Key Frames Extraction of Human Motion Capture Data Based on Cosine Similarity

Yang Li
Dalian University

Dongsheng Zhou*
Dalian University
zhoudongsheng@dlu.edu.cn

Xiaopeng Wei
Dalian University of Technology

Qiang Zhang*
Dalian University
zhangq@dlu.edu.cn

Xin Yang
Dalian University of Technology

Abstract

In this paper, we propose the key frame extraction method based on cosine similarity algorithm. The method removed the joints that have less influence on the human posture. Then we regard the human motion capture data as a vector in the Euclidean space, which is used to represent the motion frame. We extract the key frames by calculating the cosine similarity between vectors. Experimental results show that our method has high compression rate and low reconstruction error in the extraction of key frames, and has great practical value.

Keywords: motion capture, key frame, cosine similarity, reconstruction error

1. Introduction

With the rise of motion capture technology [1], the traditional animation production methods are gradually replaced by the character animation based on motion capture technology. However, human motion capture data has the shortcomings of high dimension, data redundancy, difficult to deal with. We can effectively reduce the data redundancy of video content by using the key frame extraction technology. The extracted key frames can improve the retrieval efficiency of the motion data. In order to make key frame extraction more efficient, many researchers have done a lot of work in this field.

Lim et al [2] improved the curve simplification method, simplified the motion data into the curve in space, and extracted the position sequence of extreme points on the curve as the key frame. Halitet al [3] uses the PCA reduced dimension method to get the characteristic curve of motion data, and then use Gaussian filter to get the motion sequence. Finally, the key frames are obtained by clustering method. Liu et al [4] proposed a clustering method in which the motion data are clustered into K sets according to the feature, and then the first frame in each set is regarded as the key frame. Sun [5] proposed a key frame extraction algorithm based on improved K-means algorithm, clustering the extracted feature vectors according to the similarity of artificial fish, combined with K-means algorithm to calculate the clustering results, and extracted the key frames. Yang et al [6] proposed a key frame extraction method based on quantum particle swarm optimization algorithm, Particle swarm optimization algorithm has faster search speed, and uses sequential integer coding to ensure the sequence of motion sequences. This method can extract the key frames of the motion data effectively.

The above-mentioned methods can effectively extract the key frames, but they all have their own deficiencies. Curve simplification method is easy to lose motion information. The

*Corresponding authors.

clustering method does not take into account the time series consistency, which will lead to the chaos of the motion sequences. Intelligent algorithm is complex, time-consuming and unstable. This paper presents a key frame extraction algorithm based on cosine similarity.

2. Key frame extraction based on cosine similarity

2.1 Calculate the cosine similarity

As a similarity measurement method, cosine similarity has a good effect of vector similarity measurement. Cosine similarity is widely used in text similarity detection [7], image recognition, machine learning and other fields. Cosine similarity uses the cosine value of two vectors to represent the similarity of two individuals. The cosine value is closer to 1, the more similar of the two individuals.

In n-dimensional space, the cosine value of two vectors $a(x_{11}, x_{12}, \ldots, x_{1n})$ and $b(x_{21}, x_{22}, \ldots, x_{2n})$ is calculated as follows:

$$\cos(\theta) = \frac{\sum_{k=1}^{n} x_{1k} x_{2k}}{\sqrt{\sum_{k=1}^{n} x_{1k}^2} \sqrt{\sum_{k=1}^{n} x_{2k}^2}}$$

Human motion capture data is 96 dimensional data. We regard human motion capture data as vectors in Euclidean space. There are no two motion vectors in the same direction for a continuous motion sequence. Therefore, we can use the cosine similarity to calculate the difference of the two vectors in the direction, when the difference between motion vectors is large, it can be taken as a key frame. We choose four frames from the jumping motion to calculate the cosine similarity. The selected frames are shown in Figure 1.

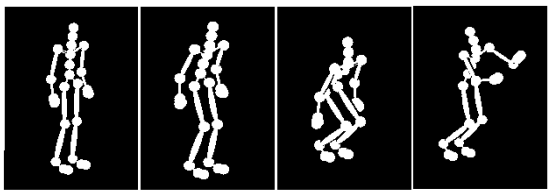

Figure 1: (a) Standing (b) Bending (c)Squatting (d) Jumping

The cosine similarity between the four motion frames is shown in Table 1.

Table 1: cosine similarity between the four frames

Frame	(a)	(b)	(c)	(d)
(a)	1.0000	0.9470	0.6321	0.5692
(b)	0.9470	1.0000	0.8210	0.7089
(c)	0.6321	0.8210	1.0000	0.7742
(d)	0.5692	0.7089	0.7742	1.0000

As can be seen from Figure 1 and Table 1, the cosine similarity between the four motions shown in Table 1 is consistent with the difference between the four motions shown in Figure 1. It can be known that the cosine similarity can be used to measure the similarity of motion frames.

