SIMILARITY MEASURE FOR MULTI-ATTRIBUTE DATA

Chuanjun Li   *B. Prabhakaran*   *Si-Qing Zheng*

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

ABSTRACT

Efficient recognition of haptic data such as 3D motion capture data and sign language sensory data can have wide applications in the interactive computer animation and sign language automatic translation areas. For this purpose, we propose a similarity measure for multi-attribute haptic data, a new form of multimedia signal. The proposed similarity measure, based on singular value decomposition, captures the most important features of the signal data, allows for different signal generating rates and reasonable variations in similar signals. Experiments with real life and synthetic data demonstrate that the proposed similarity measure can capture the similarities of motions with different speeds and different lengths and can have up to 100% recognition rates.

1. INTRODUCTION

Haptic data taken from 3D motion capture and sign language sensory devices such as CyberGlove are new forms of multimedia signals. Efficient recognition of the haptic data is important to motion analysis, human-computer interactions, computer animation and sign language automatic translation, and requires proper processing of the data. The reason is that haptic data pose new challenges to similarity search:

- The data are multi-attribute, and can have dozens or hundreds of values at each time.
- The data are of variable lengths, and different local rates cause corresponding portions to have different lengths.
- Motions can follow similar paths in different directions, and data taken from repetitious motions can hide the repetitions if not properly processed.

Due to these new challenges, Euclidean distance or dynamic time warping (DTW) distance is not suitable for measuring the similarities of multi-attribute data, although DTW distance has been widely used for speech recognition [5]. A well-defined similarity measure should be able to capture the similarities of the haptic data regardless of different data generation rates, large number of values generated each time, or different motion directions or repetitions.

Proposed Approach: We propose to exploit the structural similarity of haptic data for the similarity measure of multi-attribute data. As detailed in [4] and further explored in Section 3.1 and Section 4.2, the geometric structure of multi-attribute haptic data can be revealed by the singular vector decomposition (SVD) of the data. We investigate the main geometric structures of data matrices as exposed by SVD, especially the left singular vectors (See Section 3.1) in this paper. The left singular vectors are interpolated and considered in the similarity measure by having the corresponding DTW distances normalized. Experiments show that 100% recognition rate can be achieved, and issues of different lengths, different motion directions and repetitions can be properly taken into consideration.

2. RELATED WORK

Multi-attribute American Sign Language (ASL) motions are considered in [8], and five-word sentences are segmented at the word level and recognized by using Hidden Markov Models (HMM) with 92-98% word accuracy. The number of words in a sentence is required to be known beforehand, so are the grammar constraints or forms of sentences.

Shahabi et al. [6] applied learning techniques such as Decision Trees, Bayesian classifier and Neural Networks to recognize static signs for a 10-sign vocabulary, and achieved 84.66% accuracy. In [7], a weighted-sum SVD is defined for measuring the similarity of two multi-attribute motion sequences. All right singular vectors are involved in the similarity definition and left singular vectors of the sign data are not used for the weighted-sum SVD measure.

Support Vector Machines (SVMs) are used in [3] to classify multi-attribute motion data and up to 100% accuracy has been achieved to correctly classify and recognize discrete signs.

In [4] a similarity measure is defined as the first step to prune dissimilar sign data followed by directly comparing

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

the normalized left singular vectors to further find the most similar signs. This paper extends the similarity measure as defined in [4] to make it work for motion data with different directions or with repetitions.

3. BACKGROUND INFORMATION

In this section we discuss the SVD and DTW as background information for the proposed similarity measure.

3.1. Singular Value Decomposition

As proved in [2], any real $m \times n$ matrix A has an SVD $A = W\Sigma Z^T$, 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. For similar motions with different lengths, their left singular vectors are of different sizes, but their right singular vectors are of the equal length. The singular values of matrix A are unique, and the singular vectors corresponding to distinct singular values are uniquely determined up to the sign, or a singular vector can have opposite signs.

The geometric structure of a matrix A is revealed by its SVD. If the multi-dimensional row vectors or points in A have different variations along different directions, the SVD of matrix A gives the direction with the largest variation. Along the direction of the first right singular vector, the row vectors in A have the largest variation, and along the second right singular vector direction, the point variation is the second largest, and so on. The singular values reflect the variations along the corresponding right singular vectors. Fig. 1 illustrates the data in an 18×2 matrix with its first and second right singular vectors v_1 and v_2. The 18 points in the matrix have different variations along different directions, hence data have the largest variation along v_1 as shown in Fig. 1.

