an image plane. Segmentation is performed in two steps. In a first step, the complete observation sequence is analysed for segments where the velocity drops below a pre-set threshold to identify rest positions and pause positions that typically occur in bi-phasic gestures between transition into and out of gesture space and in tri-phasic gestures between stroke and retraction. A second step analyses the segments for discontinuities in orientation to recover strokes. We adopt a method based on Asada and Brady's Curvature Primal Sketch [2], depicted in Figure 4.

3 Component Space Representation

Each atomic component extracted from the trajectory of a person's moving hand, consists of c 2D vertices $v_c = [x_1, y_1, x_2, y_2, ..., x_c, y_c]$ with each component having a different number of vertices c. Clustering algorithms, however work on d-dimensional sets of N

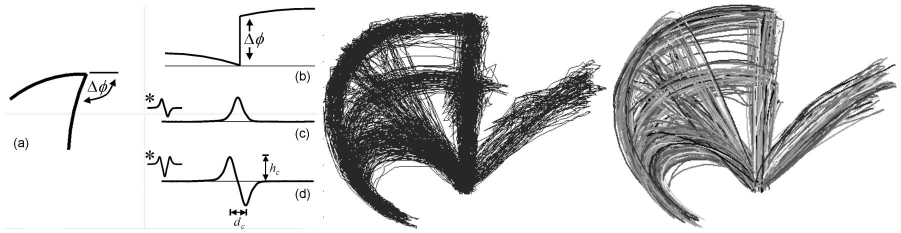

Figure 4: Detecting discontinuities in gesture trajectories. The orientation of a two dimensional hand trajectory $f(t)$ is convoluted with the first N'_σ and second derivative N''_σ of a Gaussian $N_\sigma(t) = (1/(\sqrt{2\pi}\sigma)exp(-t^2/2\sigma^2)$ at different temporal scales $\sigma = \{\sigma_{min} \cdots \sigma_{max}\}$. The filter responses are analysed for characteristic maxima and zero crossings. Only discontinuities consistent over a large scale are registered, thus taking care of noise on different levels. a) Example trajectory containing a curvature discontinuity $\Delta\Phi$. b) The trajectory in orientation space, relating the orientation of the curve to the arc length along the curve. c) Filter response $N'_\sigma * f$ of the orientation of the trajectory $f(t)$ convoluted with the first derivative of a Gaussian $N_\sigma(t)$. d) Filter response $N''_\sigma * f$ of the orientation of the trajectory $f(t)$ convoluted with the second derivative of a Gaussian $N_\sigma(t)$. As shown in d) corners give rise to a pair of peaks with a separation $d_c \approx 2\sigma$ and height $h_c \approx |\Phi|/(\sqrt{2e\pi}\sigma^2)$. Note, d_c is linearly dependent on the scale constant σ and monotonically decreases with σ, which provides a strong clue for the detection of corners. Example of a continuously recorded hand trajectory (middle). The recorded trajectory approximated by a spline and segmented into 701 atomic components (right).

input vectors $Z = [z_1, z_2, ..., z_N]$. This requires to transform the atomic components into a normalised representation, termed ”component space” (Figure 6). The transformation consists of three steps. First, the number of 2D vertices is normalised. The atomic components are approximated by splines, interpolated into d vertices and stored as $2d$-dimensional vector $z_i = [x_1, y_1, x_2, y_2, ..., x_d, y_d]$. Second, each vector z_i is concatenated with a scale factor $s_i = (c_i - c_{min})/(c_{max} - c_{min})$, the ratio of the original number of vertices minus the minimal number and the maximal number minus the minimal number of vertices to preserve information on the original length. Third, redundant dimensions are removed using Principal Component Analysis (PCA). Looking at Figure 6 we can see that atomic components projected into component space form clusters, which can be approximated by a mixture of K Gaussian, defined by mixture coefficients $W_k = [w_1, ..., w_k]$ and d-dimensional means and covariances $\Theta_k = [\mu_1, ..., \mu_k; \Sigma_1, ..., \Sigma_k]$.

$$f(z_i|W_K, \Theta_K) = \sum_{k=1}^{K} w_k N(z_i, \mu_k, \Sigma_k) \tag{1}$$

Assuming each mixture component corresponds to an atomic behaviour allows us to equate the determination of the mixture components with the determination of the atomic components itself. There is a considerable amount of literature on the estimation of mixture parameters and standard maximum likelihood methods such as k-means and EM can be used to determine the values of Σ_k, μ_k and w_k for a known model order K. There are no methods to determine the number K of parameters directly, however iterative procedures based on information criteria can be used as described in Section 4.