2.2 Extract the key frames

In the human skeleton model, there are some joint points which have little influence on human body posture, so we can remove these joint points to improve the accuracy of the calculation. In this paper, we remove the root joint points and eight other joint points (RightFingerBase/RightFinger/RightThumb/LeftFingerBase/LeftFinger/LeftThumb/RightToeBase/LeftToeBase). The original motion data is reduced from 96 dimensions to 69 dimensions.

In the experiment, we use the cosine similarity to represent the difference of human pose, the range of cosine similarity is [-1,1]. The larger the cosine value is, the more similar the two motion pose is. Set the similarity threshold, which is typically set between 0.9 and 1. The higher the compression ratio requirement, the smaller the threshold setting. First, take the first frame of the motion sequence as the first key frame. Calculate the cosine similarity between the key frame and subsequent motion frames. The frame which has large difference with the key frame is extracted as a new key frame. Then the newly extracted key frame is compared to the frames following it, Loop until the end. Specific steps are as follows:

Step1: Read the original motion capture data, remove the minor joint points, and reduce the dimension of data.

Step2: Take the first frame as the key frame, initialize the loop variable $j = 2$.

Step3: Sets the loop variable $j \le n$, where n is the total number of frames of the motion sequence. When $j \le n$, calculate the cosine similarity between the key frame and the frame represented by the number j. After the

dimension reduction, the motion data of each frame is represented by a 69 dimensional vector, the cosine similarity is calculated as follows:

$$m = \frac{x_1 y_1 + x_2 y_2 + \cdots + x_{69} y_{69}}{\sqrt{x_1^2 + x_2^2 + \cdots + x_{69}^2} \cdot \sqrt{y_1^2 + y_2^2 + \cdots + y_{69}^2}}$$

Step4: Set a threshold α , $\alpha \in (-1,1)$. When $m < \alpha$, that shows the large difference between the two motion frames. Extract the frame represented by the number j as a new key frame, save in key[j].

Step5：Set j=j+1, return step2, Loop until the end. All the key frames are saved in key[j].

3. Experiments and analysis

In the experiment, we select walking, running, jumping, kicking four kinds of sports, these sports are commonly seen in our daily life. The experimental data is from the Carnegie Mellon University (CMU) human motion capture database. After a number of tests, the extracted key frames have both low reconstruction error and high compression rate when the similarity threshold is about 0.95. In the following key frame extraction experiments, the similarity threshold is set to 0.95.

Experiment 1: The key frame extraction of walking, running and kicking sports.

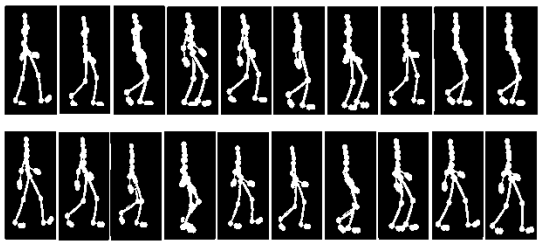

Figure 2: The key frames of walking

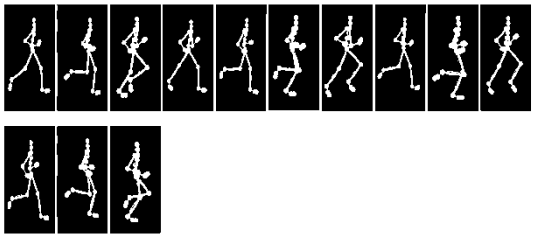

Figure 3: The key frames of running

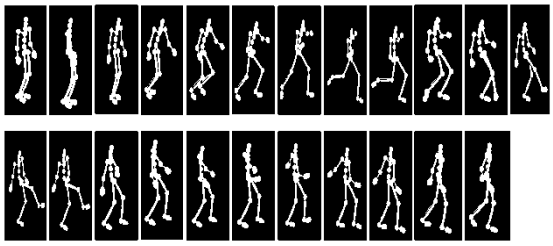

Figure 4: The key frames of playing football

The experimental results show that the key frames extracted by our method can well reflect the original motion sequence.

Experiment 2: We use the method proposed in this paper and t-SNE dimension reduction method [8] to extract the key frames of jumping movement (439 frames).

