Hindawi
Complexity
Volume 2020, Article ID 6650924, 11 pages
https://doi.org/10.1155/2020/6650924

Research Article

Human Motion Data Retrieval Based on Staged Dynamic Time Deformation Optimization Algorithm

Hongshu Bao and Xiang Yao

Department of Sports, Anhui University of Technology, Maanshan 243000, Anhui, China

Correspondence should be addressed to Xiang Yao; yaoxiang0816@ahut.edu.cn

Received 23 October 2020; Revised 16 November 2020; Accepted 3 December 2020; Published 14 December 2020

Academic Editor: Wei Wang

In recent years, with the rapid development of computer storage capabilities and network transmission capabilities, users can easily share their own video and image information on social networking sites, and the amount of multimedia data on the network is rapidly increasing. With the continuous increase of the amount of data in the network, the establishment of effective automated data management methods and search methods has become an increasingly urgent need. This paper proposes a retrieval method of human motion data based on motion capture in index space. By extracting key frames from the original motion to perform horizontal dimensionality reduction and defining features based on Laban motion analysis, the motion segment is subjected to vertical feature dimensionality reduction. After extracting features from the input motion segment, motion matching is performed on the index space. This paper designs the optimization method of the phased dynamic time deformation algorithm in time efficiency and analyzes the optimization method of the phased dynamic time deformation algorithm in time complexity. Considering the time efficiency redundancy, this paper optimizes the time complexity of the phased dynamic time deformation method. This improves the time efficiency of the staged dynamic time warping algorithm, making it suitable for larger-scale human motion data problems. Experiments show that the method in this paper has the advantage of speed, is more in line with the semantics of human motion, and can meet the retrieval requirements of human motion databases.

1. Introduction

Among the big propositions of video retrieval, the problem of video retrieval for human posture has gradually attracted the attention of researchers due to its wide application [1]. Recognizing human behavior is one of the most important topics. Human research in this field was first initiated at the end of the 20th century [2]. At present, many research results have been proposed, and some of the research methods have also been applied in real life [3, 4]. The research on video retrieval can help reduce the workload of manual annotation and at the same time greatly improve the extraction rate of information in the video so as to realize the automatic management of video human motion data [5]. For video websites, the use of better-performing video retrieval methods also means that they can provide customers with more accurate retrieval results and can retrieve suitable videos for customers according to their preferences, thereby obtaining more revenue [6].

Motion editing can usually solve the situation that the original captured motion data does not meet the needs. Common uses include changes to the original motion role or sports style. The difficulty is that the original motion can be edited to meet the needs without distortion. Since the emergence of motion capture technology, a large amount of motion capture data has also increased the difficulty of its effective management and reuse. Therefore, how to retrieve the motion database accurately and efficiently has become an important research goal in this field. The current motion retrieval methods mainly include methods such as motion template matching, content-based, index-based, and dynamic time warping. With the maturity of technology, video retrieval technology based on human gesture recognition has begun to attract attention [7]. There are many related research fields in video retrieval based on human gesture recognition, and one of the most important related fields is human gesture recognition [8]. Video positioning and key frame extraction are also more important research areas. The research goal of video positioning is to locate a segment containing the target to be retrieved from a long video. Key frame extraction improves recognition by extracting representative video frames. Accuracy can also reduce the storage space of the video. The researchers converted the video into a weighted undirected graph, and, by solving its largest connected subgraph, the human action recognition and the spatiotemporal positioning in the video were effectively combined, thereby improving the efficiency of the method [9, 10]. Many commonly used video local features are extended from the field of image processing [11]. Related scholars apply the histogram of gradient directions to the video field [12]. Certain research results have been achieved in the recognition of human poses. However, because the video contains too much information, simply extending two-dimensional features to three-dimensional features cannot describe the features of the entire video well. Therefore, the field of research is still in urgent need of breakthrough innovation. Relevant scholars proposed a human body gesture recognition method based on optical flow field and proposed a time segment representation method calculated using optical flow vector and the concept of weighted frame rate [13]. Reordering is a very important part of video retrieval. At this stage, we adjust the retrieval results according to the relative relationship between the samples in the initial retrieval results. A good reordering algorithm can greatly improve the accuracy of retrieval. At present, the most popular reordering algorithm is based on context (neighborhood relationship between samples) information [14]. Most context-based rearrangement algorithms use the neighborhood information between samples to build an undirected graph with edge weights [15]. For a relatively large human motion data set, it is difficult to save all the edge weights of N × N, and for two samples that are far apart, it can be considered that their relationship has only a slight impact on the final retrieval result. Therefore, usually, only the K-nearest neighbor information of each sample is used to construct the map, and a certain algorithm is used to adjust the edge weight or the fusion of multiple graphs and finally extract the rearranged retrieval sequence from the graph [16]. Reordering methods based on context information can be divided into global information and local information [17]. Relevant scholars have improved the manifold sorting, using anchor points to reduce the size of the relationship graph and using the adjacency matrix to speed up the calculation [18–20]. Relevant scholars build a word search tree and then use regular expressions to express the comparison between the sample features and the feature sequence in the tree [21]. However, a large number of trees need to be built to correspond to all the features and certain prior knowledge is required for retrieval. Researchers build an index table of motion sequences that are hierarchically represented by various parts of the body and then use fast string matching algorithms to match motion sequences [22]. Related scholars use a pose-based index map structure to identify the beginning and end frames of candidate motion segments and use DTW to calculate the similarity of motion segments [23]. However, the DTW algorithm only pays attention to the local scaling of the data sequence and cannot achieve good results under the global scaling and uniform scaling scale, and the calculation of this method is relatively time-consuming.

At present, most video retrieval research is more concerned with the accuracy of retrieval methods, but if it is to be applied to real life, its retrieval efficiency is far from the current level of keyword-based retrieval. How to improve the response speed of retrieval under the premise of ensuring retrieval accuracy has become a new research topic. Therefore, video retrieval still has a lot of research space. The most widely used video retrieval method on the Internet still uses keywords for text retrieval. This method has many drawbacks, such as high cost of labeling and loss of information. Therefore, it is very meaningful to study the video retrieval method based on the internal information of the video. This paper analyzes the retrieval method of human body motion capture human motion data and proposes a retrieval scheme based on index space. By extracting key frames from the original motion segment, the complexity of human motion data processing is reduced, and semantic human motion features are defined. Feature extraction is performed on motion segments, which reduces the dimensionality of human motion data and improves the semantic similarity of retrieval results. The technical contributions of this article can be summarized as follows.

First: we create an index space by extracting all the characteristics of the human body motion database and get the retrieval result through the index on the index space. The time complexity of the staged dynamic time warping algorithm is analyzed, and its time efficiency is optimized.

Second: a two-way phased dynamic time deformation is proposed to reduce the calculation of the state. Greedy thinking is applied to the staged dynamic time deformation algorithm, which reduces the state set during state transition. We analyze the relationship between the solved states and organize the solved states reasonably. By successively approaching complex state transitions, the transition time is reduced.