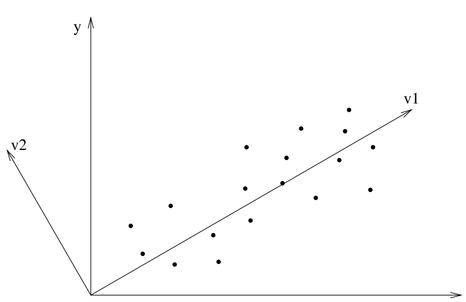

Fig. 1. Geometric Structure of Matrix Exposed by its SVD

3.2. Dynamic Time Warping (DTW)

DTW is used to align two time series of length m and n respectively. Let $d(i,j)$ be the local distance between the two sequences at i and j, and $D(i,j)$ be the global accumulative distance up to the sequence locations at i and j, respectively, then

$$D(i,j) = \min[D(i-1,j-1), D(i-1,j), D(i,j-1)] + d(i,j)$$

subject to $D(1,1) = d(1,1)$. Some other constraints may also be applied to restrict the warping window [1].

4. SIMILARITY MEASURE

In this section, we extend the similarity measure as defined in [4] by further considering the left singular vectors for their DTW distances.

4.1. No Left Singular Vectors Considered

The similarity in [4] is defined as follows.

$$\Psi(Q,P) = |u_1 \cdot v_1| \times (\vec{\sigma} \cdot \vec{\lambda} - \eta)/(1-\eta) \quad (1)$$

where u_1 and v_1 are the first right singular vectors of Q and P, respectively, $\vec{\sigma} = \sigma/|\sigma|$, $\vec{\lambda} = \lambda/|\lambda|$, and σ and λ are the vectors of the singular values of Q^TQ and P^TP, respectively. Weight parameter η is to ensure that $\vec{\sigma}$ and $\vec{\lambda}$ have similar contributions as u_1 and v_1 have to the similarity measure and is determined by experiments. η can be set to 0.9 for the multi-attribute motion data.

This similarity measure captures the most important information revealed by the first right singular vectors and the singular values, and can be applied to prune most of the irrelevant motions for similarity search, and inner products of interpolated first left singular vectors are proposed as our first attempt to consider motions with different directions or with repetitions. Below, we extend the above similarity measure by directly incorporating left singular vector information in a new similarity measure definition.

4.2. Left Singular Vectors Considered

We will show why the above similarity needs to be extended as a standalone measure and show how to extend it by applying dynamic time warping to left singular vectors.

The following theorem implies that motions following the same paths in different directions will have the same similarity as given by (1).

Theorem 1 *Let A be any real $m \times n$ matrix, and $A = U_1\Sigma V^T$. Exchange the order of any two rows of A and let the resulting matrix be B, then $B = U_2\Sigma V^T$.*

Proof. Let $C = A^TA$, then $c(i,j) = \sum_{k=1}^{m} a(k,i)a(k,j)$, and the order of rows in A makes no difference to C. Hence $B^TB = C$. It can be easily shown that $C = V\Sigma^2V^T$, hence Σ and V are independent of the order of rows in A. It follows that $B = U_2\Sigma V^T$, which completes the proof of the theorem.

Since $c(i,j) = \sum_{k=1}^{m} a(k,i)a(k,j)$, repetition of A doubles $c(i,j)$. For any positive p we have $pC = V(\sqrt{p}\Sigma)^2V^T$. Since singular values are normalized in (1), repetition of A will have no effect on the similarity measure as defined above, which will be considered further in this work.

Since $AV = U\Sigma$, hence $\sigma_i u_i = Av_i$, indicating that $\sigma_1 u_1$ are the projections of row vectors in A onto v_1, the direction in which A has the largest variations. Av_1 can also be understood as the row vector component sums weighted by the corresponding vector components of v_1. So $\sigma_1 u_1$ actually contains the motion path information in A, and different directions or repetitions can be reflected in $\sigma_1 u_1$. Similar motions have similar v_1 as shown in [4], and if they have similar directions or repetitions, their projections onto v_1 should be similar, ie, the corresponding $\sigma_1 u_1$ should be close to each other. This makes it possible for us to consider motion direction or motion repetitions by taking into account only $\sigma_1 u_1$, a uni-attribute sequence, rather than the original multi-attribute matrix A.