3.1 Removing Noise from Component Space

Not all components projected into atomic component space are part of meaningful gestures and this reinforces the common criticism of k-means and related maximum likelihood methods that they do not address the problem of noise and assign all input elements

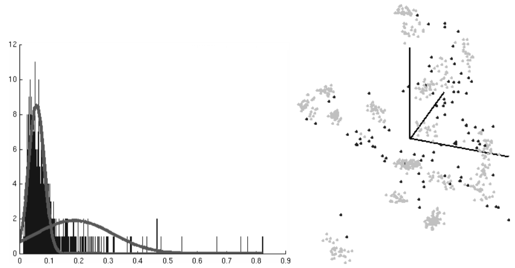

Figure 5: The average Euclidean distance calculated for all components to the nearest 15 neighbours, approximated by a mixture of two Gaussian (left).The projection of the 3 largest Principal Components of each atomic component, classified as feature (light colour) or clutter (dark colour) (right).

to one particular class. Consequently we filter the distribution based on an approach similar to that of Bayers and Raftery [12]. We assume that features and clutter can be described by two superimposed random processes and approximate the distribution of the distance d_k from a randomly chosen sample to it's k_{th} nearest neighbours by a mixture of two Gaussian defined by model parameters w and $\Theta_d = [\lambda_1, \lambda_2, \sigma_1, \sigma_2]$:

$$f(d_k|w, \Theta_d) = wN(d_k, \lambda_1, \sigma_1) + (1 - w)N(d_k, \lambda_2, \sigma_2) \tag{2}$$

where λ_1 and λ_2 are the average distances of a component to its k_{th} nearest neighbours, σ_1 and σ_2 the corresponding variances and w and $(1-w)$ the mixture probabilities for feature and noise respectively. The mixture parameters are estimated using an EM algorithm and the components (see Figure 5) are filtered by applying the following procedure:

1. For K randomly chosen components

   (a) Calculate the average Euclidean distance to its k_{th} nearest neighbours

   (b) Estimate the model parameters $[w, \lambda_1, \lambda_2, \sigma_1, \sigma_2]$ using EM [5]

2. Classify each component according to whether the average distance to it's k_{th} nearest neighbours has a higher probability under the feature or clutter.

An example for this procedure is shown in Figure 5 . Using the components extracted from a sequence, as illustrated in Figure 4, we calculate the average Euclidean distance for all components to the nearest 15 neighbours and approximate the resulting distribution by a mixture of 2 Gaussian as seen in Figure 5 (left). Clutter is represented by the distribution with the largest variance and all atomic components whose average distance to it's k_{th} nearest neighbours falls within this distribution is classified as noise.

4 Automatic Model Order Selection

The density distribution in component space can be described by a mixture of Gaussian, where each mixture component, k, models a different atomic component. Consequently, the determination of atomic components requires the determination of an optimum number of unknown clusters K, known as the problem of model order selection, and the estimation of the model parameters W_K and Θ_K.

