MOTION STREAM SEGMENTATION AND RECOGNITION BY CLASSIFICATION

Chuanjun Li   *Punit R. Kulkarni*   *B. Prabhakaran*

Department of Computer Science
University of Texas at Dallas, Richardson, TX 75083
{chuanjun, prk032000, praba}@utdallas.edu

ABSTRACT

This paper proposes a classification-based approach to segmenting and recognizing patterns in motion signals. Feature vectors are extracted based on singular value decomposition (SVD) for classification. Multi-class support vector machine (SVM) classifiers with class probability estimates are explored for segmenting and recognizing motion streams. Experiments show that the proposed approach can find patterns in the multi-attribute motion streams with high accuracy.

1. INTRODUCTION

Recognition of motions signals or streams from 3D motion capture systems or data gloves can find wide uses in many applications, such as surveillance video systems, 3D animation and simulation-based training, gait analysis and rehabilitation and gesture recognition. Distance measures can be defined as in [5] to capture motion similarities in streams, or machine learning techniques can be employed to recognize motions as in [7].

SVM classifiers with decision values, rather than class probability estimates, have been successfully applied to classify complete patterns/isolated hand gesture motions generated by using a data glove in [4]. As shown in [8] and [4], when only isolated motions or patterns are considered and multiple examples for each pattern are available, classification can have higher recognition rate than template matching with certain similarity or distance measure.

We observe two intuitive yet important facts:

- If a pattern belongs to some class, the probability of any of its sub-patterns being in the same class is less than the probability of the complete pattern being in the class.
- If all classes contain complete patterns only, the probability of a pattern being in its own class is higher than the probability of any of its sub-patterns being in any class.

Proposed Approach: Based on the observations, we propose to apply classification to stream segmentation and pattern recognition in order to take advantage of the high accuracy of classification. In order for classification to be applicable to multi-attribute matrices, feature vectors capturing the major geometric structures of motion matrices are extracted by using singular value decomposition for motion matrix classification. Multi-class SVM classifiers with class probability estimates are explored for recognizing patterns in streams. Class probability estimates are proposed not only for recognizing class labels of complete patterns, but also for rejecting incomplete patterns or sub-patterns. We propose to use complete pattern examples only, and no non-pattern/incomplete examples at all for training. For the sequentially segmented motion candidates, the best class is chosen from the two classes which give the two highest probabilities. The class to which a larger number of motion candidates belong is determined to be the best class and the motion candidate which has the highest class probability is the best motion segment.

2. FEATURE SELECTION

Multi-attribute motions can be represented by matrices in which each column represents one attribute, and each row is for one recording of the motion signal. This section describes how to extract features from motion matrices for stream segmentation.

The geometric structure of a matrix can be revealed by the SVD of the matrix. As shown in [2], any real $m \times n$ matrix A can be decomposed into $A = W\Sigma Z^T$, where $W = [w_1, w_2, \ldots, w_m] \in R^{m\times m}$ and $Z = [z_1, z_2, \ldots, z_n] \in R^{n\times n}$ are two orthogonal matrices, and Σ is a diagonal matrix with diagonal entries being the singular values of A: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$. Column vectors w_i and z_i are the i^{th} left and right singular vectors of A, respectively, and all the right singular vectors have the same dimension n, which is independent of the number of matrix rows m.

The ith largest singular value σ_i of A is geometrically the 2-norm or Euclidean length of the ith largest projected vector Ax which is orthogonal to all the $i-1$ larger orthog-

Work supported in part by the National Science Foundation under Grant No. 0237954 for the project CAREER: Animation Databases.

onal vectors as shown by

$$\sigma_i = \max_{U} \min_{x \in U, \|x\|_2 = 1} \| Ax \|_2$$

where the maximum is taken over all i-dimensional subspaces $U \subseteq \Re^n$ [2]. Hence, the right singular vectors are the corresponding projection directions of the associated singular values, and the singular values account for the Euclidean lengths of different vectors projected by the row vectors in A onto different right singular vectors.

When two motions are similar, the row vectors in the motion matrices should cover similar trajectories in the n-dimensional space, hence the geometric structures of the motion data matrices are similar. For realistic motions with variations, singular vectors associated with different singular values have different sensitivities to motion variations. If a singular value is large and well separated from its neighbors, the associated singular vector would be relatively insensitive to small motion variations. On the other hand, if a singular value is among a poorly separated cluster, its associated singular vector would be highly sensitive to motion variations.