$\sigma_1 u_1$ can be computed by multiplying A with v_1, which can be computed from SVD of $C = A^TA$. Since different motions can have different durations and their lengths can be hundreds or thousands, direct comparison of $\sigma_1 u_1$ can be computationally costly. We propose to uniformly interpolate the computed $\sigma_1 u_1$, and compare the resulting equal length interpolated $\sigma_1 u_1$.

$AV = U\Sigma$ indicates that as long as v_1 and u_1 have consistent signs, their signs are not unique as illustrated in Fig. 2. Motions (a) and (b) are similar, yet signs of their first left singular vectors can be opposite. Without loss of generosity, we let the mean of v_1 components be positive, and negate $\sigma_1 u_1$ if necessary as shown in Fig. 2. Let $\bar{u}_1 = \sigma_1 u_1$ after negation if necessary. Shifts or dis-alignment of $\bar{u}_1$ for similar motions as shown in Fig. 2 justify the use of DTW for measuring the closeness of $\bar{u}_1$ for different motions.

As Fig. 3 and Fig. 4 show, the DTW distances of $\bar{u}_1$ for similar motions are usually much smaller than those of dissimilar motions. To take the DTW distance into account in the similarity measure, we normalize the DTW distance as follows.

- Find a maximum DTW distance of $\bar{u}_1$ for similar motions, let it be $\bar{D}$.
- Convert DTW distance D into $\mathcal{D} = (\bar{D} - D)/\bar{D}$. Let $\mathcal{D} = 0$ if $\bar{D} < D$.

It can be easily verified that $0 \leq \mathcal{D} \leq 1$. We then define

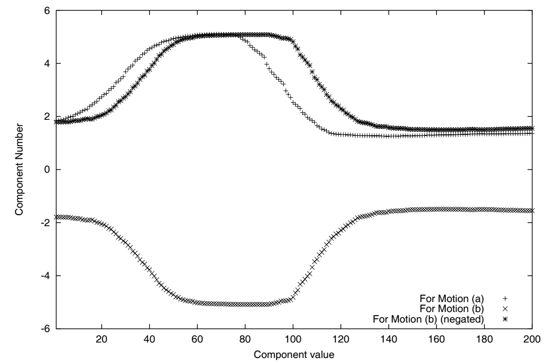

Fig. 2. Interpolated First Left Singular vectors Multiplied by First Singular Values for Similar Motions

our new similarity measure as follows.

$$\Psi(Q,P) = |u_1 \cdot v_1| \times (\vec{\sigma} \cdot \vec{\lambda} - \eta)/(1-\eta) \times \mathcal{D} \qquad (2)$$

where u_1, v_1 are the first right singular vectors of motion Q and P, respectively, $\vec{\sigma}$ and $\vec{\lambda}$ are the corresponding normalized singular value vectors, η is a weight parameter and $\mathcal{D}$ is defined as above.

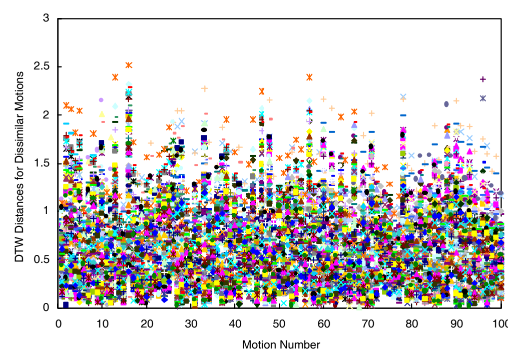

Fig. 3. DTW Distances of $\bar{u}_1$ for Dissimilar Motions

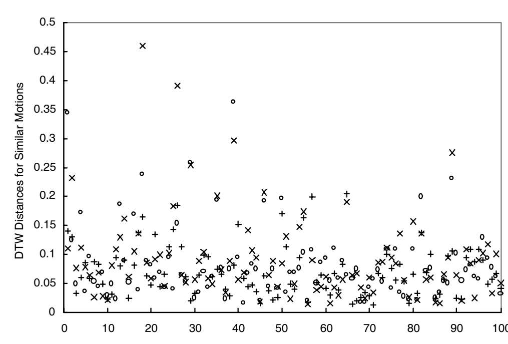

Fig. 4. DTW Distances of $\bar{u}_1$ for Similar Motions

5. PERFORMANCE EVALUATION

In this section we evaluate the performance of similarity (2) on data of one degree of freedom (DOF) and of three DOFs taken respectively from CyberGlove and Virtual Reality Model Language (VRML) models.