Third: a simulation experiment was carried out. The results show that the proposed method has better time efficiency, higher flexibility, and retrieval accuracy due to better motion semantic feature extraction. This can meet the retrieval needs of a variety of human motion data, as well as the retrieval of large-scale human motion databases.

The rest of this article is organized as follows. Section 2 discusses the key technology of human motion data retrieval. Section 3 constructs the optimization of the phased dynamic time warping algorithm in terms of time efficiency. In Section 4, experiment simulation and result analysis are carried out. Section 5 summarizes the full text.

2. Key Technologies for Human Motion Data Retrieval

2.1. Mathematical Representation of Human Bones and Joints. Euler angle is a sequence that decomposes angular displacement into three rotation angle values around three mutually perpendicular axes. "Angular displacement" means that Euler angles can be used to describe any rotation, but it can also describe the spatial orientation of an object. Euler angles divide the azimuth into rotations around three mutually perpendicular axes, which generally follow the Cartesian coordinate system and are positive in a certain order.

Initially, the object coordinate system and the inertial coordinate system coincide, heading is the amount of rotation around the y-axis, and rightward rotation is positive; that is, it is clockwise when viewed downward along the y-axis. Similarly, pitch is the amount of rotation around the x-axis in the object coordinate system, and bank is the amount of rotation around the z-axis in the object coordinate system. Both follow the left-hand rule and rotate clockwise when viewed from the positive direction of the axis to the origin.

Under normal circumstances, after the rotation of the three Euler angle components, any rotation of the object in the coordinate system can be expressed. The special situation refers to the universal lock; the locked state will appear when rotating. The three components of Euler angles can be defined differently under different conditions. For example, we obtain the data of bvh format human motion from the zxy axis sequence.

Since the three expressions have their own advantages and disadvantages, using different expressions in different scenarios is often convenient for calculation or has characteristics that other expressions cannot match. It is often necessary to convert the three forms of expression to each other.

Euler angle (x, y, z) to rotation matrix is

$$R(x) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & \sin x & -\cos x \\ 0 & \cos x & \sin x \end{bmatrix} \quad R(y) = \begin{bmatrix} \sin y & 1 & -\cos y \\ 1 & 0 & 1 \\ \cos y & -1 & \sin y \end{bmatrix} \quad R(z) = \begin{bmatrix} \sin z & \cos z & 1 \\ \cos z & -\sin z & 1 \\ -1 & 1 & 0 \end{bmatrix}. \tag{1}$$

The quaternion to Euler angle is

$$\begin{aligned} x &= \arctan\left[2\cdot(bc-aw)/1+2\cdot\left(a^2+b^2\right)\right], \\ y &= \arccos[2\cdot(wa-ab) \\ z &= \arctan\left[2\cdot(ac-ab)/1+2\cdot\left(a^2+c^2\right)\right]. \end{aligned} \tag{2}$$

2.2. Motion Feature Definition and Feature Extraction. On the one hand, feature extraction can reduce the dimensionality of the original human motion data, avoiding the direct similarity comparison of high-dimensional human motion data during retrieval, and, on the other hand, effective features can represent the semantics of human motion [24, 25]. The matching of features is more in line with people's cognition of sports, and the retrieval results are more accurate. The pose corresponding to the key frame is usually the boundary pose in human motion, and it is also the most representative pose in the adjacent frames. It plays the role of the essence and the outline in the motion segment, so a motion can use a sequence of key pose. This paper extracts the key frame sequence from the motion segment in the human motion database [26–28]. With reference to Laban motion analysis, the feature set is mainly divided into physical characteristics, dynamic characteristics, and appearance characteristics. The visual feature collection and feature extraction of human motion action are shown in Figure 1.

Body characteristics mainly describe the structure and geometric characteristics of human movement. This category can help identify which part of the human body is moving, and which parts have contact and patterns of movement of human body parts. In the definition of this part of the feature, considering that the degrees of freedom of some joints of the human body are usually less than 3, some pose features that are unlikely to appear can be directly ignored to simplify the feature space.

The action feature indicates whether the human body is in the current state of motion, by calculating whether the displacement between the adjacent posture exceeds the threshold. This feature is mainly for the displacement of the limbs and the whole:

$$G\left(t^j\right) = \left|t_s^j - t_s^{j-1}\right| > \varepsilon_1. \tag{3}$$

The relative position describes the relative position of each limb. For example, the left and right legs are in front or behind the body plane, and the characteristic value before the plane is 1; otherwise, it is 0. This feature can describe the position of each part of the body and is the most important and most part of the defined features.

The orientation information indicates whether the orientation of the upper and lower parts of the human body changes greatly from the basic posture. This feature can describe movements that change orientation information, such as turning around:

$$G\left(t^j\right) = \left|\mathrm{ori}\left(t_s^j\right) - \mathrm{ori}\left(t_s^{j-1}\right)\right| \geq \alpha_1. \tag{4}$$

The contact information between the limbs calculates whether each limb is in contact (whether the distance between the two limbs is less than the threshold), such as the separation and merging of the left and right arms, which can distinguish movements more accurately:

$$G\left(t^j\right) = \left|\mathrm{dis}\left(t_{s1}^j\right)\right| - \left|\mathrm{dis}\left(t_{s2}^j\right)\right| \geq \varepsilon_2. \tag{5}$$

The angle information calculates whether the angle between the limbs in the current posture and the angle between the upper and lower parts of the human body is less than the threshold. For example, the angle between the thigh and the calf and between the upper body and the lower body

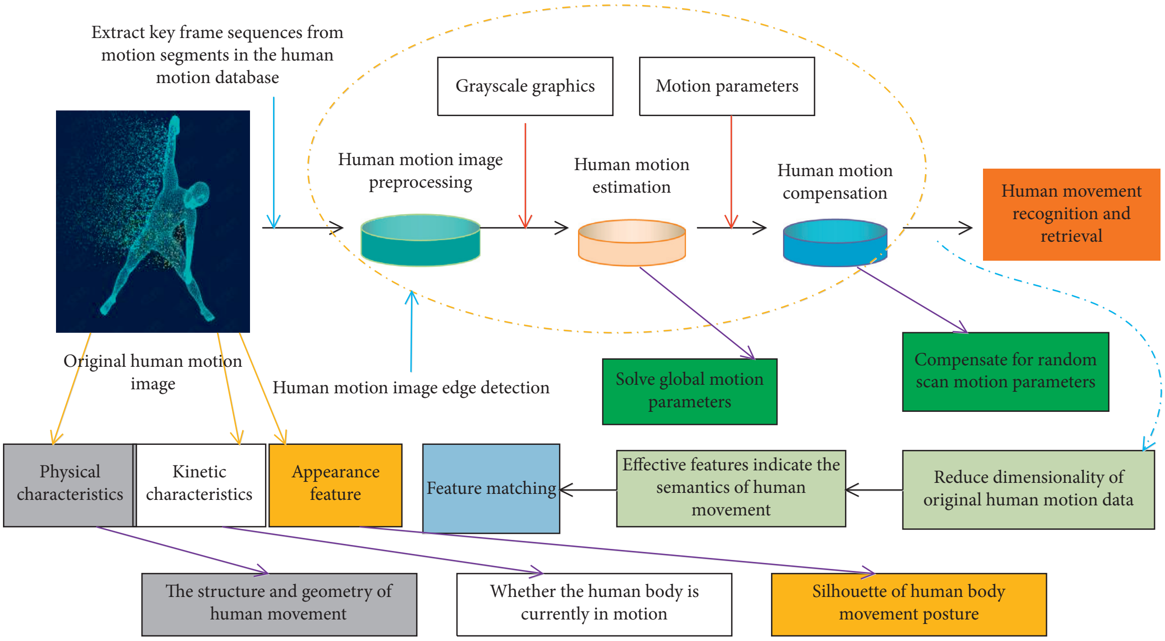
Extract key frame sequences from motion segments in the human motion database
Grayscale graphics
Motion parameters
Human motion image preprocessing
Human motion estimation
Human motion compensation
Human movement recognition and retrieval
Original human motion image
Human motion image edge detection
Solve global motion parameters
Compensate for random scan motion parameters
Physical characteristics
Kinetic characteristics
Appearance feature
Feature matching
Effective features indicate the semantics of human movement
Reduce dimensionality of original human motion data
The structure and geometry of human movement
Whether the human body is currently in motion
Silhouette of human body movement posture

