Analysis and Synthesis of Human Motion Function Data Based on Decision Tree Classification

Ying Li*
Weifang Engineering Vocational College, Qingzhou, Shandong 262500, China

*Corresponding author: 154924200@qq.com

Abstract. At present, motion capture is widely used in computer animation, games, movies and robots, but it is still a difficult problem to synthesize stylized human motion. To solve this problem, a motion synthesis method based on decision tree classification and block principal component analysis is proposed. Block principal component analysis is carried out on the motion data grouped according to the characteristics of human skeleton structure, and low-dimensional subspace parameters with specific semantics are obtained. Triangular constraint is used to block the connection between moving frames which are far apart, thus ensuring the time sequence continuity of segmentation results; In the retrieval process, the similarity of key points is calculated according to different influence degrees in turn; Finally, an efficient motion retrieval simulation system is realized.

Keywords: Decision tree; Human body movement; Motion synthesis; Functional data analysis

1. Introduction
The recognition of human motion posture is a research hotspot in the field of pattern recognition. In recent years, with the rapid development of key technologies such as human-computer interaction and wireless body area network, human motion posture recognition is widely used in sports competition, rehabilitation medicine, health monitoring, somatosensory games and other fields [1-2]. However, as far as its technology itself is concerned, motion capture technology is a copy of motion, and the acquired data is a limited set of motion sequences, which can only be applied to specific environments, so it is necessary to process the captured motion data to meet the needs of new scenes. So how to use the existing human motion capture data more effectively and realize motion analysis and synthesis of human motion capture data has become an important research topic at present.

Motion graph is another effective motion synthesis method [3-4]. Literature [5] put forward the concept of motion graph, and literature [6-7] also put forward many ideas to improve the structure of motion graph. In this method, a directed graph structure is constructed according to the motion data obtained by motion capture technology, and directed links are established between points that meet the transition conditions in different motion segments. In [8], motion segmentation is transformed into a one-dimensional search problem, and a kernel time-series segmentation algorithm is proposed. In this method, a kernel function is used to calculate the inner product between frames, and an objective function based on the relationship between frames is designed. One-dimensional linear search is used...
to find the frame that minimizes the objective function in a window, which is the segmentation point. Then, starting from the segmentation point, start a new round of searching until all the segmentation is completed. Literature [9] takes the segmentation of motion capture data as the verification of this method. This method has strong scalability. The source data set and the target data set need not be the same type of data, but only need to build the dissimilarity matrix between them. Moreover, the framework of the algorithm is not limited to clustering, but also applicable to other types of applications, such as variable selection.

Virtual reality is a comprehensive integration technology, among which motion capture technology is one of the key technologies of virtual reality. Motion capture technology can track the motion process of moving objects by installing sensors for key parts of moving objects. In this paper, we will summarize and improve the previous methods of human motion reuse, motion synthesis and motion charts, and propose a method of human motion control and synthesis based on Principal Component Analysis (PCA) based on decision tree classification, design and implement a comprehensive human motion processing system, and generate motion sequences that meet various needs by processing the prefabricated human motion capture database.

2. Principal component analysis

PCA is an important statistical method to study how to transform multi-index problems into fewer comprehensive indexes. It can transform high-dimensional space problems into low-dimensional space to deal with them, which makes the problems simpler and more intuitive. Moreover, these few comprehensive indexes are not related to each other and can provide most of the information of the original indexes [10].

The basic idea of PCA can be summarized as follows: with the help of an orthogonal transformation, the original random variables related to components are transformed into new variables unrelated to components. From an algebraic point of view, the covariance matrix of the original variables is transformed into diagonal matrix, and from a geometric point of view, the original variable system is transformed into a new orthogonal system, which points to the orthogonal direction where the sample points are scattered most widely, and then the multidimensional variable system is reduced in dimension. According to the viewpoint of feature extraction, PCA is equivalent to an extraction method based on minimum mean square error.

Let \( X = (X_1, X_2, \cdots, X_n)^T \) be an \( n \)-dimensional random variable, and the dimension of each random variable is \( m \), then \( x \) can be expressed as a matrix of \( n \times m \) [11]:

\[
X_{n \times m} = \begin{bmatrix} X_1, X_2, \cdots, X_n \end{bmatrix}^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} (1)
\]

Where, \( X_i = [x_{i1}, x_{i2}, \cdots, x_{im}] \), \( i \in (1, n) \). The mean matrix and covariance matrix of \( X \) are denoted as \( \mu = E(X) \), \( \Sigma = D(X) \) respectively.

