

Motion Key frame Extraction Based on Grey Wolf Optimization Algorithm

Yi Zhang, Juan Li^a and Min Zhang

College of Information Engineering, Dalian University, Dalian Liaoning116622, China

Abstract: In order to extract the key frames more effectively, we propose a key frame extraction method for human motion sequences based on Grey Wolf Optimization (GWO) algorithm. The fitness function is defined with the minimum reconstruction error and the optimal compression rate. The social hierarchy of grey wolves and hunting strategy are simulated to search key frames. Experimental results show that the proposed method can not only maintain the consistency of key frames between similar human motion sequences, but also effectively compress and summarize the original motion data. Under the same compression ratio, the reconstruction error is the minimum.

1 Introduction

With the maturity of motion capture technology and the rapid development of virtual reality technology, motion capture data, as an emerging multimedia data, is used in many fields such as movies, animation and entertainment games. However, motion capture data based on high-frequency sampling usually has a large amount of data redundancy. Therefore, it is particularly important to select the data that best represents the characteristics of this data from the complicated data. Key frame extraction technology refers to selecting some of the most representative motion frames from a complex human motion sequence as key frames, while other non-key frames can be reconstructed with interpolation on key frames[1]. The application of key frame technology Reduces the load of data storage and transmission, and users can quickly understand the contents of the motion database by browsing key frames.

In recent years, key frame extraction technology, as a research hotspot, has attracted the attention of scholars in all sectors. Gianluigi[2] proposed a method for selecting key frames of a video digest sequence by analyzing the difference between two consecutive frames. Zhang[3] proposed a key frame extraction algorithm based on clustering, which classified similar motion data into one class, and selected one frame in each class as the key frame. This algorithm had certain generalization ability to the original motion sequences, but ignored the timing between the motion sequences. Lim[4] regarded motion data as a trajectory curve in high-dimensional space, and proposed a single-layer curve simplification algorithm. Zhu[5] derived the explicit mapping relationship between high-dimensional motion data and low-dimensional state variables with the linear time-invariant system, divided the motion sequences into a series of cascaded segments and regarded the motion pose at these segmentation points as

the key frame. Yang[6] proposed a key frame extraction method for motion capture sequence based on quantum behavior particle swarm optimization algorithm, which could determine the number of key frames. Li[7] exploited the distance between quaternions to represent the difference between human body posture, and continuously calculated the difference between the current frame and the last frame, then regarded the frame with the difference exceeding than a threshold as the key frame. Lee[8] defined the objective function with the compression ratio and key frame and proposed an algorithm that using genetic algorithm to extract key frames from dynamic mesh sequences. Halit[9] processed the input motion sequence as a motion curve and selected the curve to be analyzed by PCA dimension reduction technique, then extracted the most important frame as the key frame of motion using frame reduction technology. Liu[10] used the frame decimation method to calculate the reconstruction error, and extracted the key frame with the optimal compression rate. However, under the optimal compression rate, the reconstruction error is not the minimum.

In this paper, the fitness function is defined with the optimal compression ratio and the minimum reconstruction error. A key frame extraction method based on grey wolf optimization (GWO) algorithm is proposed, which simulates the social hierarchy of grey wolves and hunting strategies to extract key frames. The GWO algorithm is simple, easy to implement, less parameters and fast convergence. So it is superior to other intelligent optimization algorithms such as particle swarm optimization algorithm and genetic algorithm in function optimization problems. The experiment conducted also shows that the key frames extracted using the proposed method can not only summarize the original motion sequence well, but also obtains the optimal compression rate and the minimum reconstruction error simultaneously.

