Motion Data Retrieval Model based on Self-similar Matrix DTW Algorithm

Shenshen Liu, Pu Zhou
Langfang Teacher’s University, Langfang 065000, Hebei, China

Yubing Wang
Oriental College, Beijing University of Chinese Medicine, Langfang 065001, Hebei, China

Abstract
In view of the standard DTW algorithm, there are some problems such as poor retrieval accuracy and poor robustness in human motion data retrieval. In order to overcome such problems, this paper proposes a human motion data retrieval model based on self-similar matrix optimization algorithm. A three-dimensional vector is constructed by using the first order differential and the second order differential respectively on the basis of the two time sequences of the original algorithm. The first order differential can provide the shape information of the sequence, reduce the error mapping of the points on different trends, and the second order differential can provide more information, such as the maximum value, the minimum value, the discount point and so on. Then, the self-similar matrix is used to redefine the distance matrix, so that the anti-noise performance can be improved. Simulations show that the improved algorithm has higher retrieval precision than that of standard algorithm, and it also has better robustness in the case of self-occlusion.

Key words: Human Motion Characteristics, Motion Data Retrieval, Three-dimensional Vector, Error Mapping Optimization, Self-Similar Matrix

1. INTRODUCTION

With the progress of sensor technology, human motion capture technology is becoming more and more mature, it has been successfully applied to animation, health care, medical diagnosis, motion analysis, robot control, real scene games, and even home entertainment (Wu, Xia and Wang, 2013). After 30 years of development, the gradually accumulated capture database has provided a good support for the research and application of motion capture technology (Jin and Prabhakaran, 2011). However, Since the motion capture data has the characteristics of high dimension and huge data volume, retrieving the necessary motion sequences from huge human motion capture database quickly and accurately has been one of the problems to be solved urgently in the field of human motion capture (Blei, Ng and Jordan, 2013).

At present, many experts and scholars at domestic and abroad have carried on the related research in the field of movement data detection. Yang Tao and others (Yang, Xiao and Wu, 2013) proposed a key frame extraction method based on layered curve simplification by compressing the original data in the time dimension. Meinard Muller and others (Muller and Roder, 2013) introduced a series of qualitative geometric correlation features on the basis of “Logically similar motion may not be numerically similar” to describe the motion data. Also, they achieved retrieval from huge human motion capture database by using variable retrievals. And on the basis of the above, Gao Yan and others (Gao, Ma and Chen, 2012) presented a scene description language, which mainly includes the three-tier structure of words, actions and scenes.

In order to achieve local retrieval, Yu and others (Yu, Shen and Li, 2015) proposed to use Laban dance to mark the movement of objects. Since each motion sequence of the movement database corresponds to a Laban dance sequence, the authors use the similarity measure method to find a similar Laban dance sequence, and finally these action sequence fragments are combined to retrieve the results. Weiner (Weiner, 2013) first proposed the suffix tree indexing method, and there was a certain degree of development in vector-space based index structure afterwards. In addition, the posture-based approach has also yielded remarkable results. Muller (Muller and Roder, 2015) constructed the geometry characteristics of the joint articulation of human motion in specific position to find the time period of the motion capture data, and achieved the motion retrieval. Filho (Filho, Traina and Traina, 2011) showed a method based on the reference point index (Omni), which selected the reference sequence from the convex hull of the data set, and the reference points are far away from each other. Chiu Zhiyi and others (Chiu, Chao and Wu, 2015) proposed an index map structure based on original data posture distribution on the basis of introducing an affine invariant posture feature.
In order to overcome the shortcomings of DTW algorithm in human motion data retrieval, this paper proposes a human motion data retrieval model based on self-similar matrix DTW algorithm. Simulations show the effectiveness of the improved algorithm.

2. DTW MOTION IDENTIFICATION ALGORITHM FOR JOINT ANGLE TIME SEQUENCE

The motion characteristic parameter extracted in this paper is the joint angle time sequence, in other words, it is a one-dimensional time-varying signal, so the design goal is the classification matter of time-varying feature data.

