Segmentation of Hand Gestures using Motion Capture Data

(Extended Abstract)

Ajay Sundar Ramakrishnan
University of California Davis
One Shields Avenue
Davis, CA, USA
ajay.sundar.r@gmail.com

Michael Neff
University of California Davis
One Shields Avenue
Davis, CA, USA
neff@cs.ucdavis.edu

ABSTRACT

Virtual agent research on gesture is increasingly relying on data-driven algorithms, which require large corpora to be effectively trained. This work presents a method for automatically segmenting human motion into gesture phases based on input motion capture data. By reducing the need for manual annotation, the method allows gesture researchers to more easily build large corpora for gesture analysis and animation modeling. An effective rule set has been developed for identifying gesture phase boundaries using both joint angle and positional data of the fingers and hands. A set of Support Vector Machines trained from a database of annotated clips, is used to classify the type of each detected phase boundary into stroke, preparation or retraction. The approach has been tested on motion capture data obtained from different people with varied gesturing styles and in different moods and the results give us an indication of the extent to which variation in gesturing style affects the accuracy of segmentation.

Categories and Subject Descriptors

I.3.7 [**Computing Methodologies**]: Computer Graphics—*Three dimensional graphics and realism*[Animation]

General Terms

Human Factors

Keywords

motion capture; gesture segmentation; machine learning; support vector machines; principal component analysis; character animation

1. INTRODUCTION AND RELATED WORK

Large corpora are increasingly being used to build more natural and realistic models of gesture. To be useful, these corpora need to be segmented into the phases: preparation, stroke, hold and retract (PSHR model)[7]. This paper looks at the task of automating this segmentation process (a time consuming process, when manually done, reliant on heuristics and subjective judgement), which is a challenging task because gesturing style varies greatly across people and with changes in mood, including variation in frequency, amount of movement and gesture position.

According to [3], the stroke is the central part of the gesture and carries its meaning; the preparation phase is responsible for the hand being raised to the starting point of the stroke and the retraction phase is responsible for returning the hand to a resting position. [7] noted that a gesture could have a hold phase, which is performed while waiting for coherence of speech with stroke or while extending a single movement stroke or when a stroke is completed early.

[5] performs a segmentation of gesture phases, using velocity-acceleration profiles, but on implementing the postulates of this approach, we obtain a prohibitively large number of false positive segment boundaries. [6] developed FORM, a fine-grained annotation scheme and represented gesture data based on this scheme, but a major problem in this approach is that the work needed to FORM-annotate the video segment is prohibitive for large datasets.

An approach similar to [1] is used in our approach for quantifying and segmenting motion of the human hand. A Support Vector Machine [2] which can learn underlying patterns, as used by [4], is used in our work for segmentation.

Fig 1 gives a brief overview of our approach. These steps will be described in the remainder of the paper.

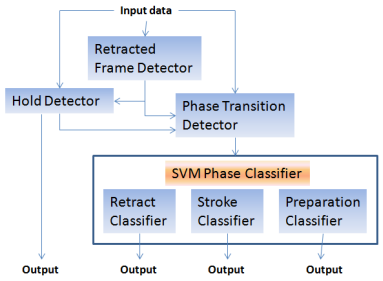

Figure 1: Flowchart of our segmentation approach

2. OUR APPROACH

The main steps involved in our approach are as detailed below:

Appears in: *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013)*, Ito, Jonker, Gini, and Shehory (eds.), May, 6–10, 2013, Saint Paul, Minnesota, USA.