Figure 1: Visual feature collection and feature extraction of human movement.

can be used to judge whether the person is upright or squatting.

Dynamic characteristics are more described as the logical dynamics of actions, and the same body posture may be represented by two actions with completely different semantics. For example, the action of reaching forward, gently handing something, and pushing angrily is completely different in semantics, which cannot be represented by only the geometrical position characteristics.

The spatial feature distinguishes whether the motion trajectory is straight or nonstraight, which is calculated from the distance ratio between the current posture and the previous frame interval:

$$G\left(t^{j}\right)=\sum_{i}^{j}\left|t_{s}^{j}-t_{s}^{i}\right|/\left|t_{s}^{j}-t_{s}^{i-1}\right|\geq\varepsilon_{3}. \qquad (6)$$

The time characteristic corresponds to the dynamic characteristic and distinguishes whether the movement changes suddenly or steadily according to the speed of the movement from small to large. This feature and the intensity feature together can describe in detail the style of the entire process from the start of motion to motion and the end of motion. For example, an eager knock on the door appears suddenly and violently, while a jump on the spot appears suddenly and gently.

The shape feature describes the outline of the human body's movement posture, and the feature mainly defines the different shape states of the human body and limbs.

2.3. Creation of Index Space. This paper defines a feature space to index the entire body motion database in a distributed manner. For a motion posture, the corresponding Boolean feature value sequence can be used as the index number of the posture in the distributed index space, and the posture feature sequence is ANDed with the sequence corresponding to a feature bit of 1 and the remaining bit of 0 to get the value of the posture in a certain feature. In the feature extraction stage, the motion segment number and the corresponding frame number of each pose are recorded as the index information of the pose. Finally, all poses are sorted according to the size of the Boolean feature value and stored in order. It can be seen that the index space extracts the different poses of all motion segments in the human body motion database and contains the information of each pose in the motion segment. The update of the index space is also very easy; it just inserts a new motion posture and index information directly or just adds new index information of the existing posture.

According to the characteristics of the defined index structure, common retrieval requirements such as motion subsequence matching and fuzzy search can be easily realized. When the query motion segment is input, it may be a subsequence of some long motion segment in the human motion database. In retrieval, by querying the feature value sequence corresponding to the first key frame posture of the motion segment, the similar posture in the index space can be quickly located; that is, the beginning of the query motion segment matching in the long motion segment can be found.

When the user needs human motion data that is similar to but not completely consistent with some features of the query motion segment, they can request matching by specifying some features. The fewer the specified features, the more the matching.

This paper proposes a video retrieval framework based on human motion, and its process is shown in Figure 2. Firstly, the human motion data in the human motion data set is preprocessed, including position and size calibration of RGB video and depth video and multiangle calibration. We

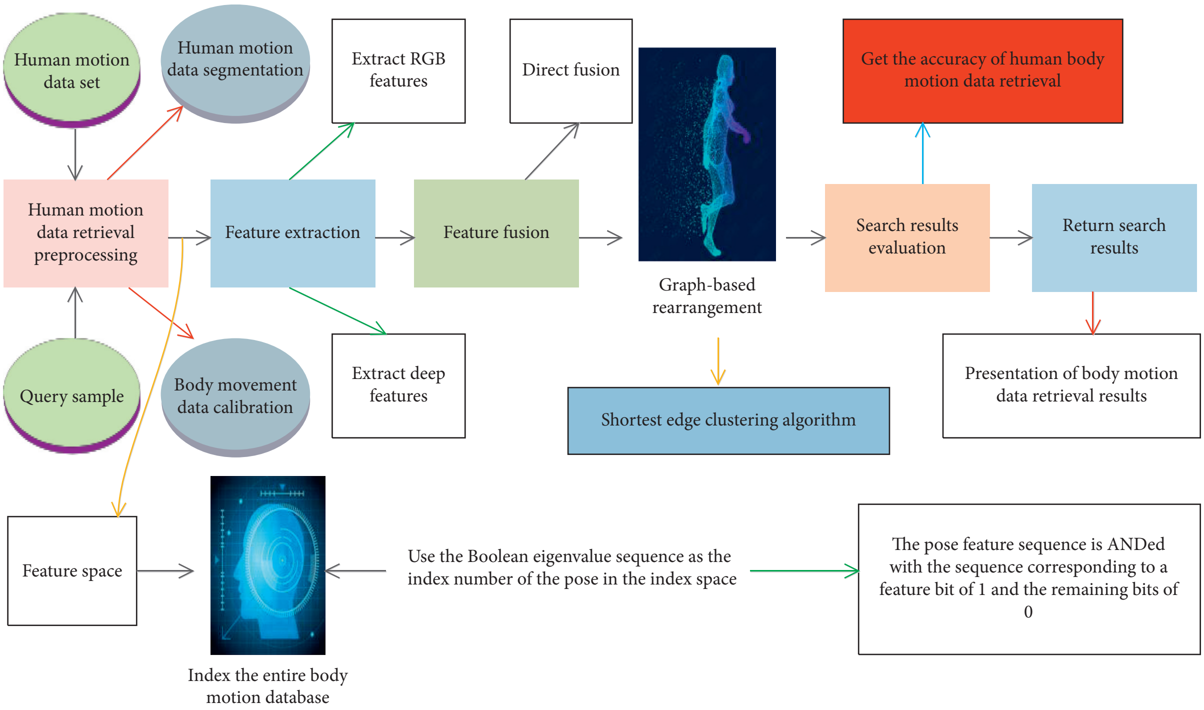
Human motion data set
Human motion data segmentation
Extract RGB features
Direct fusion
Get the accuracy of human body motion data retrieval
Human motion data retrieval preprocessing
Feature extraction
Feature fusion
Graph-based rearrangement
Search results evaluation
Return search results
Query sample
Body movement data calibration
Extract deep features
Shortest edge clustering algorithm
Presentation of body motion data retrieval results
Feature space
Index the entire body motion database
Use the Boolean eigenvalue sequence as the index number of the pose in the index space
The pose feature sequence is ANDed with the sequence corresponding to a feature bit of 1 and the remaining bits of 0

Figure 2: Video retrieval framework based on human motion.

perform feature extraction to get the corresponding RGB and depth features and use a certain metric to preliminarily sort the two features to construct a Jaccard graph. TTNG mapping based on the Jaccard diagram is used to eliminate the influence of outliers on the results. The edge weights are directly added. Using the shortest edge clustering algorithm proposed in this paper, the nodes belonging to other manifolds that are close to each other are removed from the search results. Finally, the nodes in the graph are sorted by weight to obtain the new retrieval result as the return value of the framework.

