Automatic Segmentation for Virtual Human Motion

Shi Qu*, Tinxin Xu, Liang Ma and Jianxun Liu
Department of Space and Air Early Warning, Air Force Early Warning Academy, Wuhan, 410019, China

*Corresponding author

Abstract. A technique of automatic segmentation for motion data has been proposed, which describes high-dimensional motion with low-dimensional motion character and automatically segments the motion capture data by detecting changes of motion character. Gaussian Process Latent Variable Models has been used to reduce dimension for motion data. In this model, the motion data has been mapped from observation space to latent space. The character function of motion has been constructed in the latent space, and it has excellence including sensitive to all joints, simple construction, and so on. The motion data segmentation points can be detected to complete motion segmentation after analyzing character function by geometry character. The technique in this paper has well adaptation and high correct rate as shown in experiments.

Keywords: Motion data, Hidden space map, Observation space.

1. Introduction

With development and popularization of equipment for motion capture, a new motion generation method has emerged, namely motion capture method, which uses motion capture data to generate the movement of virtual characters, which has the advantages of easy realization and high motion fidelity. At present, some scholars at home and abroad have carried out research on motion segmentation and achieved certain results. Different from the direct extraction of motion characteristics from the angle of joint, some other scholars have studied the segmentation of motion data from a statistical point[1] Three methods are proposed for motion segmentation, and they are based on principal component analysis and probability principal component analysis to realize motion segmentation. These two algorithms have a good effect on motion capture data segmentation, but their computational efficiency is low. In addition, a motion data segmentation algorithm based on a hybrid Gaussian model is also proposed, but the user needs to specify the number of motion segments to be divided, which is inconvenient to use because the number of motion segments contained in the motion sequence is often not known before the segmentation.

In this paper, from the perspective of statistical learning, the data of high-dimensional motion is reduced, the motion data is mapped to hidden space, and the motion feature function is constructed in the hidden space to perform motion segmentation. There are three main advantages to doing this. First, instead of simply considering only a few fixed joints and ignoring other joints, changes in any dimension of motion data will cause changes in hidden variables, thus changing the value of the motion characteristic function. That is, the motion characteristic function is sensitive to all joints and the motion segmentation is universal; At the same time, the motion feature function is based on low dimensional implicit variable construction, the structure is relatively simple and the computational efficiency is high; Finally, the motion data reduction processing process does not require artificial intervention at all. It can be
automatically completed in the background and carried out in parallel with motion segmentation, so it will not affect the efficiency of motion segmentation.

2. Hidden Space Map

Due to the complexity of human structure, moving data is usually of higher dimension. If we divide it directly, it will bring huge computational complexity. A feasible idea is to map high-dimensional motion data to low-dimensional space through low-dimensional technology and divide it in low-dimensional space. However, after reducing dimension, it is necessary to retain the main characteristics of motion data. It can ensure that motion segmentation point in latent space can segment the motion correctly. Principal component analysis is a commonly used low-dimensional technology first proposed by Karl Pearson[5]. However, it is a simple linear reduction dimension, lacking a probability description of the motion space. Because motion data is very complex, we need a nonlinear method to reduce dimension. Through experiments, we find that the Gaussian process implicit variable model is an effective method. Documentation[3] Using GPLVM to achieve the solution of reverse kinematics, we make the following three improvements:

- In the feature vector, remove the world coordinates that represent the position of the human body. Because the motion segmentation needs to consider the difference between the posture itself, regardless of the position of the human body.
- In the kernel function, the cosine component of the angle of the vector is introduced, and the sum of the Gaussian kernel function and the cosine of the angle of the vector is used to reflect the measure of hidden space similarity.
- When training GPLVM, literature[8] Using the method of fixed number of iterations, we introduce the convergence judgment condition and stop the iteration when the condition is satisfied.

Suppose there is a motion capture data \( \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N]^T \), where \( \mathbf{y}_i \) represents the characteristic vector of the human posture at the time \( t_i \), consisting of the rotation angles of each joint, excluding the world coordinates of the human body, \( \mathbf{y}_i \in \mathbb{R}^D \), The Gaussian process latent variable model maps motion data from the \( D \)-dimensional space \( \mathbf{y} \) to \( q \)-dimensional space \( \mathbf{x} \). The dimension of \( \mathbf{x} \) is generally 2 or 3 dimensions. The following method is used to determine:

\[
q = \begin{cases} 
2 & \text{when } D \leq 60 \\
3 & \text{when } D > 60 
\end{cases} \tag{1}
\]

When using the Gaussian process implicit variable model to reduce dimension for the motion data, we use the kernel function:

\[
k(\mathbf{x}_i, \mathbf{x}_j) = \alpha \exp(-\frac{\gamma}{2}(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)) + \eta \exp(-\frac{\mathbf{x}_i \cdot \mathbf{x}_j}{|\mathbf{x}_i||\mathbf{x}_j|}) + \delta_{i,j}\beta^{-1} \tag{2}
\]