The main idea of DTW (Dynamic Time Warping) is to use the idea of dynamic programming to find two matching paths with the smallest distance of different length sequences, and the matching path is the mapping relation between points and points of the sequences. For the two sequences with different timelines, the DTW algorithm can eliminate the differences on the time axis, so that the distortion between them is minimized by stretching or warping the time axis of one of the sequences, which can make the degree of overlap of the time axis with the other sequence maximal.

Suppose two time sequences \( Q \) and \( C \), the lengths are \( n \) and \( m \), respectively. Then,

\[
Q = \{q_1, q_2, \ldots, q_n\} \quad \text{(1)}
\]

\[
C = \{c_1, c_2, \ldots, c_m\} \quad \text{(2)}
\]

In order to align two sequences, it is necessary to create a matrix \( n \times m \), and its element \((i^a, j^a)\) is the distance between \( q_i \) and \( c_j \), which denotes as \( d(q_i, c_j) \). Typically Euclidean distance is used, i.e. \( d(q_i, c_j) = (q_i - c_j)^2 \). Each element \((i^a, j^a)\) of the matrix corresponds to alignment points \( q_i \) and \( c_j \), which is shown in Figure 1.

![Figure 1. A schematic view of a warping path \( W \)](image)

Warping path \( W \), which is a continuous set of matrix elements, defines the mapping relation of \( Q \) and \( C \). The \( k \)-th element of \( W \) is defined as \( w_k = (i, j)_k \). Then,

\[
W = \{w_1, w_2, \ldots, w_k\}, \max(m, n) \leq K \leq m + n - 1 \quad \text{(3)}
\]

The warping path must meet the following constraints:

1. Boundary constraints
   \( w_1 = (1, 1) \) and \( w_K = (m, n) \), in fact, this requires that the path start and end points be two ends of the diagonal of the matrix.

2. Continuity constraints
   If \( w_{k-1} = (a', b') \) and \( w_{k} = (a', b') \), \( a - a' \leq 1 \) and \( b - b' \leq 1 \). This limits the size of the warping path allowing only two adjacent matrix elements, including diagonal adjacent elements.

3. Monotonic constraints
   If \( w_k = (a, b) \) and \( w_{k+1} = (a', b') \), \( a - a' \geq 0 \) and \( b - b' \geq 0 \). It constrains the point on the warping path is monotonous.
There are many warping paths satisfying the above constraints, but we are concerned about the path that makes the cost smallest.

\[
DTW(Q,C) = \min \left\{ \sum_{k=1}^{K} w_k / K \right\}
\]

(4)

where the denominator \( K \) is used to compensate for the different lengths of the different paths. The optimal path can be easily determined by dynamic programming, the cumulative distance \( \gamma(i,j) \) can be obtained through distance \( d(q_i, c_j) \), that is, the nearest element with the smallest cumulative value around the current cumulative matrix should be found:

\[
\gamma(i,j) = d(q_i, c_j) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\}
\]

(5)

According to formula (4) and formula (5), we can see that the optimal warping path must pass \((q_i, c_j)\), so that the DTW algorithm will map two points. However, in fact, these two points should not exist mapping relationship.

Thus, it is often found that a point on a sequence is mapped with a plurality of points on another sequence with the same length, as the point is “nearest” to a number of points on another sequence, and we call this situation “singular”, which is shown in figure 2.

![Figure 2. Mapping effect of the traditional DTW algorithm](image.png)

Figure 2 shows two artificial time sequences for theoretical analysis. Let us consider the most special case, we assume that the length of two sequences are the same, but they are different in amplitude and local peak valley values. The mapping effect of the traditional DTW algorithm in this case, it is conceivable that the ideal situation is one by one mapping of all points.
3. DTW ALGORITHM FOR SELF-SIMILAR MATRIX OPTIMIZATION