The problem of model order selection has been widely studied in the literature (see [4] for a review). Heuristic methods have been proposed by Akaike [1], Schwarz [14] and Rissanen [11], who respectively proposed (AIC) A Information Criterion, (BIC) Bayesian

Information Criterion and (MDL) Minimum Description Length. These methods are heuristic in the sense that they do not minimise an error function between the estimated and the true model order. Instead these methods define various information criteria that only depend on the unknown model order K, which is defined as minimising value for the respective criterion. One of the most popular criteria, the information criterion of Rissanen, MDL, is defined as

$$MDL(K) = -ln[L(Z|W_K, \Theta_K)] + \frac{1}{2}Mln(N) \tag{3}$$

MDL is obtained from information-theoretic considerations, and the model order is defined as the model that minimises the description length, i.e. the model that encodes the vector of observations in the most efficient way [7]. The first term $-ln[L(Z|W_K, \Theta_K)]$, the maximised mixture likelihood of $P(Z|W_K, \Theta_K)$, measures the systems entropy and can be seen according to Shannon's Information Theorem as a measure for the number of bits needed to encode the observations $Z = [z_1, z_2, ...z_N]$, with respect to the model parameters W_K and Θ_K

$$P(Z|W_K, \Theta_K) = \prod_{i=1}^{N} f(z_i|W_K, \Theta_K) \tag{4}$$

The second term, $\frac{1}{2}Mln(N)$ measures the additional number of bits needed to encode the model parameters and serves as penalty for models that are too complex. M describes the number of free parameters and is given for a Gaussian mixture by $M = 2dK + (K-1)$ for (K-1) adjustable weights due to the constraint $\sum_k w_k = 1$ and $2d$ parameters for d dimensional means and diagonal covariance matrixes.

The optimal number of clusters and therefore number of atomic components can be determined by applying the following iterative procedure:

1. For all K, $\{K_{min} < K < K_{max}\}$

   (a) Maximise the likelihood $L(Z|W_k, \Theta_k)$ using the k-means [6] or EM [5].

   (b) Calculate the value of MDL(k) according to Eqn. (3) and Eqn. (4)

2. Select the model parameters $\{W_k, \Theta_k\}$ for the minimising value of MDL(k).

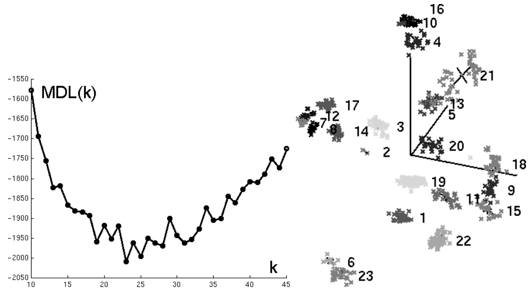

Figure 6: The Minimum Description Length MDL(k) calculated for all cluster configuration $[10 < k < 45]$, with global minimum determined for 23 clusters (left). Component space segmented into 23 clusters: The projection of the 3 largest Principle Components of each atomic component (right).

An example is given in Figure 6. The Minimum Description Length MDL(k) is calculated for all cluster configurations with $[10 < k < 45]$ clusters, extracted from the observation sequence shown in Figure 4 (right). A global minimum and therefore optimal number of clusters can be determined for k=23 clusters, consequently segmenting the projection of the atomic components in component space into 23 clusters as shown in Figure 6 (right). The corresponding trajectory segments are shown in Figure 7, small squares at the end of each trajectory indicate the direction of movement.