In addition, assume that the transformation matrix:

\[
W = \begin{bmatrix} \omega_1, \omega_2, \cdots, \omega_n \end{bmatrix}^T = \begin{bmatrix} \omega_{11} & \omega_{21} & \cdots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \cdots & \omega_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{bmatrix} (2)
\]

Among them, \( \omega_i = [\omega_{i1}, \omega_{i2}, \cdots, \omega_{im}] \).
Consider the following linear transformation:

\[
\begin{align*}
Y_1 &= \omega_{11}X_1 + \omega_{12}X_2 + \cdots + \omega_{1n}X_n = \omega_1^TX \\
Y_2 &= \omega_{21}X_1 + \omega_{22}X_2 + \cdots + \omega_{2n}X_n = \omega_2^TX \\
&\vdots \\
Y_n &= \omega_{n1}X_1 + \omega_{n2}X_2 + \cdots + \omega_{nn}X_n = \omega_n^TX
\end{align*}
\] (3)

In matrix form, it is expressed as:

\[
Y = [Y_1, Y_2, \ldots, Y_n]^T = W^TX
\] (4)

It is necessary to seek a new set of variables \(Y_1, Y_2, Y_d, d < n\), which can fully reflect the information of the original variable \(X_1, X_2, \ldots, X_n\) and are independent of each other.

3. Algorithm principle

3.1. Data preprocessing

Because of the time sequence of human motion, in animation technology, human motion is defined as a motion sequence, which consists of several frames. In each frame, the detailed features of human motion, such as the offset relative to the starting position of motion and the rotation angle of each joint of human body, are uniformly defined as posture. A database containing \(N\) motion segments can be expressed as:

\[
\{x_i(t) \mid n = 1,2,\ldots,N; t = 1,2,\ldots,T_n\}
\] (5)

\(x_i(j)\) is the posture represented by the \(j\)th frame of motion \(i\); The frame number of each motion segment is \(T_n\).

In this paper, D TW is used to regularize similar motion segments in order to normalize the length of each motion segment [12]. D TW adopts the idea of dynamic programming to find the shortest distance between the two matched modes, thus achieving the purpose of nonlinear bending of the two modes. In this paper, a motion sequence is selected as the reference motion, and other motion segments are registered with the reference motion, so as to obtain the time bending function \(w_i\) of each motion segment to the reference motion, which records the time sequence characteristics of motion segment \(x_i\).

Every motion segment \(x_i\) is composed of a matrix of \(P \times T_i\) (\(P\) is the dimension of a motion frame). After DTW normalization, \(x_i\) is changed into a one-line vector \(s_i\) with a length of \(P \times T_r\) (\(T_r\) is the length of reference motion).

Therefore, the spatial features \(S_i\) and temporal features \(W_i\) of the original motion segment \(x_i\) are separated and can be expressed as:

\[
x_i = s_i \otimes w_i \] (6)

\(W_i\) is the mapping function from the time axis of \(x_i\) to the reference motion standard time axis. \(S_i\) is the spatial representation of \(x_i\) on the standard time axis; Operator “\(\otimes\)” means that \(S_i\) is bent back to the original time axis of \(x_i\) itself according to \(W_i\). In order to facilitate the subsequent calculation and discussion, \(S = \{s_1, s_2, \ldots, s_N\}\) is expressed as matrix \([S_1^T, S_2^T, \ldots, S_N^T]^T\) in this paper.
3.2. Data dimension reduction
In this paper, the dimensions of motion data and time alignment curves are reduced. For spatial features, PCA is applied to the aligned motion vector \( z_n \) to get:

\[
z = Z(u) = z_0 + \sum_{i=1}^{K} u_i \cdot a_i
\]  

(7)

In which \( z_0 \) is the mean value of all aligned samples, \( u = (u_1, \cdots, u_K) \) is the main coefficient vector, which is the projection of motion data on each orthogonal feature motion vector \( a_i \), and \( K \leq N < T \). The posture at any standard time \( t \) after dimension reduction can be expressed as:

\[
z(t) = z_0(t) + \sum_{i=1}^{K} u_i \cdot a_i(t)
\]  

(8)

Similarly, PCA also needs to act on the time alignment curve. However, using PCA directly will bring problems, because the time alignment curve must meet the non-negative and monotonically increasing constraints, and PCA operation will break these constraints. In this paper, the transformation in reference [13] is used to map the time alignment curve \( w(t) \) to the new space \( h(t) \):

\[
h(t) = \ln(w(t) - w(t-1)), t=1,2,\cdots, T
\]  