3.1. DTW Algorithm for Self-Similar Matrix Optimization

This paper improves the error mapping of the traditional DTW algorithm. First of all, on the basis of \( q_i \) and \( c_j \), three-dimensional vectors \( [q_i, \dot{q}_i, \ddot{q}_i] \) and \( [c_j, \dot{c}_j, \ddot{c}_j] \) are constructed respectively by using its first order differential and second order differential. The first order differential can provide the shape information of the sequence, and reduce the error mapping of points on different trends. Second order differential provides more information such as maximum, minimum, and discount points. The original data \( q_i \) and \( c_j \) are also very important, because the first order differential and the second order differential are sensitive to noise, while the original data is relatively stable. The traditional DTW algorithm uses the Euclidean distance to establish the distance matrix. This paper uses the self-similar matrix to redefine:

\[
d(q_i, c_j) = w_1 (q_i - c_j)^2 + w_2 (\dot{q}_i - \dot{c}_j)^2 + w_3 (\ddot{q}_i - \ddot{c}_j)^2
\]

where the first order differential and the second order differential can be defined:

\[
\dot{q}_i = \frac{(q_{i+1} - q_{i-1}) + (q_{i+1} - q_{i-1})}{2}
\]

\[
\ddot{q}_i = q_{i+1} + q_{i-1} - 2q_i
\]

In order to ensure the effect of the first order differential and the second order differential, \( w_1, w_2, w_3 \) should meet the following:

\[
w_1 > w_2 > w_3
\]

3.2. Algorithm Effectiveness Analysis

The best warping path can be found using dynamic programming, and the method is the same as DTW algorithm. The effectiveness of the improved method is verified using the two curves in figure 3. As shown in figure 3, the mapping accuracy between sequences is significantly improved.

![Sequence mapping effect of the improved DTW algorithm](image)

In the actual situation, the joint angle sequence cannot be smooth and stable. The camera’s acquisition frequency, the freedom of the tester action and randomness, as well as a variety of uncertainties of noise, will make the sequence shape uneven. In order to simulate this situation, two sequences in figure 2 are added to the random Gaussian noise, which is shown in figure 4.

The mapping effect of the traditional DTW algorithm is shown in figure 5, we can find that the mapping results of the sequence are extremely confusing. “one to many” situation is more serious. In the arrow pointing place in figure 4, two curves coincide, in other words, two curves have the same Y-axis at this point. According to the previous analysis of the characteristics of the traditional DTW algorithm, there will be “one to many” error mapping, and the results of figure 5 verified this idea. The mapping effect of the improved algorithm is shown in figure 6. It can be found that the mapping effect of the sequence is obviously superior, because the improved DTW algorithm can take full account of the trend of the sequence and the overall situation, while ignore the minutiae.
4. ANALYSIS OF MOTION DATA RETRIEVAL

In order to verify the performance of the improved algorithm proposed in this paper, a ten-bar plane truss is taken as an example to simulate. First of all, the characteristic extraction is performed towards the static posture of the human body. Static posture joint angle calculation results are shown in table 1.
Table 1. Joint angle characteristics of static postures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Numerical</th>
<th>Variable</th>
<th>Numerical</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RS = {\theta, \phi}$</td>
<td>[112.14,135.65]</td>
<td>$LS = {\theta, \phi}$</td>
<td>[175.57,176.58]</td>
</tr>
<tr>
<td>$RE = {\alpha}$</td>
<td>144.58</td>
<td>$LE = {\alpha}$</td>
<td>158.79</td>
</tr>
<tr>
<td>$RH = {\theta, \phi}$</td>
<td>[137.47,163.18]</td>
<td>$LH = {\theta, \phi}$</td>
<td>[93.96,176.26]</td>
</tr>
<tr>
<td>$RK = {\alpha}$</td>
<td>166.15</td>
<td>$LK = {\alpha}$</td>
<td>174.89</td>
</tr>
<tr>
<td>$Co = {\beta}$</td>
<td>19.12</td>
<td>$Sa = {\beta}$</td>
<td>92.98</td>
</tr>
</tbody>
</table>

From table 1, we can see that the improved algorithm proposed in this paper has been quite accurate in measuring the joint angle.