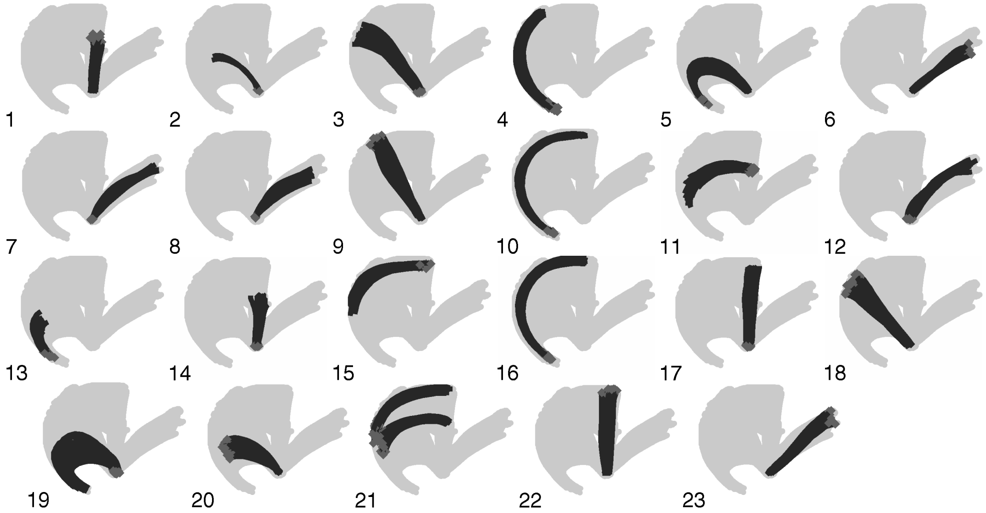

Figure 7: The overlaid trajectory segments corresponding to the 23 extracted clusters.

5 Extracting High Level Knowledge

We assume that gestures are repetitive sequence of atomic components starting and ending in a rest position and are governed by a high level structure controlling their temporal sequence. Consequently the extraction of gestures involves in a first step the determination of rest positions and in a second step the enclosed sequence of atomic components.

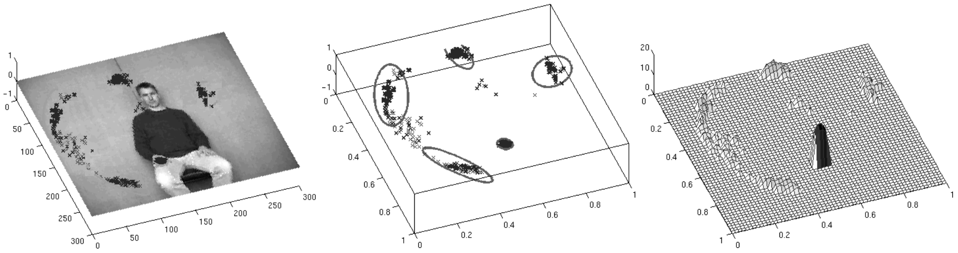

Figure 8: Extracting Rest Positions. Extracted areas with little or no hand motion (left). The distribution approximated by a mixture of Gaussian with 5 mixture components determined using MDL (middle). Probabilities for the rest positions in logarithmic display (right).

Rest positions are defined as areas where the hands undergo little or no motion. We identify rest positions by analysing the complete observation sequence for segments where the velocity drops below a pre-set threshold as shown in Figure 8(left). The corresponding $[x, y]$ co-ordinates are stored and the resulting distribution is approximated by a mixture of Gaussian, as shown in Figure 8(middle). Mixture components are estimated using the EM algorithm and the number of mixtures determined using the MDL framework introduced in the previous section. This is equally true for pause positions, that typically occur between transitions into and out of gesture space, however hands remain longer in rest positions than in a pause position and this results in higher probabilities, as can be seen in Figure 8(right). In real life scenarios we can encounter multiple rest positions and people can change in between them. For simplicity we assume there is only one rest position and consequently select the cluster with the highest probability.