Table 1: Features chosen for segmentation using SVM for each phase

Phase type	Features chosen
Retract	Position at next inflexion frame, Avg. vertical velocity during the phase, Avg. rate of proj. error change during the phase
Stroke	Avg. acceleration during the phase, Avg. rate of proj. error change during the phase
Preparation	Position of hand at present inflexion frame, Avg. vertical velocity during the phase, Avg. rate of proj. error change during the phase

2.1 Preprocessing Segmentation Data

The positions, velocities and accelerations of the wrist joints are first calculated. Then, complex hand data obtained from cybergloves is processed in a novel approach to be able to provide a measure of the change in handshape, which aligns well with phase transitions. Our approach for this is based on the fact that simple motions have lower inherent dimensionality than more complex ones and thus a change in inherent dimensionality at a particular frame would represent a transition at that frame. Each segment in our case would correspond to a different hand-shape. We use Principal Component Analysis to reduce the 21 dimensional (there are 21 parameters obtained at each frame of capture from the cybergloves) data of hand motion to 5 dimensions (calculated using our training data by ensuring that 90 % of the information is retained in this lower dimensional space). Then, projection errors of the data for this dimensionality reduction is calculated for each frame. A large change in projection error at a frame indicates a possible segmentation boundary at that frame.

Retracted frames are the frames a gesturer brings his hands to, while not performing a gesture. Identifying these retracted positions help in reducing the frames we have to consider for segmentation, apart from helping in identifying other phases like retraction. Hold phases are then found by looking for phases where the hand is stable and not in a retracted position.

All potential inflexion frames are found by looking for abrupt change in the acceleration of the hand and handshape change, in the frames that have not been classified as retracted frames or holds. Frames too close to each other due to a change in both these parameters are then merged.

2.2 The Segmentation Process

The inflexion frames calculated above need to be classified as either false positives or labeled as a particular type of phase change. A Support Vector Machine is used for this classification. The features chosen for the segmentation are listed in Table 1.

The list of feature vectors and their labels at the inflexion points calculated above, are fed as training data to the SVM algorithm. The data first has to be scaled to the [-1,1] interval, then the RBF kernel is used for the SVM process. The values of the parameters of the RBF function are found by n-fold cross validation. The SVM is then trained and a model summarizing the data is generated, with which the hyperplane to separate the data is calculated. Three such SVMs are trained- one each for the preparation, stroke and retract phases. All potential inflexion points are tested with each SVM and it produces a "yes"/"no" label as to whether the inflexion frame is the start of a particular type of phase.

3. RESULTS AND CONCLUSION

The algorithm was tested on about 15 minutes of hand annotated motion capture data of 3 people with varying gesturing styles, performing gestures in different moods. Training and testing on singleton gesture data of the same person provided a high (80 %) total accuracy percentage. Training on similar real gesturing data of the same person also gave encouraging results with a 75 % total accuracy of prediction. Training and testing on similar gesturing data of different persons gave a 75% for all phases other than the preparation phases which had a poor prediction accuracy of 47 % due to many inflexion frames not being detected. Training and testing on stylistically different gesturing data of the same person gave very poor accuracy (20%), again due to many inflexion frames being missed due to the difference in fundamental characteristics of training and test data.

To summarize, this paper presented a method of segmenting gesture sequences automatically, using motion capture data. It provided a method of quantitatively representing handshape change using PCA. It introduces the concept of retracted positions which provide reference points to find information about other phases. The paper showed that a majority of the phase change frames could be detected using acceleration and hand-shape information for singleton gestures and stylistically similar gesturing data.

4. REFERENCES

[1] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard. Segmenting motion capture data into distinct behaviors. In *In Graphics Interface*, pages 185–194, 2004.

[2] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, 1992.

[3] D. Efron. *Gesture and environment*. King's Crown Press, Morningside Heights, NY, 1941.

[4] C. Li, P. R. Kulkarni, and B. Prabhakaran. Segmentation and recognition of motion capture data stream by classification.

[5] A. Majkowska, V. B. Zordan, and P. Faloutsos. Automatic splicing for hand and body animations. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 309–316, 2006.

[6] C. Martell. *FORM: An experiment in the annotation of the kinematics of gesture*. PhD thesis, University of Pennsylvania, 2005.

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