Where \( \mathbf{x}_j \) is hidden variable corresponding to \( \mathbf{y}_j \), \( \mathbf{x}_j \in \mathbb{R}^q \), \( \alpha, \eta \) are the proportional factors, represents the degree of association of the two points of hidden space, \( \gamma \) represents the Gaussian distribution variance, \( \beta \) represents noise, \( \delta_{i,j} \) is the Kronecker Delta function, and when \( x_i, x_j \) is the same point \( \delta_{i,j} = 1 \), Otherwise \( \delta_{i,j} = 0 \). Given \( N \) vector \( \{\mathbf{x}_i\} \), a nuclear matrix \( K \) can be defined by formula(2), \( K \) is a \( N \times N \) Matrix, \( K_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) \). In the GPLVM model, both motion data and hidden variables obey the Gaussian distribution. Through the integral and Bayesian rules, the conditional probability can be obtained:

\[
p(\mathbf{Y} | \mathbf{X}, \alpha, \beta, \gamma) = \frac{1}{(2\pi)^{DN/2}|K|^{1/2}}\exp(-\frac{1}{2}tr(K^{-1}\mathbf{YY}^T)) \tag{3}
\]
Where $X = [x_1, x_2, \ldots, x_N]^T$ is the implicit variable matrix. Assuming that hidden variable $x$ obeys standard q-dimensional Gaussian distribution, $x \sim N(0, I)$, we get logarithmic likelihood function:

$$L_{GP} = - \ln p(X, \alpha, \beta, \gamma, \eta | \mathbf{Y}) = \frac{D}{2} \ln | \mathbf{K} | + \frac{1}{2} \sum_i Y_i^T K^{-1} Y_i + \frac{1}{2} \sum_i x_i^T x_i + \ln(\alpha \beta \gamma \eta) \quad (4)$$

Here $D$ is dimension of motion data, $Y_k$ is the $k$ column of $Y$, and $\alpha, \beta, \gamma, \eta$ has the same meaning (2). By minimizing (4), the maximum likelihood estimate of parameter $\alpha, \beta, \gamma, \eta$ can be obtained. After $\alpha, \beta, \gamma, \eta$ is determined, it is possible to use (5) to make a maximum likelihood estimate of $x$, achieve reduced dimensions, and map motion capture data to Q dimensional hidden space.

$$L_{L(x,y)} = \frac{\| y - g(x) \|^2}{2\sigma^2(x)} + \frac{D}{2} \ln \sigma^2(x) + \frac{1}{2} \| x \|^2 \quad (5)$$

Where,

$$g(x) = Y^T K^{-1} k(x) \quad (6)$$

$$\sigma^2(x) = k(x, x) - k(x)^T K^{-1} k(x) \quad (7)$$

Here $k(x)$ is $N$ dimensional column vector and the $i$ element is $k(x_i, x)$. Motion capture data is generally long sequence data. In order to improve the computational efficiency of the reduced dimension, at each iteration, we choose a subset of motion capture data to participate in the calculation, which is called an active subset.

### 3. Automatic Motion Segmentation

In the long sequence motion capture data, different types of motion segments have different motion characteristics, and in the feature function they show different geometric characteristics. Motion data is mapped to hidden space by descending dimensions. It has two characteristics: a point with similar hidden space, and the corresponding motion posture is similar; Hidden space is sensitive to all joints, and changes in any dimension of motion data can cause changes in hidden variables. Therefore, we can use the q-dimensional hidden variable of hidden space to construct a motion feature function to represent the characteristics of the motion, and use the geometric characteristics of the motion feature function to automatically detect the motion segmentation point. The process of automatic segmentation of motion data is shown as Figure 1.

![Figure 1. Automatic Segmentation Process for Motion Capture Data.](image)

The vector and matrix are used to formalize the data of the motion capture to be divided. First, the order of the joints that make up the motion and the degree of freedom of each joint are determined. The variable of freedom of all joints constitutes the feature vector $y$, and $y$ is the $D$ dimensional column vector. Then each frame of the motion capture data is read in turn, and each element of the feature vector is assigned in sequence according to the determined joint order. The feature vector of frame $i$ is labeled
and $i$ is a positive integer. For the motion capture data to be divided containing $N$ frames, it can be represented by Matrix $Y' = [y'_1, y'_2, \cdots, y'_N]^T$ and $N$ is a positive integer.

The mean vector calculation formula for matrix $Y'$ is:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y'_i$$  \hspace{1cm} (8)

Subtract the mean vector from each row of matrix $Y'$ to obtain the 0-mean motion capture data $Y$:

$$Y = [y_1, y_2, \cdots, y_N]^T = [y'_1 - \bar{y}, y'_2 - \bar{y}, \cdots, y'_N - \bar{y}]^T$$  \hspace{1cm} (9)

With $Y$ as the training sample, the Gaussian process implicit variable model is trained, and the motion data is mapped from the observation space to hidden space:

$$[y_1, y_2, \cdots, y_N]^T \rightarrow [x_1, x_2, \cdots, x_N]^T$$  \hspace{1cm} (10)