(9)

Let \( w(0) = 0 \), whose inverse transformation is \( w(t) = \sum_{i=1}^{L} \exp[h(i)] \). Note that \( w(t) \) is always non-negative and monotonically increasing. PCA in \( h(t) \) space will not destroy the constraint condition of time alignment curve in \( w(t) \) space. Finally, \( w(t) \) is expressed as:

\[
w(t) = W(v) = \sum_{i=1}^{L} \exp[h(v(i)) + \sum_{j=1}^{L} v_j(i) \cdot b_j]
\]  

(10)

In which \( h_0 \) is the mean value of all alignment curves in \( h(t) \) space, \( b_j \) forms a low-dimensional space, \( v = (v_1, \cdots, v_L) \) is the low-dimensional representation of \( h(t) \) in this space, and \( L \leq N < T \). The result of dimensionality reduction of motion data and time alignment curve is to obtain two vectors \( u \) and \( v \), and two subspaces constructed by vectors \( a_i \) and \( b_j \), which respectively describe the spatial and temporal characteristics of motion.

3.3. Feature extraction
Human skeleton information contained in human motion captured by different motion capture devices is different. In order to adapt to different data structures, in this section, only 15 joints are used for feature extraction without losing key information. These 15 joints are: Head, Neck, Root, Right/Left Shoulder, Right/Left Elbow, Right/Left Hand, Right/Left Hip, Right/Left Knee, Right/Left Foot.

Usually, two motion data are not equal in length in time, so it is necessary to use dynamic time planning (DTW) to align the two data in time dimension, but this causes certain information loss and unnecessary interference. Literature [13] proposes a similar motion retrieval method based on human posture coding. According to the feature representation, the motion data is coded frame by frame, and the continuous frames with the same coding are divided into equivalent segments. Then, the inverted index based on the equivalent segments is established to retrieve the candidate motion, and the final retrieval result is determined by the coding matching algorithm. In reference [12], based on JRD, the variance of joint relative distance (VJRD) is proposed as a feature, and good retrieval results are obtained. Variance indicates the fluctuation range of each JRD in its mean value. Intuitively, two
logically similar movements and some joints move in the same way, which are directly reflected in its VJR D value.

In this section, \( VGPD \) is used as the feature of a motion sequence. For a motion sequence \( M \), it can be formally expressed as \( M = \{F_1, F_2, \ldots, F_r\} \), which contains \( T \) frames of motion data, and \( F_i \) represents one of them. For each frame, 1493 features need to be calculated. The GPD of each frame of a motion sequence is constructed. When a motion sequence obtains its GPD, its \( VGPD \) can be simply calculated by formula (11).

\[
VGPD_M(p) = \frac{1}{T} \sum_{t=1}^{T} (nGPDM(t,p) - \overline{nGPDM(p)})^2
\]  

In the above formula, \( nGPDM(t,p) \) represents the \( GPDM(t,p) \) of normalized N [0,1], and \( \overline{nGPDM(p)} \) represents the average value of \( T \) frame sequence. \( VGPD \) can not only describe the logical correlation between motion and motion, but also is not affected by the length of motion sequence, which simplifies the intermediate process, improves the retrieval efficiency and is beneficial to realize the automation of motion retrieval.

### 3.4. Motion synthesis

The shortest distance between any node in the graph can be obtained by labeling the behavior with the line of the motion graph. However, by directly connecting the motion segments connected by the shortest path, the synthesized motion will have obvious posture mutation, which can not meet the requirements of smooth transition. Therefore, in order to solve this problem in the process of motion synthesis, it is necessary to consider the smooth transition between the motions to be synthesized and deal with it accordingly. In this paper, interpolation is used to deal with it.

Different human motion capture data segments have different spatial positions and orientation information, so before interpolation operation, they need to be aligned to make the human motion capture data segments to be synthesized have the same orientation and position.