We generated 300 motions using CyberGlove, a fully instrumented glove with 22 sensors that provides 22 high-accuracy joint-angle measurements representing the angles of the physical hand joints at different parts of a hand. Each of the 100 different motions has other two similar ones, one more accurate, and one less accurate. Motions can have different durations and lengths, and motion velocities are different at different time for even similar motions.

Similarities between similar motions are computed for the 100 motions, and similarity between each motion and other 99 dissimilar motions are also computed. Fig. 5 shows the similarities of more accurate motions and the highest similarities between each motion and the other 99 different motions. For more accurate motions, similarities between similar motions are higher than those between the same motion and all other dissimilar motion, achieving 100% recognition rate for the 100 different motions. For the less accurate motions, 92% similar motions have higher similarities than dissimilar motions.

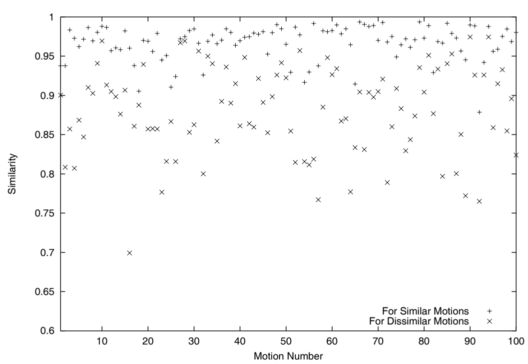

Fig. 5. Motion Similarities

In addition to the one DOF data from CyberGlove, three DOFs data from VRML model NANCY and JAKE have also been generated. These two models have different geometries, and can represent people with different body sizes and proportions. Fourteen common joints such as shoulders, elbows, wrists, knees, and ankles are chosen as each model's sampling points. For each model, 4 basic motions are generated: STRETCH, KICK, WALK, and RUN. KICK and STRETCH involve the motions of arms and legs of the models in different directions. One similar motion is extracted from each of the 4 basic motions for each model by removing alternative rows for different portions of the 4 basic motions, and then repeated for the whole duration, and another set is obtained by reversing the motion directions of the basic motions. Then 4 basic motions and 8 extracted motions of one model are computed for similarities with different motions of the other model. Experiments show that 100% recognition rate can be achieved for the VRML model data.

It's worth noting that high recognition rates have been achieved despite the following challenges:

- Multi-attribute data have 22 attributes for one DOF CyberGlove data, and 42 attributes for three DOF VRML data.
- Data matrices have different lengths and motions have different local rates.
- Motions can have different directions and repetitions.

6. CONCLUSION

We have proposed a new similarity measure based on SVD of multi-attribute motion data, a new medium for signal processing. The first left singular vectors should be similar for similar motions. They are interpolated, and DTW is applied to the interpolated equal length left singular vectors for measuring the closeness of the left singular vectors. The computed DTW distances are normalized so as to be incorporated into an existing similarity measure as proposed in [4]. Experiments from one DOF and three DOF data show that for well controlled accurate similar motions, 100% recognition rate can be achieved, and for less accurate motions, 92% recognition rate can be achieved. The proposed similarity measure can address the motion direction and repetition issues of multi-attribute data with high recognition rates as our motion datasets have demonstrated.

7. REFERENCES

[1] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI-94 Workshop on Knowledge Discovery in Databases*, pages 229–248, 1994.

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

[3] C. Li, P. R. Kulkarni, L. Liu, B. Prabhakaran, and L. Khan. Real-time classification of multivariate motion data using support vector machines (SVM). In *Proceedings of the 5th Int'l Workshop on Multimedia Data Mining (MDM/KDD2004)*, pages 1 – 7, Aug. 2004.

[4] 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.

[5] L. Rabiner and B. Juang. *Fundamentals of speech recognition*. Prentice-Hall, 1993.

[6] C. Shahabi, L. Kaghazian, S. Mehta, A. Ghoting, G. Shanbhag, and M. McLaughlin. Analysis of haptic data for sign language recognition. In *Proceedings of the 9th International Conference on Human Computer Interaction*, pages 441 – 445, Aug. 2001.

[7] C. Shahabi and D. Yan. Real-time pattern isolation and recognition over immersive sensor data streams. In *Proceedings of The 9th Int'l Conference on Multi-Media Modeling*, pages 93–113, January 2003.

[8] 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.