In order to construct a motion feature function, first, calculate the reference point coordinates of the hidden space. The calculation formula is:

$$\bar{x}_j = \min_{1 \leq i \leq N} X_{i,j}$$  \hspace{1cm} (11)

Where $X = [x_1, x_2, \cdots, x_N]^T$. We use the distance between the implicit variable and the reference point as the motion feature to construct the following motion characteristic function:

$$f(x) = ||x - \bar{x}||^2 = \sum_{j=1}^{D} (x_j - \bar{x}_j)^2$$  \hspace{1cm} (12)

Each $x_j$ in the hidden space is substituted into the motion feature function, and the motion characteristic function value corresponding to each frame of the motion capture data is calculated. The motion data is converted into a series of discrete values:

$$Motion(N) = [f(x_1), f(x_2), \cdots, f(x_N)]$$  \hspace{1cm} (13)

With the change of the geometric characteristics of the motion feature function, the segmentation point of the motion segment can be detected and the automatic segmentation of the motion capture data can be realized.

4. Results and Analysis
To verify the validity and correctness of our algorithm, the data of human motion capture is divided by using this algorithm. The experiment used the mocap human motion capture database at Carnegie Mellon University in the United States[6]. The feature vector is 56 dimensions and the hidden space is 2 dimensions, i.e., $D = 56, \quad q = 2$. As shown in Figure 2A, the capture data for the split motion consists of two sports segments, kick and boxing, for a total of 980 frames. Figure 2B is the result of automatic detection of motion segmentation points, of which the local maximum value is 9 and the local minimum value is 8. Two motion segmentation points are detected, which are located at 643 frames and 660 frames, respectively. One of them is arbitrarily selected as a motion segmentation point. Both can be divided into two sports segments: kick and boxing.
a. Motion Capture Data To Be Split

b. Automatic Detection for Segmentation Points

Figure 2. Automatic Segmentation of Human Motion Data.

Table 1 is a comparison of the performance of this method with that of other methods. The performance index includes the checking rate and checking rate of the segmentation point, and the result of manual segmentation is used as the evaluation standard. The checking rate refers to the proportion of the correct segmentation point in the segmentation point automatically detected by the invention. The total checking rate refers to the proportion of the correct segmentation point to the manual segmentation point automatically detected by the present invention. Compared with the method of extracting motion characteristics directly from a number of fixed joints, the feature function constructed in this paper is sensitive to all joints, so it has a higher search rate, but the detection rate is slightly lower. Compared with the Gaussian hybrid model method, the accuracy rate and the completion rate are both higher, and the need not to specify the number of motion segments in advance by the user is another obvious advantage.

| Methods                                | Accuracy rate(%) | Total rate(%) |
|----------------------------------------|------------------|---------------|
| Method of feature function of          | 86.4             | 91.3          |
| structural motion of fixed joint       |                  |               |
| Method Based on Gaussian Mixed Model   | 76.7             | 71.5          |
| Model                                  |                  |               |
| Our method                             | 85.3             | 96.6          |

5. Conclusion
The method of automatic segmentation of motion data overcomes the shortcomings of the lack of universality of the method of constructing motion feature function by fixed joints, and overcomes the limitation that the method based on Gaussian mixed model requires the user to specify the number of motion segments in advance. The automatic segmentation of motion capture data can be accomplished well. Further improving the training algorithm of the Gaussian process implicit variable model and improving the convergence speed are the contents of the next research.
Acknowledgement
This work is supported by the National Nature Science Funding No. 61602506, No. 11805278 and the Hu Bei Province Nature Science Funding No. 2016CFB307.

References
[1] Hervé Abdi, Williams L. J. Principal component analysis[J]. Wiley Interdisciplinary Reviews Computational Statistics, 2010, 2(4):433-459.
[2] Ding C, He X. Cluster Structure of K-means Clustering via Principal Component Analysis[J]. Lecture Notes in Computer Science, 2004, 46(4):414-418.
[3] Reda Elbasiony, walid gomaa. WiFi Localization For Mobile Robots based on Random Forests and GPLVM[C]// 13th International Conference on Machine Learning and Applications (ICMLA'14). 2014.
[4] Wilson, Andrew Gordon, Adams, Ryan Prescott. Gaussian Process Kernels for Pattern Discovery and Extrapolation[J]. Microtome Publishing, 2013, 13(4):págs. 129-135.
[5] Brian Anderson, Shaoshuai Mou, A. Morse,etc. Decentralized gradient algorithm for solution of a linear equation[J]. IEEE Transactions on Parallel & Distributed Systems, 2017, 6(3):319-328.
[6] Dunia R, Qin S J , Edgar T F , et al. Identification of Faulty Sensors Using Principal Component Analysis[J]. Aiche Journal, 2010, 42(10):2797-2812.
[7] Kolenikov S , Angeles G. Socioeconomic Status Measurement with Discrete Proxy Variables: Is Principal Component Analysis a Reliable Answer?[J]. Review of Income and Wealth, 2009, 55(1):128-165.
[8] Kwak N. Principal Component Analysis by, -Norm Maximization[J]. IEEE Transactions on Cybernetics, 2014, 44(5):594-609.