It is assumed that there are two human motion capture data segments \( A \) and \( B \) to be synthesized, and \( A_m \) is the last frame for motion segment \( A \), translation information \( T_A(t) = (x_A, y_A, z_A) \) exists at \( A_m \) root node, rotation information \( R_A(t) = (\alpha_A, \beta_A, \gamma_A) \) and \( B_n \) are the first frame of motion segment \( B \), and translation information \( T_B(t) = (x_B, y_B, z_B) \) and rotation information \( R_B(t) = (\alpha_B, \beta_B, \gamma_B) \) exist at \( B_n \) root node. Therefore, the alignment operation is shown in formula (12):

\[
T = T_A(t) - T_B(t) = (x_B - x_A, y_B - y_A, z_B - z_A)
\]
\[
(x', y', z') = (x, y, z) + T
\]
\[
R = R_A(t) - R_B(t) = (\alpha_B - \alpha_A, \beta_B - \beta_A, \gamma_B - \gamma_A)
\]
\[
(\alpha', \beta', \gamma') = (\alpha, \beta, \gamma) + R
\]  

Then, for the motion segments \( A \) and \( B \) to be synthesized, the root node describing the position information adopts linear interpolation, while the joint information describing the rotation adopts spherical linear interpolation SLERP algorithm, as shown in Formula (13). \( P_i \) is the position information after interpolation, \( R_i \) is the rotation information after interpolation, and the record interpolation start point is \( m_i \), end point is \( n_i \), \( \delta \) represents the transition coefficient, and \( L \) describes
the transition length.

\[
\delta = \frac{(i - m_i + 1)}{L} \\
p_k = (1 - \delta)p_i + \delta p_j, i \in (m_i, m_i + L), j \in (n_j - L, n_j) \\
r_k = slerp(R_i, R_j, \delta) = (R_i \sin(1 - \delta \theta) + R_j \sin(\delta \theta))/\sin \theta
\] (13)

4. Experimental results and analysis

4.1. Efficiency analysis

In order to verify the effectiveness of this method, we carried out experiments. A virtual character model with 24 bones was adopted, with a total of 72 degrees of freedom. We collected samples of five kinds of movements, namely, moving in small steps, walking, jogging, going up and down stairs, and obtained 22 complete motion cycles, with a total of 1936 frames, by manual segmentation. In the preprocessing stage, we set the number of resampled frames to be 100, and the dimension after dimension reduction of posture is

![Figure 1 Eigenvalue size corresponding to characteristic motion](image)

After preprocessing, we analyze their functional data. Figure 1 shows the eigenvalues corresponding to feature functions, which are arranged from large to small. It can be seen that the energy contained in the first five-dimensional feature functions accounts for the vast majority (95.32%) of the total energy. From it, we can clearly see the law of motion change: from left to right, the motion posture of human body changes from walking to running; From top to bottom, the motion sequence changes from going up the stairs to going down the stairs, which shows that through the analysis of function data, the change rule of sample motion is extracted.

4.2. Human motion synthesis experiment based on block PCA

In this experiment, the coarse and fine-grained grouping methods in Table 1 and Table 2 are adopted respectively. In which \( z_c \) and \( z_f \) are low-dimensional semantic parameters under coarse-grained grouping and fine-grained grouping respectively, and \( z_{l,m} \) represents the \( m \)-th parameter in the \( l \)-th group of low-dimensional semantic parameters.
Table 1 Coarse-grained grouping $L = 7, M = 22$

| Group number | Grouping semantics                  | Number of basis vectors in a group |
|--------------|-------------------------------------|------------------------------------|
| 1            | Root joint position+orientation     | 3                                  |
| 2            | Trunk                               | 5                                  |
| 3            | Head+neck                           | 1                                  |
| 4            | Right upper limb                    | 4                                  |
| 5            | Left upper limb                     | 3                                  |
| 6            | Right lower limb                    | 3                                  |
| 7            | Left leg                            | 3                                  |

Table 2 Fine-grained grouping $L = 10, M = 25$

| Group number | Grouping semantics                  | Number of basis vectors in a group |
|--------------|-------------------------------------|------------------------------------|
| 1            | Root joint position+orientation     | 1                                  |
| 2            | Lower back half                     | 3                                  |
| 3            | Upper half back                     | 2                                  |
| 4            | Chest                               | 3                                  |
| 5            | Lower half neck                     | 2                                  |
| 6            | Upper half neck                     | 4                                  |
| 7            | Right big arm                       | 4                                  |
| 8            | Right elbow joint                   | 2                                  |
| 9            | Right wrist joint                   | 2                                  |
| 10           | Right foot                          | 2                                  |

Through this experiment, it can be proved that the PCA-based human motion synthesis method can synthesize realistic and natural human motion sequences, but there are obvious problems, that is, the low-dimensional parameters used to control the motion details in this method do not have obvious semantic features, and users can’t visually distinguish the motion details determined by each low-dimensional parameter through observation. In addition, due to the problem of different weights in the principal components obtained by PCA method, the higher the weight, the greater the influence of the low-dimensional parameters corresponding to the principal components on the motion details.