Figure 1 shows the accumulative singular values for hand gestures and captured human subject motions. It shows that the first two singular values account for more than 95% of the sum of singular values, while the others might be very small. Accordingly, the corresponding first two singular vectors of similar motions, especially the first singular vectors would be close or parallel to each other as shown in Figure 2.

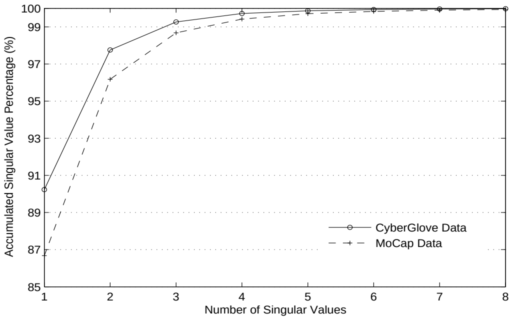

Fig. 1. Accumulated singular value percentages in the sums of singular values for two data sources: CyberGlove data and captured human subject motion data. There are 22 singular values for one CyberGlove motion, and 54 singular values for one motion capture motion.

Since the right singular vector u_1 can have opposite signs, the following steps can be taken to obtain consistent signs for u_1 of similar patterns.

1. Generate a matrix S with rows being the first right singular vectors u_1 of all known patterns.

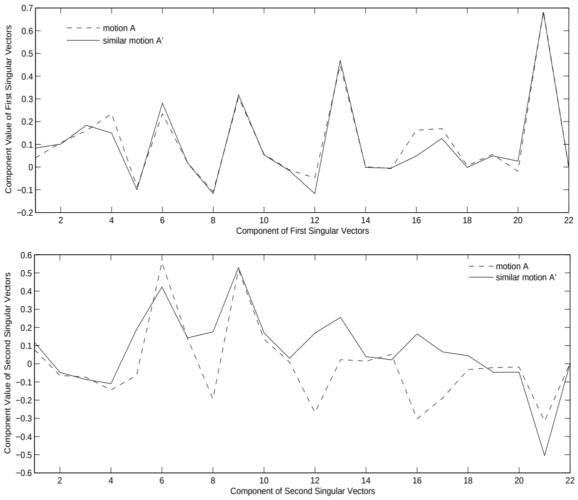

Fig. 2. Singular vectors of similar patterns. The first singular vectors are similar to each other, while other singular vectors, even the second vectors as shown in the bottom, can be quite different.

2. Subtract the elements of S by their corresponding column means, and update S to be the resulting matrix with zero column means.
3. Compute the SVD of S and let its first right singular vector be s_1.
4. Project the first right singular vector u_1 of patterns (or pattern candidates) onto s_1 by computing $u_1 \cdot s_1$.
5. Negate all components of any u_1 if the corresponding the inner product $u_1 \cdot s_1 < 0$, and let u_1 be the negated vector.

Since the projections of u_1 onto s_1 have the largest variances among projections on any unit vectors, we can expect that $u_1 \cdot s_1$ will not cluster around zero. Our experiments with hundreds of patterns of different sources show that no pattern has $|u_1 \cdot s_1| < 0.3$. Because similar patterns should have close projections $|u_1 \cdot s_1|$, reasonable variations in similar patterns would not result in $u_1 \cdot s_1$ projections of opposite signs if their u_1 signs are the same. That is, only if the u_1 signs of similar motions are opposite can their $u_1 \cdot s_1$ projections have different signs. Hence, u_1 of similar motions would have the same sign by requesting $u_1 \cdot s_1 > 0$.

Similarly, the above steps can be repeated for u_2 with all u_1 replaced by u_2, resulting in consistent signs for the second singular vectors of similar motions.

We can extract the feature vectors from the singular vectors and singular values. The first two singular vectors are the most dominating factors contributing to the similarity of two motions due to their associated large singular values. Other singular vectors are less reliable in capturing the similarities due to their associated singular values which might be small and approach zero. Hence we can use singular values as weights to reflect the reliability of the associated singular vectors. Feature vectors are thus constructed by *concatenating the weighted first singular vectors* $w_1 u_1$ *with the weighted second singular vectors* $w_2 u_2$, where $w_i = \sigma_i / \sum_{k=1}^{n} \sigma_k$. These feature vectors are extracted by using only the prominent information from the right singular vectors and singular values, and have the same dimension n irrespective of the variable number of rows of different motion matrices.