3. Optimization of the Staged Dynamic Time Deformation Algorithm in Time Efficiency

3.1. Optimization of the Total Number of States. The staged dynamic time deformation algorithm can be implemented efficiently in two ways, namely, forward and backward. In some cases, the same phased dynamic time deformation algorithm adopts different implementation methods, and there will be certain differences in the time efficiency of the program. Generally, if the initial state of the problem is determined but the end state is uncertain, it can be achieved through the "top-down" sequential method. On the contrary, if the end state of the problem is determined but the initial state is uncertain, you can consider the "bottom-up" inverse method.

The breadth-first search algorithm can be combined with the two-way expansion method to reduce the total amount of state, and the staged dynamic time warping algorithm can also be used. Similar to the two-way breadth-first search algorithm, when the state space of the problem is large and the initial state and end state of the problem are determined, one-way expansion will lead to a large number of invalid state calculations. Next, in order to reduce the scale of the state, you can expand in two directions from the initial state and the end state and perform optimal judgments at the intersection of the two expansions to get the solution of the problem.

Figure 3 shows the difference between the two-way expansion and the one-way expansion in the total number of states that need to be calculated. In actual problems, the initial and end states of the problem can be determined. The number of problem states is huge, and the state variables in each stage grow rapidly. At this point, you can consider using a two-way planning method to reduce the total amount of states that need to be calculated in the problem.

The two-way phased dynamic time deformation also has better applications in practical problems, such as the combined service selection problem; that is, among multiple services that provide the same function, specific services are selected for combination, so as to maximize the user's service quality. The service selection problem is abstracted into a graph model. Finally, the problem is transformed into the longest path problem in the graph. The problem is solved with two-way and staged dynamic time deformation, which is more time-efficient than one-way programming.

The number of states to be solved in the problem is directly related to the state representation method, and the total number of states in the problem can usually be changed by improving the state representation method. When using the staged dynamic time deformation idea to solve the problem, the algorithm designed with different state representation methods requires a different total number of states to be solved. When the number of states for each state transition and the time for each state transition are

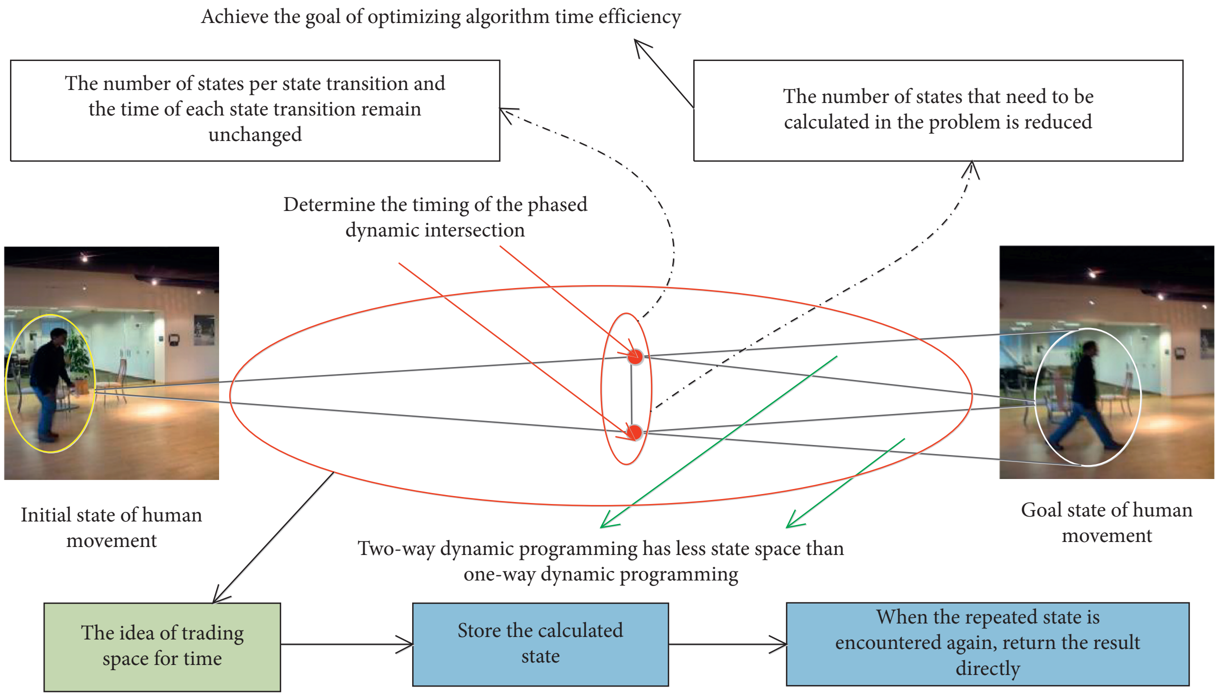
Achieve the goal of optimizing algorithm time efficiency
The number of states per state transition and the time of each state transition remain unchanged
The number of states that need to be calculated in the problem is reduced
Determine the timing of the phased dynamic intersection
Initial state of human movement
Goal state of human movement
Two-way dynamic programming has less state space than one-way dynamic programming
The idea of trading space for time
Store the calculated state
When the repeated state is encountered again, return the result directly

Figure 3: State space diagram of two-way and one-way planning.

unchanged and the number of states that need to be calculated in the problem is reduced, the time efficiency of the algorithm can be improved as a whole.

3.2. Optimization of the Number of States Involved in Each State Transition. The process of solving a problem with a phased dynamic time deformation algorithm is to calculate all the states (subproblems) in the problem. The calculation of the current state is usually through the state that has been solved and the decision in this state, which is the state transition process. The number of states involved in each state transition when calculating states is a key factor affecting the time efficiency of the phased dynamic time deformation algorithm. This section will discuss some optimization methods to reduce the number of states involved in each state transition.

We use the phased dynamic time deformation method to solve the problem when the problem planning model is

$$G(i,j) = \begin{cases} \min[G(i,k-1) + G(k,j) + w(i,j)], & i<k<j, \\ 0, & \text{others}. \end{cases} \tag{7}$$

W(i, j) represents the sum of the metric values from the i-th stage to the j-th stage in the problem. This metric has different meanings in different problems and can be profit, output, and resource consumption. G(i, j) represents the state of the problem in the (i, j) stage. The meaning of the state transition equation is to decompose the original problem into two subproblems and then solve them. The optimal binary search tree and other problems all use this state transition equation.