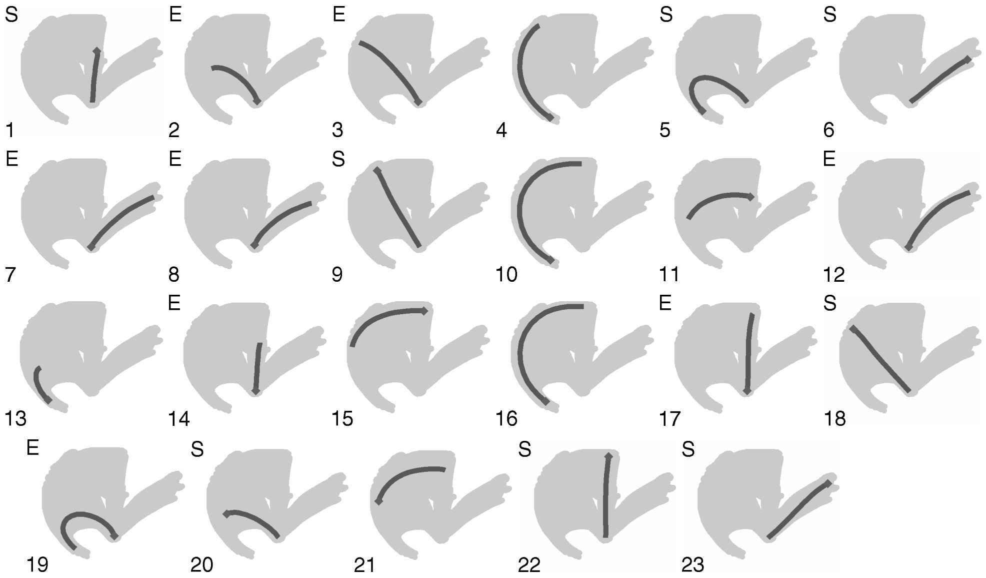

Figure 9: Atomic Components: The mean trajectories corresponding to the 23 extracted clusters, labelled according to whether they start or end in a rest position.

In a second step atomic components were identified that start or stop in a rest position. The mean trajectory for all components belonging to a particular class is computed and a component defined to start or stop in a rest position in case the first $[x_1, y_1]$ or last $[x_N, y_N]$ co-ordinate tuple is within 3 standard deviations from the rest positions mean position. The mean trajectories are shown in Figure 9. The complete observation sequence is then analysed for components that start and stop in a rest position and the corresponding component identifier of the enclosed sequence of atomic components stored. Sequences containing components that were previously classified as clutter are discarded. The derived gesture models and the atomic components they consist of are shown in Figure 10. The labels next to each atomic component indicate the corresponding component identifier, labels at the bottom left indicate the gesture model identifier whereas labels at the top left indicate the number of extracted sequences corresponding to each gesture model.

5.1 Determining the Number of Gesture Models

Looking at components [6,23],[7,8,12],[9,18],[10,16] in Figure 9 we can see a general tendency of MDL to overestimate the total number of atomic components. This results in identical gestures being described by multiple models containing different, however similar atomic components as can be seen in Figure 10 for models [2,7,11,12,13], [3,9] ,[4,5]. To reduce the number of models, we take the complete observation sequence, concatenate consecutive trajectory segments corresponding to the extracted gesture models, approxi-

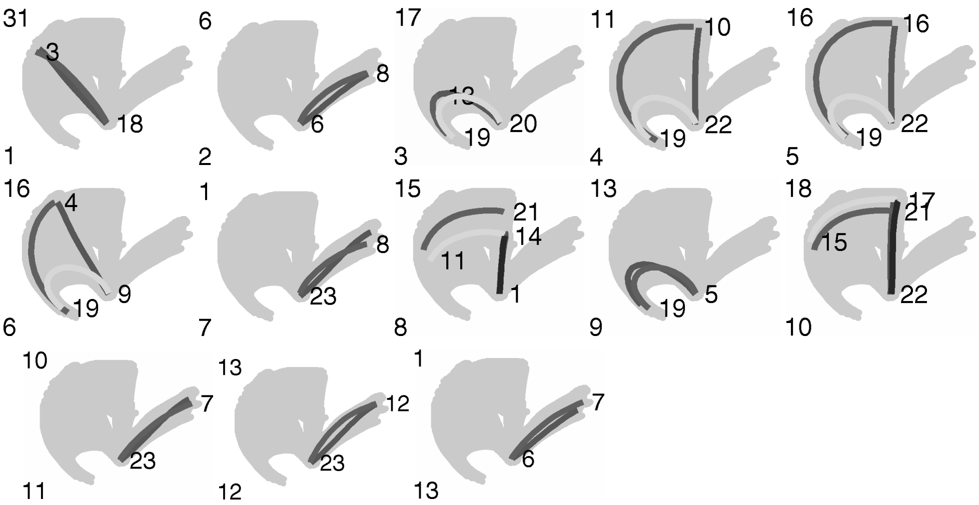

Figure 10: The 13 extracted gesture models depicted by their atomic components. The small numbers in the lower left indicate the model identifier and the numbers in the upper left indicate the number of times each particular sequence has been seen.

mate the concatenated components with splines and interpolate them into $2d$-dimensional vectors $z_i = [x_1, y_1, x_2, y_2, ..., x_d, y_d]$. The corresponding gesture model identifiers are stored and the resulting distribution is approximated by a mixture of Gaussian. Clusters are extracted using k-means and their number determined using the already introduced MDL. Models corresponding to the model identifier in each of the extracted cluster are grouped together to one model, thus reducing the total number of models, as seen in Figure 11.