4.3. Comparison of retrieval results of different retrieval algorithms

Figure 2 shows the comparison of the precision of motion retrieval after motion analysis between the traditional decision tree and the improved decision tree. It can be seen that the retrieval precision after the improved decision tree analysis has been significantly improved.

![Comparison of retrieval accuracy between traditional and improved decision tree analysis](image-url)
Table 3 Comparison of retrieval results of different retrieval algorithms

| Motion sequence | Recall ratio(%) | Precision ratio(%) |
|-----------------|----------------|-------------------|
| Walk            | 84             | 98                |
| Punch           | 75             | 95                |
| Kick            | 86             | 89                |

Table 3 lists the recall and precision of the retrieval algorithm based on eight-segment bone features and key frame extraction for the same database. It can be found that motion retrieval based on time series features and decision tree learning has better effect and efficiency, and is also very practical for more complex motions.

5. Conclusion

In this paper, we propose a method of human motion synthesis based on PCA of function data of decision tree classification. Using PCA of function data, we can obtain a function subspace composed of a group of characteristic motions from motion capture data. This method encapsulates the characteristics of motion data in two dimensions of time and space by using an independent spatio-temporal feature space model, which can better deal with the dynamic characteristics of motion. It solves the problem that the low-dimensional parameters in the parametric model depend on all the high-dimensional data, which makes the parameters difficult to understand. Experiments show that this method can provide simple and real-time interactive motion synthesis, and users only need to adjust specific semantic parameters to generate the required motion results. Although this paper has achieved good results in walking motion, the next step is to extend this method to other types of motion, which requires improving the time alignment and low-dimensional motion model.

References

[1] Pang zhihong. Analysis of automatic capture of human motion data and generation of dance spectrum. China cable TV, vol. 406, no. 05, pp. 509-511, 2019.
[2] Li Chunsheng, Jiao Haitao, Liu Peng, et al. Improvement and application of decision tree classification algorithm based on C4.5 . Computer Technology and Development, vol. 030, no. 005, pp. 185-189, 2020.
[3] Dong R, Cai D, Ikuno S. Motion Capture Data Analysis in the Instantaneous Frequency-Domain Using Hilbert-Huang Transform. Sensors, vol. 20, no. 22, pp. 6534, 2020.
[4] Liu Jian. Research on intelligent boundary recognition method of low-dimensional human motion data. Microelectronics and Computer, vol.35, no. 12, pp. 147-150, 2018.
[5] Ren Yi. Research on massive medical image data mining method based on decision tree. Electronic Design Engineering, vol. 027, no. 006, pp. 33-36, 2019.
[6] Loreen P, Thomas N, Markus W, et al. Analysis and Quantification of Repetitive Motion in Long-Term Rehabilitation. IEEE Journal of Biomedical & Health Informatics, 2018, PP:1-1.
[7] Li Shaoshuai, Luo Yufeng, Shi Zhixin, et al. Configuration optimization and simulation analysis of three-rotation parallel mechanism based on ankle joint motion characteristics . Mechanical Transmission, vol.43, no. 03, pp. 82-88, 2019.
[8] Pogrzeba L, Neumann T, Wacker M, et al. Analysis and Quantification of Repetitive Motion in Long-Term Rehabilitation. Biomedical and Health Informatics, IEEE Journal of, vol. 23, no. 3, pp. 1075-1085, 2019.
[9] Lu Jinze, Zhang Yuan, Han Xie. Improved human motion detection method. Computer engineering and design, vol .4, no. 09, pp. 68-73, 2019.
[10] Rühl, Ria Maxine, Bauermann T, Dieterich M, et al. Functional correlate and delineated connectivity pattern of human motion aftereffect responses substantiate a subjacent visual-vestibular interaction. Neuroimage, 2018, 174;22-34.
[11] Lu yuying. Research on motion recognition based on 3d human motion capture data. Electronic production, vol. 368, no. Z1, pp. 71-73+77, 2019.
[12] Chen Yan, Wang Qiang. Research on data format of human motion capture. Journal of Jiujiang University no. Natural Science Edition), vol. 033, no. 001, pp. 58-60, 2018.
[13] Sun Yuhang, Zhou Jianqin, Zhang Xuefeng. Human motion pattern recognition based on acceleration sensor. Computer System Application, vol. 029, no. 006, pp. 196-203, 2020.