3. STREAM SEGMENTATION BY CLASSIFICATION

This section discusses how to recognize patterns in multi-attribute streams by classifying the feature vectors extracted as above using SVM classifiers.

3.1. Classifier Selection

SVM classifiers have found widespread successful uses in many pattern recognition problems [1]. The good classification performances are due to the optimal hyperplane which maximizes the margin, or the distance between separating hyperplane and the training examples nearest to the hyperplane. Efforts of mapping standard SVM outputs to posterior probabilities have been made in [6, 8]. For multi-class classification, class probabilities can be estimated from binary class probabilities by pairwise coupling. Wu et al. [9] propose a multi-class probability approach which is more stable than other popular existing methods and the class with the largest posterior is chosen to be the winning class for a test vector: $\arg\max_i [p_i]$.

It has been shown by experiments on large-scale problems in [3] that in general the accuracy rate of one-versus-one multi-class SVM is higher than that of one-versus-rest, and the training time of one-versus-one multi-class SVM is less than that needed for one-versus-rest classifiers. Due to these reasons, we propose to use the one-versus-one multi-class SVM classifiers with estimated class probabilities for classification, and the Gaussian radial basis function (RBF) is used as the kernel function for classifier training. We use multi-class SVM classifiers for distinguishing patterns from incomplete patterns or sub-patterns segmented from a stream, and for classifying patterns. No sub-patterns or non-patterns are needed as negative examples, and all classes include only complete patterns.

3.2. Stream Segmentation

We assume that a pattern in a stream has a minimum length l and a maximum length $\mathcal{L}$. A stream is segmented into multiple segments ending between l and $\mathcal{L}$ with segment length difference $\Delta > 0$. The feature vector for each motion segment is constructed as described in Section 2.

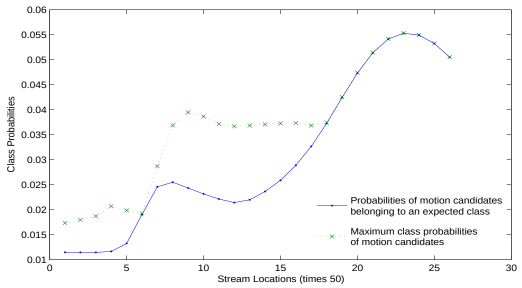

Fig. 3. Changes in estimated probabilities. The probability estimate reaches maximum as a pattern is completed, and decreases as the stream continues.

The completely segmented pattern in a stream will have the highest probability of being recognized as its corresponding correct class in the training set. When the optimal hyperplane divides the space between the correct and incorrect classes, this completely segmented pattern will lie on the side of the correct class. Similarly as the lengths of the motion candidates increase from l and the motion candidates approach a motion pattern, their feature vectors would become closer to the optimal hyperplane if they are on the side of any other classes, and move away from the optimal hyperplane if they are on its correct class side. The corresponding probability of the incomplete motion candidates being classified to the correct pattern class will eventually increase to the maximum as shown in Figure 3. Hence, the point with the highest probability should be the segmentation point for a complete pattern, and the class with the highest probability should be the right class for the segmented pattern.

Extracting a feature vector from a motion matrix unavoidably loses some minor information of the motion, and this information loss can be reflected in the class probabilities of motion candidates when stream segmentation is considered. It might be possible that the probability of the feature vector of certain motion candidate belonging to some class is higher than the probability of the feature vector of the best motion candidate belonging to the correct class. To address this issue, we obtain two classes with the highest probabilities instead of only one class with the highest probability. To choose the correct class, we consider the number of motion segment candidates belonging to each of the two classes. As Figure 3 indicates, the candidates which are close to the best candidate also have the highest probabilities of belonging to the expected class. This might not be true for the feature vector of some other motion candidate, which happens to have higher probability of belonging to some class than that of the best candidate belonging to the expected class. Hence we choose the class *to which more* *motion candidates belong* to be the best one, and choose the segment, which has the highest probability among all the candidates belonging to the same class, to be the best motion candidate.