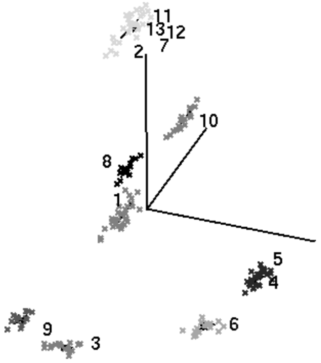

Figure 11: Reducing the number of gestures. All trajectory segments corresponding to the extracted gesture models projected into gesture space and re-clustered using k-means, grouping models [4,5] and [2,7,11,12,13] together.

6 Experiments

To evaluate our approach, we recorded a participant performing 7 gestures. The recording included deictic gestures such as *”pointing left”* and *”pointing right”* , metaphoric gestures such as *”he bent a tree”* and *”there was a big explosion”* and communicative gestures such *”waving high”* , *”waving low”* and *”please sit down”* . Examples are shown in Figure 1, Figure 2 and Figure 3.

A time limit of 15 minutes was set and each gesture was approximately performed 30 times in arbitrary order. Gestures were recorded with 15 frames per second and the $2D$ positions of the participants hands and the head extracted using a tracker developed by Sherrah and Gong [15]. The 2-dimensional observation trajectory containing the $[x,y]$ position of the participant's right hand was approximated by a spline and automatically partitioned into 701 atomic components as shown in Figure 4. All atomic components were transformed into a component space. Each component was interpolated into 20 $2D$ vertices, stored as 41-dimensional vector (20 $2D$ vertices plus scale factor) and reduced to 5 dimensions using PCA, still containing 95% of the original information. The average Euclidean distance for each component to its nearest 15 components was calculated. The resulting distribution was approximated by a mixture of two Gaussian as seen in Figure 5 and all components with higher probability under clutter than feature removed, thus reducing the number of components by 118 to 583. The remaining components were approximated by a mixture of Gaussian with $[10 < k < 45]$ mixture components, determined using a k-means clustering algorithm. The Minimum Description Length MDL(k) was calculated for each configuration and a global minimum was determined for $k = 23$ mixtures as shown in Figure 6 (left). Figure 6 (right) shows the projection of the largest 3 principal components of each atomic component segmented into 23 clusters and Figure 7 shows the corresponding trajectory segments. The small numbers next to each diagram indicate the mixture identifier with the small squares at the end of each trajectory, indicating the direction of movement. Looking at components [6,23], [7,8,12], [9,18], and [10,16] we can see a general tendency of MDL to overestimate the number of atomic components. They are similar however were split into different components.

Figure 8 shows the extracted rest and pause positions. The complete observation sequence was analysed for sequences where the velocity dropped below 1/20th of the maximal velocity. The corresponding $[x, y]$ co-ordinates were extracted and the resulting distribution approximated by a mixture of Gaussian. Five mixture components were determined using MDL and the mixture with the highest probability chosen as rest position. Figure 9 shows the mean trajectories corresponding to each extracted atomic component. The small numbers in the lower left indicate the component identifier and the letter 'S' or 'E' indicate whether a component starts or ends in a rest position. The extracted gesture models can be seen in Figure 10. The small numbers in the lower left indicate the model identifier, the numbers in the upper left indicate the number of times each particular sequence has been seen and the number next to each of the trajectory segments the atomic component identifiers. A total of 13 gesture models were extracted corresponding to the 7 gestures due to a tendency of MDL to overestimate the number of atomic components which can be seen in models [2,7,11,12,13], [3,9] and [4,5]. They are identical however consist of different atomic components. For example, gesture [2] consists of components [6] and [8], while gesture [11] consists of components [23] and [7] similarly gesture [3] and [9] consist of components [20,13,19] and components [5,19] respectively. Some of the components are shared by two or more gestures. The atomic component [22] is shared by the metaphoric gestures [4,5] "*he bent a tree*" and the communicative gesture [10] "*waving high*" or component [19], is shared between the communicative gestures [3,9] "*please sit down*" and the metaphoric gestures [4,5] "*he bent a tree*" and [6] "*there was a big explosion*", thus demonstrating the functionality of atomic components as building blocks of gestures.