Among them, if the weight function w satisfies the following formula, it is said that the function w satisfies the monotonicity of the interval inclusion relationship:

$$w(i',j') > w(i,j) \quad (i,j) \longrightarrow (i',j'). \tag{8}$$

If the function w satisfies the following formula, it is said that the function w satisfies the quadrilateral inequality:

$$w(b,c) + w(a,c) > w(a,b) + w(b,d) \quad a<b<c<d \tag{9}$$

When the state f(a, b) satisfies the quadrilateral inequality property, the decision variable v(a, b) satisfies

$$|v(a-1,b)| < |v(a,b-1)| < |v(a+1,b)|. \tag{10}$$

Using the properties of decision variables, the optimized state transition equation is obtained as follows:

$$G(i,j) = \begin{cases} \min[G(i-1,k) + G(k,j-1) + w(i,j-1)], & i<j, \\ 0, & i=j, \\ v(i-1,j) < k < v(i,j+1), & i>j. \end{cases} \tag{11}$$

We use the state transition equation to satisfy the quadrilateral inequality, analyze the relationship between the states, and then deduce that the optimal decision is monotonic. When calculating the state, the monotonicity of the optimal decision is used to reduce the number of states that need to be considered for each state transition, thereby reducing the time complexity of the algorithm. We can get inspiration from this optimization measure. In the process of using the staged dynamic time deformation algorithm to solve the problem, not only can the algorithm be optimized by reducing the total number of states, but also the nature of the optimal decision can be fully utilized to optimize the algorithm.

In the process of solving the problem with the phased dynamic time deformation method, the calculated state in the problem will be continuously quoted. Therefore, in-depth analysis of the relationship between the states that have been solved and a reasonable organization of it can sometimes simplify the dependence of the state in the problem, which helps to improve the time efficiency of the staged dynamic time deformation algorithm. To illustrate through an example model, the problem planning model is as follows:

$$G(i) = \begin{cases} \max\left[G(j)|m(j) < m(i) + 1\right], & i > 1, \\ 1, & \text{others}. \end{cases} \tag{12}$$

In the formula, $m(i)$ represents the condition value at the i-th stage of the problem, G(i) represents the state value at the i-th stage of the problem, and the state transition equation means that only when the condition value of the current stage of the problem is greater than the condition value of the previous stage, the state will be transferred.

There are O(n) states to be solved in the state transition model, the time for each transition is O(1), the number of states for each state transition is O(i), and the total time complexity of the algorithm is

$$O(1 + 2 + 3, \ldots \infty, + n - 1 + n) = O\left(n^2\right). \tag{13}$$

The optimized state transition equation is

$$G(i) = \begin{cases} G(k), & m(k) > m(i) > m(j-1), \\ G(j), & m(i) = m(j-1), \\ 1, & m(j) > m(i-1), \\ \max|G(j)| + 1, & m(j) > m(j-1). \end{cases} \tag{14}$$

The element $m(i)$ in the set D is monotonically increasing. Therefore, in the above state transition process, an efficient search algorithm-binary search can be used, so that the time for each transition is O(log n), and the number of states involved is only O(1). After optimization, the time for each state transition is increased, but the number of states in the state transition is reduced. This reflects the contradiction between time efficiency factors in the optimization process. After optimization, the time complexity of the algorithm is O(nlog n). Therefore, the overall time efficiency of the algorithm is improved.

3.3. Use Greedy Thinking to Optimize the Staged Dynamic Time Warping Algorithm. Greedy thinking will select the optimal solution in the current state of the problem according to a specific greedy strategy when solving a problem, instead of considering the global optimal problem. Therefore, the algorithm cannot get the optimal solution for all optimization problems, but the best practice can often achieve the effect of simplifying the problem model.

We use the phased dynamic time deformation method to solve the problem. When the state space of the problem is huge and the model is complex and the conventional phased dynamic time deformation method is not efficient, you can consider using the greedy idea in the phased dynamic time deformation. You analyze the essence of the problem in depth, find out the redundancy in the algorithm, reduce the range of the allowable decision set that may produce the optimal solution, and then achieve the goal of optimizing the time efficiency of the algorithm.

In actual problems, you will encounter the state transition equation shown in the following equation:

$$G(i,j) = \begin{cases} G(i-1,k) + G(k,j-1) + w(i,j), & i < k < j-1, \\ 0, & \text{others}. \end{cases} \tag{15}$$

W(i, j) represents the sum of the metric values from the i-th stage to the j-th stage in the problem. This metric has different meanings in different problems and can be profit, distance, and resource consumption. G(i, j) represents the state of the problem to stage (i, j). This is a staged dynamic time deformation of a typical interval model. The problem is generally to find the optimal value of the entire interval. The basic feature of this type of problem is that the problem can be decomposed into the form of combining two subproblems. The solution is to enumerate the merge points, decompose the problem into two left and right subproblems, and then merge the optimal solutions of the left and right parts to obtain the optimal solution of the original problem.

The optimization process is to analyze the optimal solution composition of the problem to find out the relationship of the problem state dependence, thereby avoiding the enumeration of some invalid states and reducing the time complexity of the algorithm. This is a classic embodiment of greedy thinking. "Greedy phased dynamic time warping" is not a specific algorithm, but an optimization idea. If you want to flexibly use greedy thinking in the phased dynamic time warping algorithm, the key lies in the in-depth analysis and analysis of the problem. You start with the original staged dynamic time deformation model, analyze the overall structure of the algorithm, and then cleverly use the greedy idea to solve the redundancy in the original staged dynamic time deformation algorithm, so as to achieve the purpose of optimizing the algorithm.

4. Experimental Simulation and Result Analysis

4.1. Human Motion Data Preprocessing. The experimental human motion data in this paper includes about 400 motion segments from the human motion capture database of Carnegie Mellon University, which are divided into 8 categories, as shown in Figure 4. The number of joints of all motion captured human motion data is 23, the frame frequency is 24fps, and the motion segment length ranges from 40 to 5000 frames. This paper implements the whole process of preprocessing and retrieval of human motion data before retrieval. In the experiment, the system runs on a PC with Intel P4 2.7 GHz CPU and 8 GB memory.