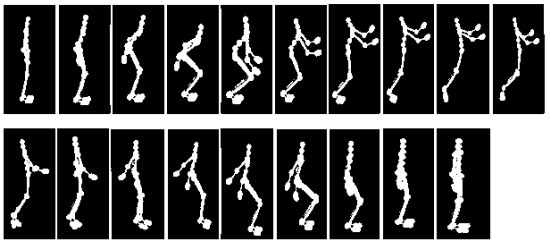

Figure 5: The method used in this paper (19 frames)

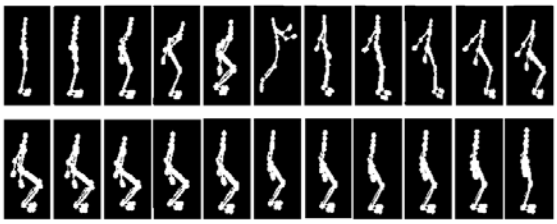

Figure 6: t-SNE dimension reduction method (22 frames)

As can be seen from Figure 5, our method extracts the key frames sampled uniformly and has a good generalization of the original motion sequence. The t-SNE method had the problems of over sampling at the end of the movement and under sampling when strenuous movement. In comparison, the method proposed in this paper is better.

Experiment 3: The method proposed in this paper is compared with t-SNE method.

This paper adopts the method of linear interpolation to reconstruct the motion sequence. The reconstruction error between the reconstructed sequence and the original sequence is calculated as follows:

$$E = \frac{1}{n}\sum_{i=1}^{n}(M_1(i) - M_2(i))^2$$

Where $M_1(i)$ is the original motion data, $M_2(i)$ is the reconstructed motion data, n is the total number of the motion frames.

The compression ratio and the reconstruction error of the extracted key frames are shown in the following tables:

Table 2: Compression ratio and reconstruction error using our method

Motion type	Total frames	key frames	Compression ratio (%)	Reconstruction error
Walk	343	20	5.83	5.16e+03
run	165	13	7.87	3.35e+03
Jump	439	19	4.32	6.28 e+03
Kick	801	23	2.87	1.65 e+04

Table 3: Compression ratio and reconstruction error using t-SNE method

Motion type	Total frames	key frames	Compression ratio (%)	Reconstruction error
Walk	343	29	8.43	6.86e+03
run	165	14	8.48	3.97 e+03
Jump	439	22	5.01	8.02 e+03
Kick	801	29	3.62	2.71 e+04

Seen from Table 2 and Table 3, for several common sports, our method extracts fewer key frames, which means higher compression efficiency. At the same time, the reconstruction error obtained by our method is lower than that of t-SNE method. Therefore, the proposed method can get better results compared to t-SNE method.

4. Conclusion

In this paper, we propose using cosine similarity algorithm to extract key frames. The experiment result shows that our method has a good effect in the key frame extraction. And the proposed method only need to set a similarity threshold, the key frame extraction is simple and reliable. How to improve the application of cosine similarity in human capture data, reduce the similarity error and make it have good adaptability, which is the direction of our research and improvement.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos.61370141, 61300015), the Program for Dalian High-level Talent's Innovation(2015R088), Program for Changjiang Scholars and Innovative Research Team in University (No.IRT_15R07), and by the Program for Liaoning Innovative Research Team in University (No.LT2015002).

References

[1] Dong, Y., & Desouza, G. N. A new hierarchical particle filtering for markerless human motion capture. Computational Intelligence for Visual Intelligence, CIVI '09. IEEE Workshop on, 2009:14-21

[2] Lim I S, Thalmann D. Key-posture extraction out of human motion data. In Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, 2001, 2:1167-1169

[3] Halit C, Capin T. Multiscalemotion saliency for keyframe extraction from motion capture sequences. Computer Animation and Virtual Worlds, 2011, 22(1): 3-14

[4] Liu F, Zhuang Y, Huang T S, et al, 3D motion retrieval with motion index tree. Computer Vision and Image Understanding, 2003, 92(2): 265-284

[5] Sun S, Zhang J, Liu H. Key frame extraction based on Artificial Fish Swarm Algorithm and k-means. International Conference on Transportation, Mechanical, and Electrical Engineering. IEEE, 2011: 1650-1653

[6] Yang T, Sun H J, Jun Y E. Extraction of keyframe from motion capture data based on quantum-behaved particle swarm optimization. Application Research of Computers, 2014, 2(205): 526-530

[7] Oskina K. Text Classification in the Domain of Applied Linguistics as Part of a Pre-editing Module for Machine Translation Systems. Speech and Computer. Springer International Publishing, 2016: 691-698

[8] Zhang Q, Yao Y, Zhou D, et al. Motion Key-Frame Extraction by Using Optimized t-Stochastic Neighbor Embedding. Symmetry, 2015, 7(2): 395-41