To reduce the final number of gestures all trajectory segments belonging to each of the 13 extracted gesture models were concatenated and normalised into a 40 dimensional vector. The corresponding model identifiers were stored and the resulting distribution was approximated by a mixture of Gaussian. Eight mixture components were determined using MDL, grouping gesture models [2,7,11,12,13] and [4,5] together as can be seen in Figure 11. Thus reducing the total amount of gestures to 8, containing the models [1], [2,7,11,12,13], [3], [4,5], [6], [8], [9] and [10].

7 Summary and Conclusions

We presented a systematic approach to automatically segment and label a continuous observation sequence of hand gestures for a complete unsupervised model acquisition. We assumed that gestures can be viewed as a repetitive sequence of atomic components that can be modelled by a mixture of Gaussian in a component space. Mixture components were determined using k-means clustering and the number of components was automatically determined using the MDL criterion. Experiments were performed on a training sequence containing 7 different gestures. A first low level analysis slightly overestimated the number of models and extracted 13 partly identical gestures due to a tendency of MDL to overestimate the number of atomic components. In a second step, we analysed the gestures on a high level and were able to reduce the number of models to 8. Visual inspection shows that the extracted gestures model all 7 original gestures.

References

[1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[2] H. Asada and M. Brady. The curvature primal sketch. *Technical Report. MIT AI memo 758*, 1984.

[3] A. Bobick and A. Wilson. A state-based technique for the summarisation of recognition of gesture. *ICCV*, pages 382–388, 1995.

[4] H. Bock. Probability models and hypothesis testing in partitioning cluster analysis. *Clustering and Classification*, pages 377–453, 1996.

[5] N. Dempster, A.P.and Laird and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:1–38, 1977.

[6] J. Hartigan and M. Wong. A k-means clustering algorithm. *Appl. Statist.*, 28:100–108, 1979.

[7] D. Hirshberg and N. Merhav. Robust methods for model order estimation. *IEEE Transactions on Signal Processing*, 44:620–628, 1996.

[8] S. McKenna and S. Gong. Gesture recognition for visually mediated interaction using probabilistic event trajectories. *BMVC*, pages 498–507, Southampton, England, September 1998.

[9] D. McNeill. Hand and mind: What gestures reveal about thought. *University of Chicago Press*, 1992.

[10] Y. Raja, S. McKenna, and S. Gong. Colour model selection and adaptation in dynamic scenes. *ECCV*, Freiburg, Germany, 1998.

[11] J. Rissanen. Modelling by shortest data description. *Automatica*, 14:465–471, 1978.

[12] S.Byers and A. Raftery. Nearest neighbour removal for estimating features in spatial point processes. *Technical Report no 305*, Department of Statistics, University of Washington, April 27, 1996.

[13] J. Schlenzig, E. Hunter, and R. Jain. Recursive identification of gesture inputs using hidden markov model. *Workshop on Applications of Computer Vision*, pages 187–194, 1994.

[14] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[15] J. Sherrah and S. Gong. Resolving visual uncertainty and occlusion through probabilistic reasoning. *BMVC*, 1:252–261, September 2000.

[16] T. Starner and A. Pentland. Visual recognition of american sign language using hidden markov models. *International Workshop on Automatic Face and Gesture Recognition*, pages 189–194, Zurich, 1995.

[17] M. Walter, S. Gong, and A. Psarrou. Stochastic temporal models of human activities. *International Workshop on Modeling People*, pages 87–94, Corfu, Greece, 1999.