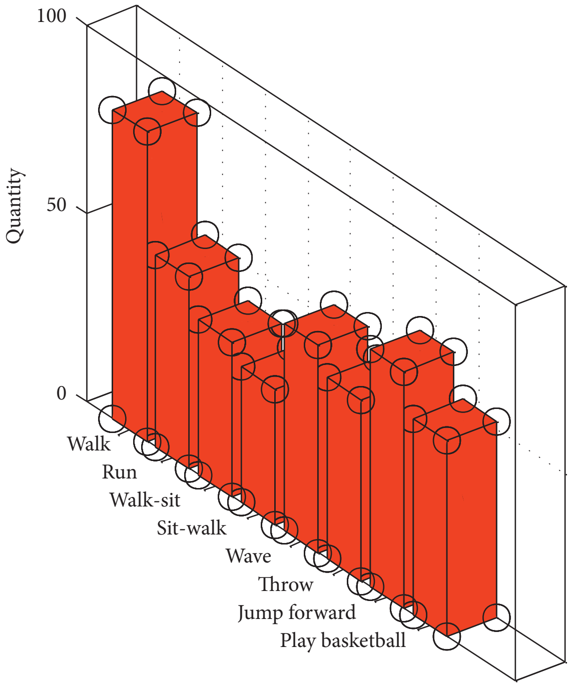

Figure 4: Data classification of the human body movement database.

The preprocessing of human motion data is the offline processing before retrieval of the original human motion data set, including three steps of key frame extraction, feature extraction, and index space construction.

(1) Taking into account the requirements of offline processing, this article adopts a phased dynamic time deformation optimization algorithm that is more computationally intensive but more accurate and stable. The experiment uses the method in the article to extract key frames from all the original human motion data. The optimal compression rate for different actions is different. The optimal compression rate for simple and slow actions is high, while the optimal compression rate for complex and violent actions is lower. For example, the average compression rate of walking slowly is 7%, while the average compression rate of dancing is 16%. In the end, the total number of frames after all human motion data is extracted is 40460 frames, and the total average compression rate is about 10%. It can be seen that the amount of human motion data processing in the later stage is greatly reduced.

(2) A total of 36 Boolean features are defined in this paper. Therefore, in the feature extraction stage, each key frame of the motion segment will be converted into a Boolean sequence of length 36, and the corresponding value range is [0, 224]. Finally, each motion segment is transformed into a feature vector.

(3) We construct an index space to extract each feature of all feature vectors, record the sequence number and frame number of the motion segment, and sort all the features. It can be seen from the feature definition that the size of the index space will not exceed 224, but in practice, some common postures, such as standing still and ordinary walking, will repeatedly appear in multiple different sports; at the same time, some theoretical postures in the feature definition do not basically appear, so the final index space will be much smaller than 224.

In this experiment, a total of 8000 different feature values were finally extracted, and the feature distribution and posture repeatability are shown in Figure 5. According to different human motion data sets, the number of posture repetitions and the distribution of peaks will be different, but it will basically be reflected in a few postures with high repetition, and most postures only appear once or a few times. In the end, although there are not many repetitive postures reduced, it can still effectively reduce the repetitive comparison of the same posture in the retrieval stage when retrieving some motion segments with repetitive actions. The cumulative matching characteristic error curve of the data set is shown in Figure 6.

4.2. Comparison of Similarity Calculation Functions. We select the human motion data test set collected and processed in this paper, use the traditional collaborative filtering similarity function to retrieve the human motion data, and then use the verification set to evaluate the retrieval effect of the algorithm. We draw the changes of accuracy rate and recall rate under different similarity functions, respectively, as shown in Figures 7 and 8.

According to Figure 7, it can be seen that, under different similarity calculation functions, the accuracy based on modified cosine is generally the best. It can be seen from Figure 8 that when using the Pearson correlation coefficient and the modified cosine similarity calculation function, the recall rate is higher than when using the cosine similarity function. Since the recall rate reflects the proportion of the human body motion data purchased by the user that is retrieved, it may be more likely that the algorithm cannot calculate the similarity when the cosine similarity is used. On the whole, the recall rate of the modified cosine similarity calculation function is the highest.

This paper compares the modified cosine similarity calculation function and the dynamic time deformation calculation function and compares the retrieval effect of the function in the interval of Top N from 1 to 25. We draw the changes in accuracy and recall under different similarity functions and different Top N, respectively, as shown in Figures 9 and 10.

It can be seen from Figure 9 that when Top N changes from 1 to 25, the traditional modified cosine similarity retrieval accuracy rate is lower. Only when Top N is selected as 20, the retrieval accuracy is the same as the improved function of this article. On the whole, each similarity calculation function has the highest retrieval accuracy when Top N is 15 selected. This shows that Top N cannot be selected too much because the accuracy rate evaluates the proportion of the correct human motion data to the total number of retrievals, so there will be more retrievals that are

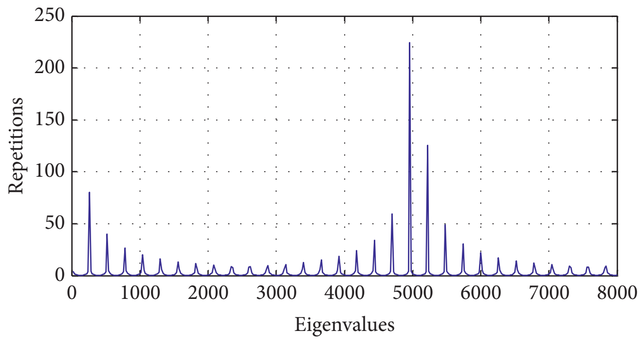

Figure 5: Feature distribution and posture repeatability.

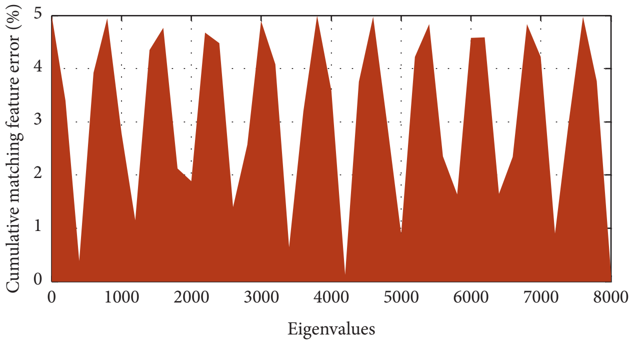

Figure 6: Cumulative matching feature error.

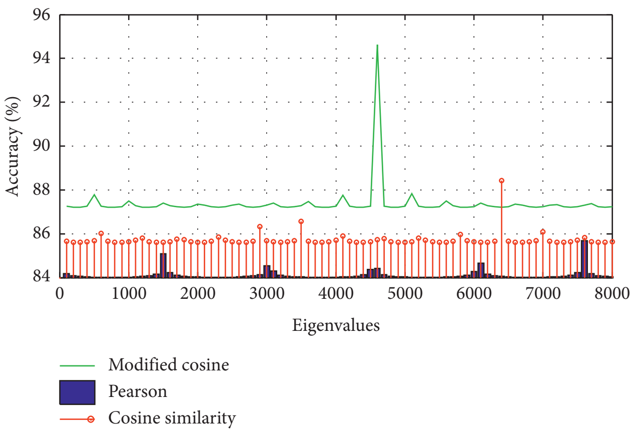

Figure 7: Comparison of accuracy of different similarity calculation functions.

more likely to be inaccurate, and the accuracy will be relatively lower.

It can be seen from Figure 10 that when Top N is selected from 1 to 25, the retrieval recall rate of the improved function proposed in this paper is higher, and when Top N is selected 11, the retrieval recall rate of the modified cosine similarity is the same as the improvement proposed in this paper. The retrieval recall rate of the function is the same. On the whole, although the improved function of this article is slightly worse when Top N chooses 11, it is not much worse. The search effect of the algorithm must consider the accuracy rate and recall rate comprehensively, so it will not have much impact on the final result.

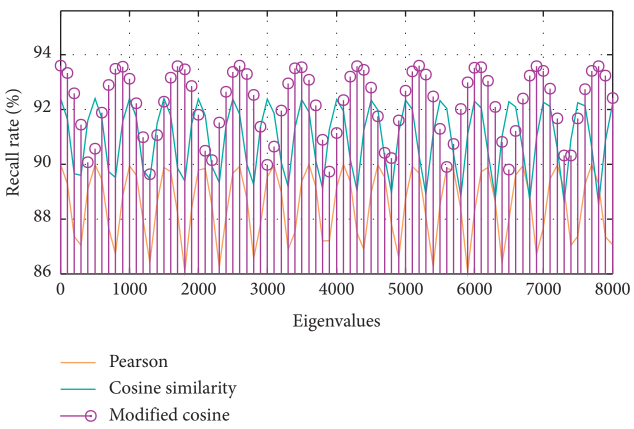

Figure 8: Comparison of recall rates of different similarity calculation functions.

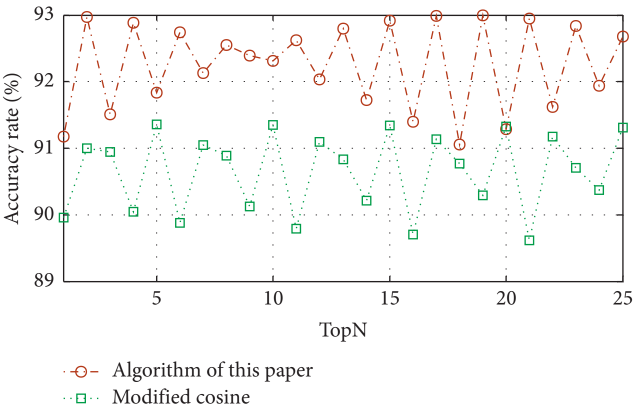

Figure 9: Comparison of accuracy of different similarity functions.

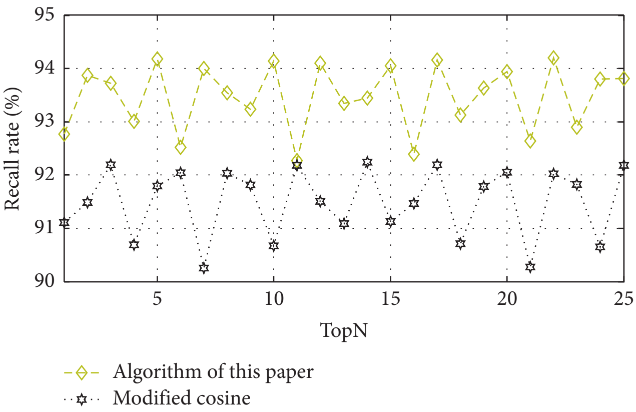

Figure 10: Comparison of recall rates of different similarity functions.

4.3. Human Motion Data Retrieval. After constructing the index space offline, you can search online by entering query fragments and specifying the features to be matched. Among them, query fragments are also transformed into feature vectors through key frame extraction and feature extraction. The retrieval algorithm in this paper uses the binary search to search the index space when searching for the feature sequence corresponding to a certain frame. For the experimental environment of this paper, when the original query motion segment is 300 frames, the average search time is 0.034 s.

In order to compare the retrieval effects of different algorithms in the test human motion database, Figure 11 shows the PR (precision-recall) curve of each algorithm on multiple motion categories. It can be seen that the algorithm in this paper uses feature extraction to convert the original angle value into logical semantic feature values, and, during retrieval, the query motion is adapted to select the features to be matched for retrieval, so it has relatively good precision and recall. Since each dictionary tree corresponds to only one feature, the results need to be merged after matching multiple dictionary trees. The process of merging cannot guarantee the consistency of the timing between the features, and this will also improve the chance of a false match. The method of defining eight-segment bone angle features is more sensitive to human motion data sets. Different styles of motion are difficult to form matching paths in the matching network, so some matching motions with the same semantics but different styles will be lost. In addition, each retrieval requires a large amount of calculation to obtain the similarity and matching path between the query motion and the human motion database motion, which is relatively time-consuming. This paper improves the feature definition and proposes a new index structure and retrieval method, which has the characteristics of flexible, fast, and accurate retrieval and can be used in large-scale human movement database retrieval.

Figure 12 shows the comparison of retrieval time between the algorithm in this paper and the other two methods on different sizes of human motion databases. The experiment uses multiple query fragments of different types and lengths of 400 frames and averages the multiple retrieval times of each method under the same size human motion database. It can be seen from the figure that the method of eight-segment skeletal angle feature has the longest retrieval time because each retrieval has the calculation cost of matching path and similarity. The dictionary tree and the word search tree constructed by the method of regular expression retrieval can reduce repeated matches. But with the increase of the human body motion database, the tree constructed will increase greatly, so the retrieval time will be longer. The method in this paper constructs an index space that has nothing to do with the size of the human body motion database, and the retrieval is performed on the index space, so the retrieval time of the algorithm in this paper is the least.

5. Conclusion

This paper proposes a new method for retrieving human motion data for motion capture, which effectively reduces the dimensionality of human motion data through key frame extraction and custom feature extraction. An index space is defined to record all motion posture distribution information in the human body motion database. In the retrieval phase, the fast query index space is used to match motion segments. This paper discusses the optimization methods of the phased dynamic time deformation algorithm in terms of time efficiency. In the process of optimizing the algorithm, on the one hand, it is necessary to conduct an in-depth analysis of the properties of the problem to find out the essence of the problem and, on the other hand, it can start to improve from the shortcomings of the original algorithm. Designing algorithms using the staged dynamic time deformation idea requires strong creativity. The staged dynamic time deformation algorithm is an idea, and there is no unified standard state model for all problems. The phased dynamic time deformation state model of the actual problem may be completely different, so a specific analysis of the specific problem is required. Similarly, the optimization of the algorithm is also very flexible. What this article discusses is only some conventional optimization measures, and efficient optimization methods need to continue to dig in specific problems. The index space constructed by the method in this paper does not depend on the size of the human motion database and is easy to modify. It can be matched with varying degrees of accuracy by flexibly specifying feature matching during retrieval, so it has good

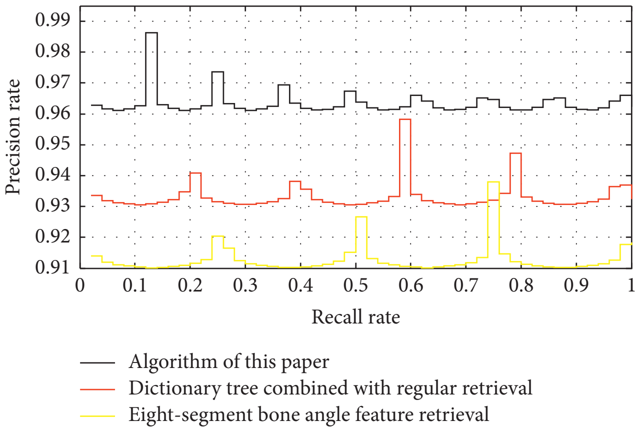

Figure 11: P-R comparison of search results of multiple methods on several different sports categories.

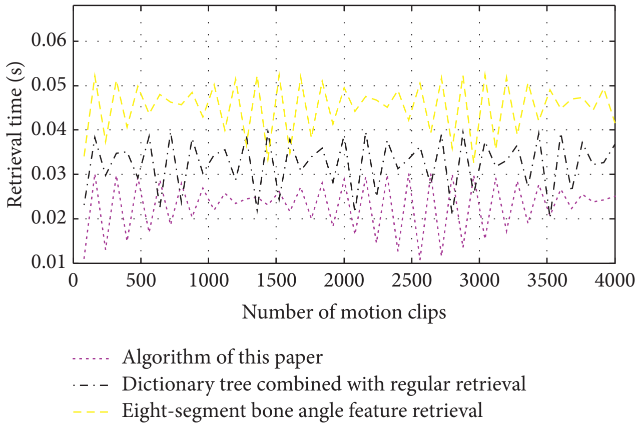

Figure 12: Retrieval time of various methods under different body motion database sizes.

retrieval performance and effects. Since the designation of parameters that need to be matched during retrieval will affect the retrieval results, the user needs to have certain prior knowledge. The next goal is to adaptively match the prominent motion characteristics of different categories of motion to achieve the purpose of intelligent retrieval.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Voulodimos, I. Rallis, and N. Doulamis, "Physics-based keyframe selection for human motion summarization," *Multimedia Tools and Applications*, vol. 79, no. 5-6, pp. 3243–3259, 2020.

[2] S. Ding, S. Qu, Y. Xi, and S. Wan, "A long video caption generation algorithm for big video data retrieval," *Future Generation Computer Systems*, vol. 93, pp. 583–595, 2019.

[3] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: a survey," *The International Journal of Robotics Research*, vol. 39, no. 8, pp. 895–935, 2020.

[4] Y. Chang, "Research on de-motion blur image processing based on deep learning," *Journal of Visual Communication and Image Representation*, vol. 60, pp. 371–379, 2019.

[5] S. Li, Y. Zhou, H. Zhu, W. Xie, Y. Zhao, and X. Liu, "Bidirectional recurrent autoencoder for 3D skeleton motion data refinement," *Computers & Graphics*, vol. 81, pp. 92–103, 2019.

[6] Y.-T. Liu, Y.-A. Zhang, and M. Zeng, "Sensor to segment calibration for magnetic and inertial sensor based motion capture systems," *Measurement*, vol. 142, pp. 1–9, 2019.

[7] W. Mrabti, K. Baibai, B. Bellach, R. O. Haj Thami, and H. Tairi, "Human motion tracking: a comparative study," *Procedia Computer Science*, vol. 148, pp. 145–153, 2019.

[8] I. Ajili, M. Mallem, and J.-Y. Didier, "Human motions and emotions recognition inspired by LMA qualities," *The Visual Computer*, vol. 35, no. 10, pp. 1411–1426, 2019.

[9] J. Sedmidubsky, P. Elias, and P. Zezula, "Searching for variable-speed motions in long sequences of motion capture data," *Information Systems*, vol. 80, pp. 148–158, 2019.

[10] J. Qu, F. Zhang, Y. Wang, and Y. Fu, "Human-like coordination motion learning for a redundant dual-arm robot," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 379–390, 2019.

[11] L. Xia, J. Lv, and D. Liu, "A motion classification model with improved robustness through deformation code integration," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8519–8532, 2019.

[12] M. A. Khan, T. Akram, M. Sharif et al., "Improved strategy for human action recognition; experiencing a cascaded design," *IET Image Processing*, vol. 14, no. 5, pp. 818–829, 2020.

[13] F. Bailly, J. Carpentier, M. Benallegue, B. Watier, and P. Soueres, "Estimating the center of mass and the angular momentum derivative for legged locomotion-A recursive approach," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4155–4162, 2019.

[14] J. S. Park, C. Park, and D. Manocha, "I-Planner: intention-aware motion planning using learning-based human motion prediction," *The International Journal of Robotics Research*, vol. 38, no. 1, pp. 23–39, 2019.

[15] B. Erol and M. G. Amin, "Radar data cube processing for human activity recognition using multisubspace learning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3617–3628, 2019.

[16] C. Veinidis, I. Pratikakis, and T. Theoharis, "Unsupervised human action retrieval using salient points in 3D mesh sequences," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2789–2814, 2019.

[17] S. Arivazhagan, R. N. Shebiah, R. Harini, and S. Swetha, "Human action recognition from RGB-D data using complete local binary pattern," *Cognitive Systems Research*, vol. 58, pp. 94–104, 2019.

[18] M. E. Montchal, Z. M. Reagh, and M. A. Yassa, "Precise temporal memories are supported by the lateral entorhinal cortex in humans," *Nature Neuroscience*, vol. 22, no. 2, pp. 284–288, 2019.

[19] N. Jaouedi, N. Boujnah, and M. S. Bouhlel, "A new hybrid deep learning model for human action recognition," *Journal of King Saud University—Computer and Information Sciences*, vol. 32, no. 4, pp. 447–453, 2020.

[20] N. Nikolakis, K. Alexopoulos, E. Xanthakis, and G. Chryssolouris, "The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 1, pp. 1–12, 2019.

[21] M. Guo and Z. Wang, "Segmentation and recognition of human motion sequences using wearable inertial sensors," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 21201–21220, 2018.

[22] J. Sedmidubsky, P. Elias, and P. Zezula, "Effective and efficient similarity searching in motion capture data," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12073–12094, 2018.

[23] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2018.

[24] M. Ramezani and F. Yaghmaee, "Motion pattern based representation for improving human action retrieval," *Multimedia Tools and Applications*, vol. 77, no. 19, pp. 26009–26032, 2018.

[25] Y. Feng, P. Zhou, J. Xu et al., "Video big data retrieval over media cloud: a context-aware online learning approach," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1762–1777, 2019.

[26] H. R. Dimsdale-Zucker, M. Ritchey, A. D. Ekstrom et al., "CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields," *Nature Communications*, vol. 9, no. 1, pp. 1–8, 2018.

[27] U. Iqbal, A. Doering, H. Yasin, B. Krüger, A. Weber, and J. Gall, "A dual-source approach for 3D human pose estimation from single images," *Computer Vision and Image Understanding*, vol. 172, pp. 37–49, 2018.

[28] F. Patrona, A. Chatzitofis, D. Zarpalas, and P. Daras, "Motion analysis: action detection, recognition and evaluation based on motion capture data," *Pattern Recognition*, vol. 76, pp. 612–622, 2018.