^aCorresponding author: JL0071993@163.com

2 Grey wolf optimization algorithm

The GWO is a novel intelligent optimization algorithm proposed by Seyedali[11]. It is a meta-heuristic algorithm developed by simulating the behavioral mechanism of grey wolves hunting prey. In the GWO algorithm, the grey wolf with the strongest leadership is marked as α wolf (called the king of wolves), which is the decision-maker in the whole hunting activity. Its fitness is the best, and it is also closest to the prey. The grey wolves with the second and the third best fitness are marked as β wolf and δ wolf respectively, they assist α wolf in management and making decisions while hunting, and they are also candidates for α wolf, the remaining grey wolves are defined as ω wolves, assisting the first three grey wolves to hunt[12]. In the GWO algorithm, the hunting activities of the grey wolves can be divided into three stages: encircling, pursuing and attacking. The specific descriptions are as follows:

1)**Encircling**:During the period of encircling, the wolves first search for the prey, and the distance between the prey and the grey wolves is expressed as:

$$D = |C \cdot W_p(t) - W(t)| \quad (1)$$

$$W(t+1) = W_p(t) - A \cdot D \quad (2)$$

Where, t represents the number of iterations; W_p is the position of the prey (optimal solution); W indicates the position of a wolf; the coefficient constant A and C are expressed as follows:

$$A = 2a \cdot r_1 - a \quad (3)$$

$$C = 2r_2 \quad (4)$$

When the wolves approach the prey, a decreases linearly from 2 to 0; r_1 and r_2 are random number which belong $[0,1]$.

2)**Pursuing**:In this stage, the position of the individual wolf must be updated as the movement of the prey, so the prey can be re-determined according to the updated positions of the α wolf, β wolf and δ wolf. The updating mechanism is as follows:

$$\begin{cases} D_\alpha = |C_1 \cdot W_\alpha - W| \\ D_\beta = |C_2 \cdot W_\beta - W| \\ D_\delta = |C_3 \cdot W_\delta - W| \end{cases} \quad (5)$$

$$\begin{cases} W_1 = W_\alpha - A_1 \cdot D_\alpha \\ W_2 = W_\beta - A_2 \cdot D_\beta \\ W_3 = W_\delta - A_3 \cdot D_\delta \end{cases} \quad (6)$$

$$W_p(t+1) = \frac{W_1 + W_2 + W_3}{3} \quad (7)$$

Among the formula(5-7), D_α , D_β and D_δ respectively indicate the distances between α wolf, β wolf, δ wolf and ω wolves.

3)**Attacking**:After the stage of pursuing, wolves find the prey and attack it, which is realized by the decreasing of a in the formula (3). A changes as the changing of a ,

and $|A| \leq 1$ indicates that the positions of the wolves after updating are closer to the prey. On the contrary, $1 < |A| < 2$ indicates the wolves are moving away from the prey.

3 key frame extraction

3.1 Coding and initialization

In this paper, the wolves are coded with binary code. Each wolf is a set of candidate key frames. The key frame is encoded as 1 and the non-key frame is encoded as 0, the length of code is equal to the number of frames. A wolf W_i represents a sequence of movements:

$$W_i = [w_{i1}, w_{i2}, \dots, w_{iN}], i = 1, 2, \dots, K \quad (8)$$

In the formula(8), N is the total number of frames; $w_{ij} = 1$ indicates the frame is a key frame, while $w_{ij} = 0$ means that the frame is a non-key frame.

The random binary code is exploited as the initial encoding of the motion sequence, and we regard the first frame of the original motion sequence of the human body as the key frame of the motion sequence, so does the last frame.

3.2 Definition of the fitness function

In the process of finding the key frame of human motion sequence, we expect the number of key frames is the minimum which means the compression ratio is optimal, the reconstruction error is also minimum at the same time. But the compression ratio and reconstruction error have mutual contradiction in nature. So we must use a weight to make a compromise.

$$f(\text{keyframes}) = lE(\text{keyframes}) + (1-l)R(\text{keyframes}) \quad (9)$$

Where, the weight $l \in [0,1]$; $R(\text{keyframes})$ represents the compression ratio, that is the ratio of the number of key frames to the number of total motion frames, which is expressed as follows:

$$R(\text{keyframes}) = \frac{\text{number}(\text{keyframes})}{\text{totalnumber}} \quad (10)$$

$\text{number}(\text{keyframes})$ is the number of the key frames and totalnumber is the total number of frames of human motion data.

Since the dimensionality between the reconstruction error and the compression ratio are inconsistent, therefore, the reconstruction error needs to be standardized as formula (11):

$$E(\text{keyframes}) = \frac{E_{\text{key}}}{E_{\text{max}}} \quad (11)$$

E_{key} represents the reconstruction error of the key frame extracted from the human motion sequence and E_{max} is the maximum reconstruction error. In general,we think the reconstruction error between the first frame and the last frame is the maximum.

3.3 Calculation of the reconstruction error

The linear interpolation method is used to interpolate and reconstruct the extracted key frame sets, and the uniform inter-frame distance between the original human motion sequence and the reconstructed motion sequence is used as the reconstruction error expressed as formula(12):

$$E = [\sum (F(n) - F'(n))^2] / N \quad (12)$$

$F(n)$ is the original human motion data and $F'(n)$ is the human motion data reconstructed with linear interpolation, N is the total number of frames of the human motion data.

3.4 Description of the key frame extraction algorithm based on grey wolf optimization

Based on the introductions about the key frame extraction above, the steps of the key frame extraction algorithm based on GWO proposed in this paper are summarized as follows:

Step1: Initializing and encoding. the number of wolves is K , the total number of frames from the bvh file is N . set the number of iterations M , the group of the grey wolves $W = [W_1, W_2, \dots, W_K]^T$ and the sequence represented by each grey wolf $W_i = [w_{i1}, w_{i2}, \dots, w_{iN}]$.

Step2: The fitness f_i of each grey wolf is calculated according to the formula (9-12) and the weight value is set as $l = 0.3$. the grey wolves ranked in the top 3 of the fitness value are chosen and marked W_α , W_β , W_δ respectively. The wolf marked W_α is the best candidate key frames set and it's fitness is the maximum among the fitness of all wolves.

Step3: The distances between the wolves α, β, δ and the remaining individuals ω are calculated according to equation (5); and then the positions of the wolves and prey (key frames) are updated according to the updating mechanism of formula(6-7).

Step4: The parameters A , a and C are updated according to formula(3-4).

Step 5: The algorithm ends up with the number of iterations t reaching the maximum M , otherwise, the steps(2-5) are repeated. Finally we choose the set W_α obtained by the last iteration as the key frame set and extract the key frames according to the position index of the "1" in W_α .

4 Experiment and analysis

In order to verify the validity of the method proposed above, we test a series of motion capture data which comes from the Carnegie Mellon University (CMU).

4.1 Key frame extraction of similar motions

We choose four walking motions with different styles, which all have the same size (130 frames). The extracted results of key frames are shown in Figure 1(a-d).

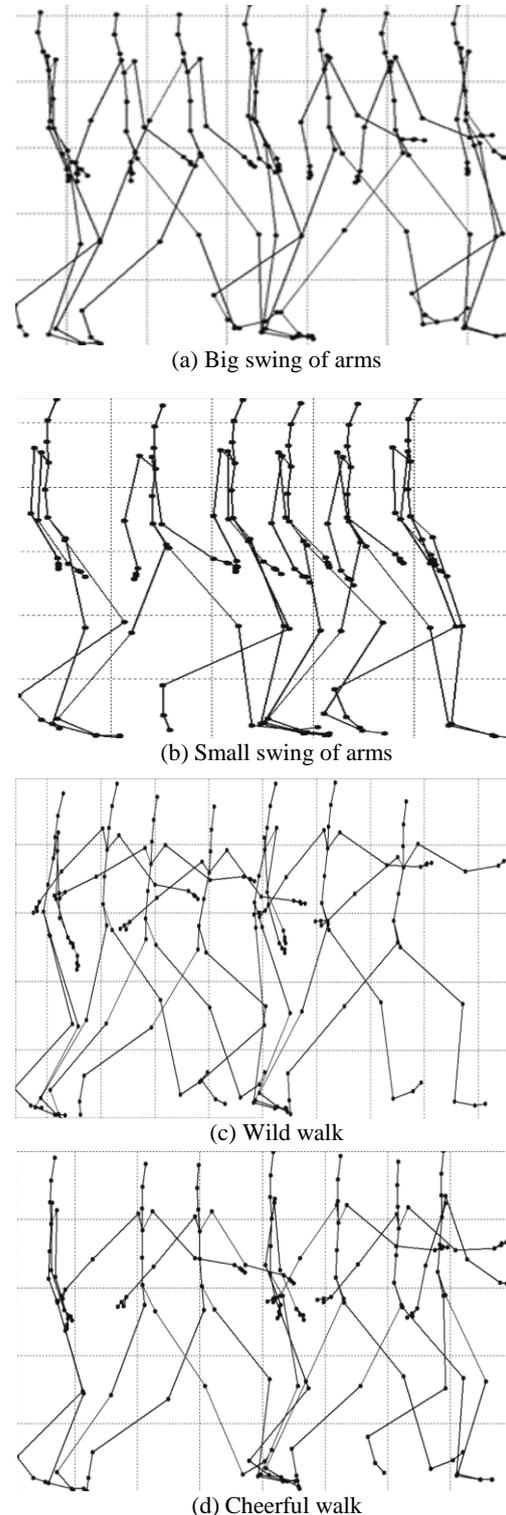


Figure1.Key frames of four different walking styles

From the figures above, we can easily distinguish the Big swing of walks from the Small swing of walks through the swing of arms, and the back is straight in Cheerful walk while it is not in the Wild walk. It can be seen that GWO algorithm maintains the consistency of key frames between similar motion sequences.

4.2 Comparison on the ability of different algorithms to detect key frames

As for the motion “kicking the ball”, an aperiodic motion sequences, we use the proposed method to extract the key frames and make a comparison with Particle Swarm Optimization(PSO) algorithm and Genetic algorithm (GA) respectively. The results are shown as Figure 2(a-c).

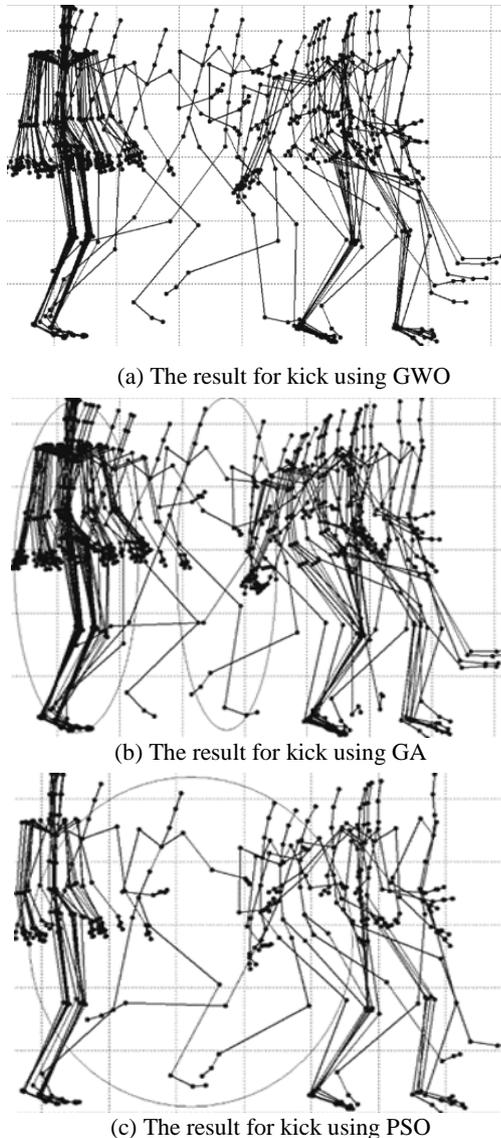


Figure2.The results of different algorithms extracting key frames

We can find that the key frames extracted using GWO can well summarize the original “kicking the ball” motion, while PSO lose a lot of drastically-changing frames in the middle part(as in ellipse), namely under-sampling. GA is also under-sampling in this part and it has much redundancy in the slowly-changing frames(over-sampling) at the same time. So on the whole, GWO possesses of the best summarizing ability.

4.3 Compression ratio and reconstruction error

The number of key frames extracted using GWO and the compression rate for five different motion behaviours are as Table1.

Table1.The compression rate of different motions

Category	Total frames	Key Frames	Compression rate(%)
Walk	343	18	5.2
Jump	439	14	3.1
Run	801	16	1.9
Dance	1033	22	2.1
Walk-jump-walk	1199	29	2.4

Under same compression ratio, we still adopt five different motions as Table1 and calculate the absolute reconstruction errors about the GWO method, Curve Saliency, Quaternion and Distance Curve methods respectively. The results are as shown in Table2

Table2.The reconstruction errors of different methods

category	Curve Saliency	Quaternion	Distance Curve	GWO
Walk	485	461	186	179
Jump	781	859	583	307
Run	3723	670	1821	486
Dance	1720	4089	1773	781
Walk-jump-walk	1788	3573	1506	1089

We use four methods to get the same key-frames from five motions and reconstruct the motion sequences by the linear interpolation method.It can be seen from Table2 that under the same compression rate,the reconstruction error of GWO is the minimum.

Conclusion

In this paper, an extraction method based on grey wolf optimization algorithm for human motion capture data is proposed, and the fitness function with the minimum reconstruction error and the optimal compression rate is defined. The binary code is used to encode the motion sequence. The key frames of the motion sequence are obtained by simulating the social rank of the grey wolves, the division of labour and the cooperation during hunting the prey. Compared with other traditional algorithms, the reconstruction error of the GWO method is minimum under the same compression rate and the compression ratio is optimal (all less than 6%). It is well visually summarized and maintains the consistency of key frames between similar motion sequences.

References

1. G Xia, H Sun, X Niu, et al. Keyframe Extraction for Human Motion Capture Data Based on Joint Kernel Sparse Representation. *IEEE T IND ELECTRON*, **64**:1589-1599(2017).
2. C Gianluigi, S Raimondo. An Innovative Algorithm for Key Frame Extraction in Video Summarization. *J REAL-TIME IMAGE PR*, **1**:69-88(2006).
3. Q Z, S.P Y, D.S Zhou, et al. An Efficient Method of Key-frame Extraction Based on a Cluster Algorithm. *J HUM KINET*, **39**:5-14(2013).
4. I.S Lim, D Thalmann. Key-posture extraction out of human motion data. *Proceedings of the 23rd Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society(IEEE). 1167-1169(2001).

5. D.M Zhu, Z.Q Wang. Extraction of Keyframe from Motion Capture Data Based on Motion Sequence Segmentation. *Journal of Computer-Aided Design & Computer Graphics*, **20**:787-792(2008).
6. T Yang, H.J Sun, J Y. Extraction of keyframe from motion capture data based on quantum-behaved particle swarm optimization. *Application Research of Computers*, **31**:2523-2527(2014).
7. S.Y Li, J Hou, L.Y Gan. Extraction of Motion Keyframe Based on Inter-frame Pitch. *Computer Engineering*, **41**:242-247(2015).
8. T.Y Lee, C.H Lin, Y.S Wang, et al. Animation Key-Frame Extraction and Simplification Using Deformation Analysis. *IEEE T CIRC SYST VID*, **18**:478-486(2008).
9. C Halit, T Capin. Multiscale motion saliency for keyframe extraction from motion capture sequences. *COMPUT ANIMAT VIRT W*, **22**:3-14(2011).
10. Y.G Liu, J.G Liu. Keyframe Extraction from Motion Capture Data by Optimal Reconstruction Error. *Journal of Computer-Aided Design & Computer Graphics*, **22**:670-675(2010).
11. S Mirjalili, S.M Mirjalili, A Lewis. Grey Wolf Optimizer. *ADV ENG SOFTW*, **69**:46-61(2014).
12. E Emary, H.M Zawbaa, A.E Hassanien. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, **172**:371-381(2016).