The next pattern recognition starts from the end of the last recognized pattern in the stream, and the same process repeats until the remaining stream has length less than the minimum stream length l.

4. EXPERIMENT EVALUATION

4.1. Data Generation

Hand gesture streams and human subject motion streams are used for performance evaluation. Hand gestures were generated by suing a data glove called CyberGlove, and the human subject motions were captured by using 16 Vicon cameras and the Vicon iQ Workstation software.

CyberGlove Data: The data for a hand gesture contain 22 angular values for each time instant/frame, one value for a joint of one DOF. The motion data are extracted at 120 frames per second. One hundred and ten different isolated motions were generated as motion patterns, and each motion was repeated for 3 times. That is, each of the 110 classes has 3 examples. Twelve different motion streams were generated for segmentation and recognition purpose. A gesture stream contains 5 to 10 gestures.

Motion Capture Data: The global 3D joint coordinates have been transformed by translations and rotations to be positions of different joints relative to a moving coordinate system with the origin at some fixed point of the subject, for example the pelvis, and the transformed data would be translation and rotation invariant. Motion matrices have 54 columns for coordinates of 18 joints.

One hundred isolated motions including Taiqi, Indian dances, and western dances were performed for generating captured motions, and each motion was repeated 5 times. Every motion repetition has a different location and can face different orientations. Hence we have 100 classes of motion patterns, and each of the classes has 5 examples for SVM training. Twelve motion streams were also generated for stream segmentation. The motion streams include 3 to 5 different length motion patterns each and the patterns in the motion streams have various-length transitions.

4.2. Performance of Classification

k-fold cross validations are used for training of SVMs, where k is 3 for the hand gestures and 5 for the 3D captured motions. The average cross validation accuracy is 96.7% for isolated hand gestures, and is 97.7% for the isolated captured motions.

We use the recognition accuracy as defined in [7]. For the CyberGlove data streams, there are 74 patterns in the 12 streams. Out of the 74 patterns, there are 11 insertions (I), 1 deletion (D) and 1 substitution (S). The accuracy is 82.43%. For the motion capture data streams, there are 45 patterns in the 12 streams. There are 5 deletions (D). The accuracy is 88.9% as shown in Table 1.

Table 1. Pattern Recognition Accuracy (%)

Data Source	Isolated Motions	Motion Streams
CyberGlove	96.7	82.43 (N=74, I=11,D=1,S=1)
Motion Capture	97.7	88.9 (N=45, D=5)

5. CONCLUSIONS

In this paper, multi-class SVMs with probability estimates are proposed for segmenting streams and recognizing the patterns in streams. SVD is applied to extract feature vectors for multi-attribute motion patterns. Two classes with the highest probabilities are chosen to be the best class candidates for the motion candidates of one motion pattern, and the class to which more motion candidates belong is the winning class. The candidate which has the highest probability of belonging to the best class is the best segmented motion candidate. Experiments with CyberGlove data and 3D motion capture data show that SVMs combined with SVD can segment and recognize motion patterns in multi-attribute motion streams with high accuracy.

6. REFERENCES

[1] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, (2):121–167, 1998.

[2] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore,Maryland, 1996.

[3] C.-W. Hsu and C.-J. Lin. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, (13):415–425, 2002.

[4] C. Li, L. Khan, and B. Prabhakaran. Real-time classification of variable length multi-attribute motion data. *International Journal of Knowledge and Information Systems (KAIS)*, Accepted.

[5] C. Li, P. Zhai, S.-Q. Zheng, and B. Prabhakaran. Segmentation and recognition of multi-attribute motion sequences. In *Proceedings of the ACM Multimedia Conference 2004*, pages 836–843, Oct. 2004.

[6] J. C. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A. J. Smola, P. L. Bartlett, B. Scholkopf, and D. Schuurmans, editors, *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA, 2000. URL citeseer.nj.nec.com/platt99probabilistic.html.

[7] T. Starner, J. Weaver, and A. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, 1998.

[8] V. N. Vapnik. *Statistical Learning theory*. Wiley, New York, 1998.

[9] T.-F. Wu, C.-J. Liu, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, (5):975–1005, 2004.