Suppose the existing action template feature matrix is: $AR = (R_1, ..., R_k)$, the feature matrix of the sample to be tested is: $AT = (T_1, ..., T_k)$. $R_k$ and $T_k$ are column vectors, ($k \in [1,M]$). The length are normally not equal. Let $D_k$ be the distance between $R_k$ and $T_k$ of the improved DTW algorithm, then, the distance between $AR$ and $AT$ is:

$$D(AR, AT) = \{D_1, ..., D_k, ..., D_M\}$$

(10)

Expected distance is calculated as the final judgment distance:

$$ED = \sum_{k=1}^{M} w_k D_k$$

(11)

where $w_k$ represents the corresponding weight. Assuming that there is a $C$ class template in the reference template library, a test sample $AT$ is given. The template which can make the minimum value of $ED$ has its corresponding action category, and take this action category as the search result.

$$Decision = \arg \min_{i \in C} \{ED_i\}$$

(12)

In addition, the similarity of the curves is observed. More attention needs to be paid to the overall trend of the sequence rather than the specific value of a time point. Due to the influence of the camera, tester and other uncertainties of noise, data changes and unnecessary fluctuations can be produced. Thus, Smooth processing of the sequence is carried out by using $n$ order moving average:

$$\frac{q_1 + q_2 + ... + q_n}{n}, \frac{q_2 + q_3 + ... + q_{n+1}}{n}, \frac{q_3 + q_4 + ... + q_{n+2}}{n}, ... ,$$

(13)

Figure 7. $n$ order moving average smoothing

The greater $n$, the smoother the curve is, which is shown in figure 7. However, in order to retain the characteristics of the original data, $n$ is generally taken as: $n = 3$.

Algorithm validation is performed in the human motion database. The database contains six kinds motions collected indoor, and these motions are given specific meaning: arm beat, head hands together, salute, up and
down cross arm, one hand punches, sideways bent over. 6 motions are abbreviated as A1~A6. Each motion is completed by 6 people, each motion is repeated 5 times. The first 3 times of motions face straight to the camera, and the latter 2 times face the camera with ±45°angle, respectively. The acquisition rate is 30FPS, total groups of motions are 180 times, and the data has 6300 units.

In this paper, 1/6 samples are randomly selected as the reference template and 5/6 as test samples. The total tests are 6 times. The average value is taken as the final result. In this paper, we all select the static features, that is, $M = 16$. For simplicity, take $w_1 = 1$, $w_2 = 2$, $w_3 = 3$, and all the weight values are $1/M$. The search results of the database are shown in table 2. The comparison with the traditional DTW algorithm is also shown in table 2.

<table>
<thead>
<tr>
<th>Action category</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTW</td>
<td>92.5</td>
<td>91.6</td>
<td>95.4</td>
</tr>
<tr>
<td>Improved DTW</td>
<td>98.7</td>
<td>99.4</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action category</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTW</td>
<td>91.6</td>
<td>93.6</td>
<td>94.8</td>
</tr>
<tr>
<td>Improved DTW</td>
<td>99.8</td>
<td>98.6</td>
<td>99.7</td>
</tr>
</tbody>
</table>

By using the algorithm in this paper, the average retrieval rate of the algorithm is 99.37%. Because of small sample data types of self-made samples, the recognition rate reaches 100% in some motions. The recognition rate of other motions by using the improved algorithm is higher than that of the traditional method.

In addition, the self-occlusion of human motion is very common, but the distance information of the scene can be obtained from the depth image. With the help of posture estimation by Kinect machine, to a certain extent, the occluded joints can be identified effectively. Therefore, this article also carried out experiments on the motion under occlusion, which is shown in figure 8. Of which, 0° shows the tester is facing the camera straight with no occlusion, and ±45° shows the tester is facing the camera sideways with ±45° angle, which has occlusion.

![Figure 8. Motion recognition effect under self-occlusion](image)

From figure 8, we can see that there is still a high retrieval rate in the case of self-occlusion. The advantage of depth image in solving human motion self-occlusion is obvious.

5. CONCLUSIONS:

With the rapid development of multimedia technology and the widespread use of human motion capture equipment, a large number of realistic three-dimensional human motion capture data are gradually accumulated. They are effectively applied to three-dimensional animation, film production games and so on. However, how to manage and reuse the large number of data effectively has become the new problem for the developers. In order to overcome the shortcomings of DTW algorithm in human motion data retrieval, this paper proposes the human motion data retrieval model based on self-similar matrix DTW algorithm. Simulations show that the improved algorithm has higher retrieval precision than that of standard algorithm, and it also has better robustness in the case of self-occlusion.
REFERENCES:


