

## Motion Key-frames extraction based on amplitude of distance characteristic curve

Qiang Zhang\*

*Key Laboratory of Advanced Design and Intelligent Computing of Ministry of Education,  
Dalian University, Dalian 116622, China*

Xiang Xue, Dongsheng Zhou, Xiaopeng Wei\*\*

*Key Laboratory of Advanced Design and Intelligent Computing of Ministry of Education,  
Dalian University, Dalian 116622, China  
E-mail: zhangq30@yahoo.com  
www.dlu.edu.cn*

Received 14 July 2012

Accepted 15 December 2012

### Abstract

The key frames extraction technique extracts key postures to describe the original motion sequence, which has been widely used in motion compression, motion retrieval, motion edition and so on. In this paper, we propose a method based on the amplitude of curve to find key frames in a motion captured sequence. First we select a group of joint distance features to represent the motion and adopt the Principal Component Analysis (PCA) method to obtain the one dimension principal component as a features curve which will be used. Then we gain the initial key-frames by extracting the local optimum points in the curve. At last, we get the final key frames by inserting frames based on the amplitude of the curve and merging key frames too close. A number of experimental examples demonstrate that our method is practicable and efficient not only in the visual performance but also in the aspect of the compression ratio and error rate.

*Keywords:* key-frames extraction, distance features, PCA, insert and merge, amplitude.

### 1. Introduction

In the last few decades, the technique of human motion capture develops rapidly and the significance of motion capture for several applications is also rising. Movies and games have started to widely use motion capture systems. At the same time, there exists a problem that storage of the motion capture database is enlarged due to their tremendous size. The need for more compact representation has led domestic and foreign scholars to investigate the ways of handling these huge data, which is achieved by Key framing.

The key-frames are selecting the most meaningful and vital frames of motion and others are not important

which are computed via the interpolation techniques by using the key frames. So the key-frames extraction technique plays an irreplaceable role in representation of the whole animation sequence and is very beneficial for storage, compression, retrieval, browse and reuse for human Mocap data.

Up to now, there have been a considerable number of approaches proposed for extraction of key frames. They essentially differ from each other in the way that they treat motion sequences. We compare those solutions in the related work section.

This paper focuses on key-frames extraction method for human motion capture data. Key-frames collection should meet the following requirements: On the one hand, key-frames set should be able to generalize the

---

\*Corresponding authors: {zhangq,xpwei}@dlu.edu.cn

original motion efficiently in certain compression ratio. On the other hand, key-frames collection can be utilized for reconstructing the original motion sequence as precisely as possible. There are two evaluation criteria for key-frames extraction: error requirement and compression ratio rather than just relying on the subjective visual judgment.

In this paper, we choose a new distance characteristic curve to reflect motion essential features. According to the magnitude of the characteristic curve, we carry out twice keyframe extraction. Finally, we automatically extract a certain number of key frames.

## **2. Related work**

There have been various proposed methods to extract key frame in literature [1]. The simplest idea of key frame extraction is uniform sampling, which extracts key frames with uniform sampling intervals. This method is simple and costs less time, but it cannot summarize the original human motion effectively, because it suffered from over-sampling in segments with less pose changes(resulting in data redundancy) or under-sampling in segments with great pose changes(resulting in loss of details). Besides that, many researchers also come forward methods based on adaptive sampling, which according to the pose changes rather than motion time, this method extracts less key frames in motion segments with less pose changes and more key frames in motion segments with great pose changes, so this method can tackle over-sampling and under-sampling problems, and we fall this method into two mainly categories: curve simplification and clustering.

Curve simplification techniques are usually used in key frame extraction field. One frame of motion sequence could be considered as a point in high-dimensional space. These points compose a trajectory curve as the consecutiveness of motion sequence. Lim et al.[2] extended a curve simplification algorithm which is typically used for planar curves in a high-dimensional space and then applied a simple curve simplification algorithm to extract key-frames. Ogawa et al.[3] proposed a position-based key frame extraction scheme. Treating animation as a set of curves which give the value of the rotation of all joints in each frame, frames are decimated one by one with measuring the positions of all the joints to detect a number of position-based key frames. At the same time, Li et al.[4]

proposed an automatically key frames detection method by using frame decimation. The frames are ranked by their importance and thus users can specify any number of key frames from one data set. Assa et al.[5] used a multi-dimensional scaling method for high dimensional motion data is mapped to low-dimensional space, then used the curve simplification method to extract key frames. Same curve simplification is also used in video sequences for key frame extraction[6]. Xiao et al.[7-8] improved the curve simplification algorithm and introduced a parameter adaptive adjustment mechanism to meet the different compression rates. They selected a group of human limb bones with the center bone angle as the movement characteristics, and then the use of layered curve simplification algorithm was developed for key-frames. Bulut et al. [9-10] proposed curve saliency algorithm, which treated the input motion sequence as a curve, and find the most salient parts of this curve which are crucial in the representation of the motion behavior, then apply key frame reduction techniques to obtain the most important key frames of the motion.

Clustering is also used for extracting key frames. Zhuang et al.[11] adopted an unsupervised clustering method to extract video key frames adaptively for video content analysis. Liu et al.[12] proposed a method based on clustering of extracting key-frames adaptively. They clustered each frame into a corresponding group in defined similarity. Then all of the first frames in those group composite key frames set. Park and Shin et al.[13] chose quaternion as their representation for motion data, and used PCA and k-means clustering to linearize quaternions and cluster them. Then they utilized sporadic data interpolation to extract key frames from clustered motion data. Zhang et al.[14] used Artificial Fish Swarm Algorithm and k-means to extract key frames in video. Xia et al.[15] proposed an new way of extracting key frame is Self-Splitting Competitive Learning(SSCL). It is effectively to calculate the numbers of key-frames by SSCL that get global optimal clustering numbers validly. At the same time, key frame will be exacted by clustering center.

However, there are some methods for extracting key-frames based on optimal error. Huang et al.[16] proposed a novel constraint-based key frame extraction technique named Key Probe. In contrast to previous curve simplification or clustering methods, the key frame extraction problem is cast as a matrix

factorization problem and the problem is solved based on the least-squares optimization technique. Clifford et al.[17-18] proposed an entropy-based motion extraction method, which is based on entropy metrics, specifically the mutual information measure. The most significant key pose identifies the largest directional change in the motion data which will have the lowest mutual information level from all the candidate key poses. Less significant key poses are then listed with higher mutual information levels. Cuntoor et al.[19]used anti-eigenvalue method to extract and detect key frames. Anti-eigenvalues depend critically on the turning of the operator, whereas eigenvalues represent the amount of dilation along the eigenvector directions aligned with the direction of maximum variance. Liu et al.[20] proposed an optimization algorithm for extracting key-frames, which use of the genetic algorithm to determine the optimal solution and use of a probabilistic simplex method to expedite the speed of convergence. The resulting set of key frames is obtained by iterative application of the algorithm. Avrithis et al.[21]extracted key frames from MPEG video databases based on a multiresolution implementation of the recursive shortest spanning tree (RSST) algorithm, and using a fuzzy multidimensional histogram.

### 3. Our approach for key frames extraction

#### 3.1. Motion feature representation

This paper uses the following human model as showed in Fig.1(a).The model is composed of 31 joints(some nodes are close and almost overlap, so the graph does not see all nodes), which uses node as joint and uses the connection as a body. We label some joints that we will select them as distant features in the following.

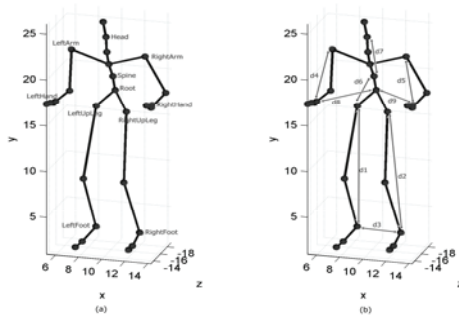


Fig.1. (a) The model of human skeleton (b) Nine distance features extracted

Table 1. Logical meaning of nine distance features extracted

d1: the left leg bending degree	d4: the left arm bending degree	d7: looked up/shook head degree
d2: the right leg bending degree	d5: the right arm bending degree	d8: the degree of left arm swing
d3: feet the size of level	d6: bent down the degree	d9: the degree of right arm swing

As showed in Fig.1 (b) and Table 1, we extract a group of new joint distances with logical meaning, which are  $\theta = (d1, d2, d3, d4, d5, d6, d7, d8, d9)$  as the distance characteristics. Above-mentioned features can better reflect the movement of the human body posture and greatly reduce dimensions of the original motion.

#### 3.2. The potential key-frames collection

We need to further reduce these features by PCA algorithm because selected features have some certain relevance (e.g. the left leg bent as the right leg).

The specific steps of PCA [14,22] are described as follows:

(i) Compute the average value of the nine features samples;

$$\bar{\theta}_i = \frac{1}{T} \sum_{j=1}^T \theta_{ji} \quad (i = 1, \dots, 9; j = 1, \dots, T) \quad (1)$$

Where T is the number of frames;

(ii) The difference  $\Delta\theta_{ji} = \theta_{ji} - \bar{\theta}_i$  produce a matrix  $D = [\Delta\theta_{j1}, \dots, \Delta\theta_{j9}]$ , where  $i = 1, \dots, 9; j = 1, \dots, T$ ; T is the number of frames;

(iii) Calculate the covariance matrix  $C = DD^T$ , and then calculate the eigenvalues  $\lambda$ , the corresponding eigenvectors  $\mathcal{G}$ ;

(iv) Finally, extract the feature vector  $\mathcal{G} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9)$ .

We sort the each component of the vector  $\mathcal{G}$  by PCA from the largest to the smallest of the contribution rate. But we just take the first dimension of greatest contribution(  $\theta_1$  ) as a feature represent the whole movement.

The feature  $\theta_1$  we obtained can be seen as a discrete one-dimensional time signal, which also is a feature curve. Then we employ the 'Lowess' smooth method to reduce noise of the curve. Potential key-frames can be obtained through collecting those frames that local extreme points occur.

Next is introduced how to get the local extreme points. The steps as follow:

(i) We get the local maximum using the function of 'findpeak'(in MATLAB);

(ii) Then continue to use the 'findpeak' function in after inverted data to obtain the local minimum. The data then was inverted for continuing to use;

(iii) Merge the local maximum and minimum, we gain the local extremum points set in accordance with the order from small to large.

### 3.3. Split and merge of key-frames

There are some candidates appeared far away from each other. Meanwhile, we do not take into account for the intensity of the movement. For lower error requirement and stronger ability to summarize human motion, one or more frames should be selected as candidate key-frames between those adjacent candidates. We believe that the motion is severe and posture change large if the curve has a large range, and posture change small if the amplitude is small.

Split and merge of key-frames based on amplitude of distance curve characteristic is described as follows:

(i) According to the curve, we can calculate the amplitude value of adjacent candidate key-frames.

(ii) We set a threshold. The swing is less than it, which illustrates the adjacent key frames posture change not huge, otherwise, which account for the section of the motion is intense, need to insert one or more frames between two adjacent candidates depend on the amplitude of the curve. The larger the magnitude to insert the more frames. For instance, the range is 0.4, we don't need to insert any frames; If the range is 10, we need to insert 10 frames with uniform sampling intervals, then add to the candidate key-frames collection.

(iii) The method results in a need to merge some candidate key-frames near to each other. We use a way to restrict the numbers between the adjacent candidate key-frames. If the number of frames between the adjacent candidates is greater than a threshold, we save the candidate to the final key-frames collection.

The typical case in point is a 'kick ball' to verify our method above-mentioned, as showed in Fig.2.

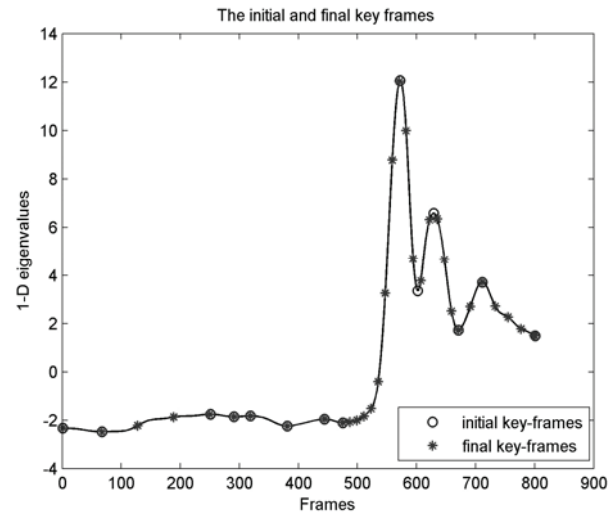


Fig.2. Distribution of key-frames with 'kick ball' motion. Here we use blue circle to represent initial key-frames and use red asterisk to denote the final key-frames.

We extract local extreme frames as potential key-frames represented by blue circle and we gain the final key-frames denoted by red asterisk, illustrated above Fig.2. By research the motion, we find that the first half of the movement is slow, and the middle part of the motion is fiercely. When motion is slow, the amplitude change small also, we don't need to insert extra key-frames. However, if the motion is fiercely, we need to split out of the new key frame between two adjacent extremes because the potential key-frames not a suitable response to the change. By using this method, we obtain the key-frames collection efficiency.

The advantage of our method is simple besides spending less time, and having a good performance in vision. And, when insert frames, we can compute the number of sampling automatically rather than determine the number of sampling interval frames manually in adjacent frames.

## 4. Experiment and analysis

In order to verify the validity of the above-method for motion key-frames extraction, we need to test a series of motion data. The motion capture data used for the experiments is from the Carnegie Mellon University (CMU) motion capture database [23]. All experiments are implemented in MATLAB R2009a on a PC with 2GB memory and a 3.10 GHz Core i3 processor.

**Experiment 1.** (Comparison of key-frames set of similar motion). We choose four different styles of a

certain frame size (130 frames) of walking motion to extract key frames. Fig.3 shows the result of key extraction for the similar motion with our method.

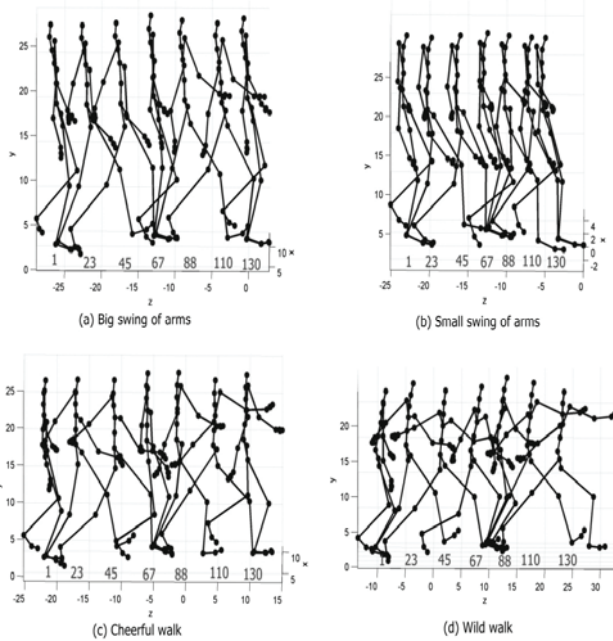


Fig.3. Key extraction with the same kind of similar motion (as example of walking motion) (a) Big swing of arms (b) Small swings of arms (c) Cheerful walk (d) Wild walk

It can be seen from Fig.3(a)-(d), extracted key frames can be a better response to a similar motion with different styles and key frame sequences of four similar motions correspond to each other basically.

**Experiment 2.**(Comparison of the different key frames extraction algorithm) We apply different methods: our method, uniform sampling, curve saliency [9] and only quaternion distance, extract the same key-frames(compression ratio) from one motion sequence ‘kicking ball’ with 802 frames(10\_01.bvh). The comparison result of different approaches is shown in Fig 4.

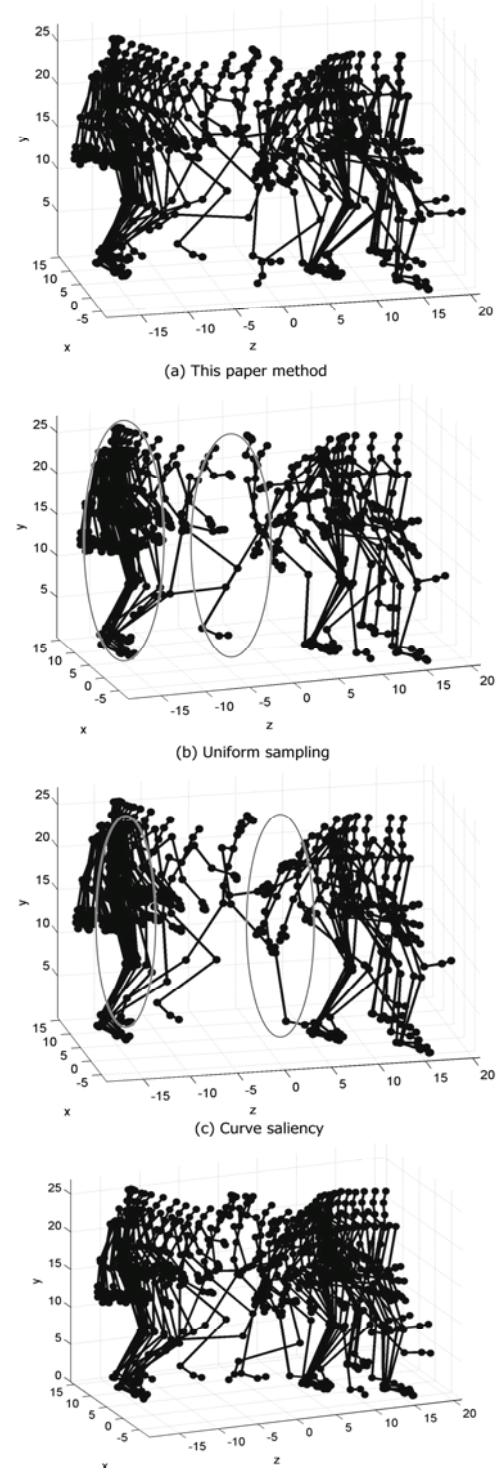


Fig.4. Comparison of key frame extraction algorithm for non-regularly motion (a) This paper method. (b) Uniform sampling method, elliptic represents over-sampling and under-sampling.



(c) Curve saliency [9], elliptic represents over-sampling and under-sampling. (d) Only quaternion distance.

As showed in Fig.4, ‘kick-ball’, one non-periodical motion, this has a dramatic change in the middle part. In (a), our method can select extreme and important transitional postures from motion sequences from a visual perspective. We extract 33 key-frames, excellent summarized in the movement, including the transition of the middle movement, to avoid over-sampling and under-sampling problems. It can be seen from (b), the use of uniform sampling it has also extracted 33 key-frames, but cannot be achieved on the original motion data summarized due to over-sampling and under-sampling. Over-sampling result in frame redundancy in the beginning of the motion with less pose change, under-sampling result in loss of frame in the middle of motion with great pose change. In (c), we found that the curve saliency method also have over-sampling and under-sampling problems when we extract 33 key-frames. In (d), the quaternion distance method also can tackle over-sampling and under-sampling under the extract 33 key-frames, but exist a higher reconstruction error than our method we will introduce in the following section.

We can see that under the same compression ratio situation, our method achieve the best result with good generalization ability of the original motion data than other methods.

**Experiment 3.** We test six different kinds of motion sequences, which is kicking ball, jumping, running-stop, walking, dancing, and walking-jumping-walking, apply our approach and other methods on those motions.

Table 2. Comparison of compression ratio in six different motion types

Category	Kickin g ball	Jum ping	Running -Stop	Walk ing	Dan cing	Walking- Jumping- Walking
Total number	802	439	239	343	1033	1200
Key- frames	33	24	11	16	37	50
Compres sion ratio	4.1	5.5	4.6	4.6	3.5	4.3

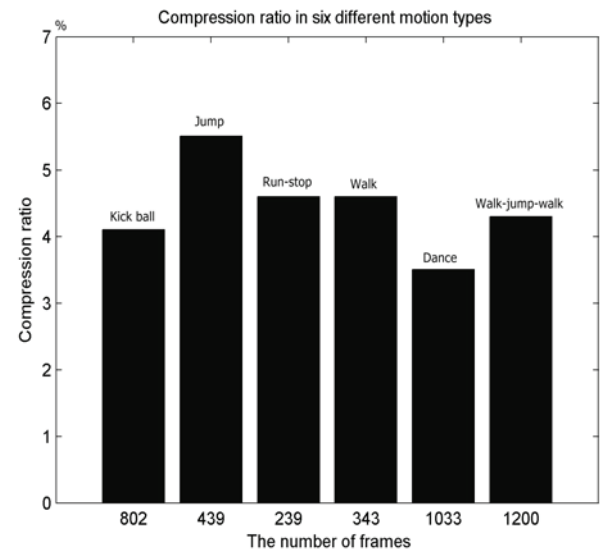


Fig.5 Comparison of compression ratio in six different motion types

From Table 2 and Fig.5, we found that our method achieved lower compression ratio above lower 10%.

Then, we compute the mean absolute error with the following formula:

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

Here,  $F(n)$  is the value and joints in the original motion data and  $F'(n)$  is the values of joints in reconstructed motion data respectively.  $N$  is the number of frames multiplied by 96 dimensions.

The reconstructed the mean absolute errors with three methods in six sampling motion sequences are shown in Table 3, in which the number in brackets represents the number of key frames extracted. At the same time, Fig.4 show the comparison figure of this error obviously.

Table 3. Comparison of the error rate in four methods

Category	Kick ball (33)	Jump (24)	Run-Stop (11)	Walk (16)	Dance (37)	Walk- Jump- Walk(50)
Our method	2.75	2.02	7.58	2.73	4.90	4.58
Quaternion	4.09	1.97	16.24	6.68	9.40	10.66
Curve saliency	14.71	2.48	7.61	5.20	7.17	7.12
Uniform sampling	15.25	2.53	7.79	2.94	7.49	4.86
Ratio	4.11	5.5	4.6	4.6	3.5	4.1

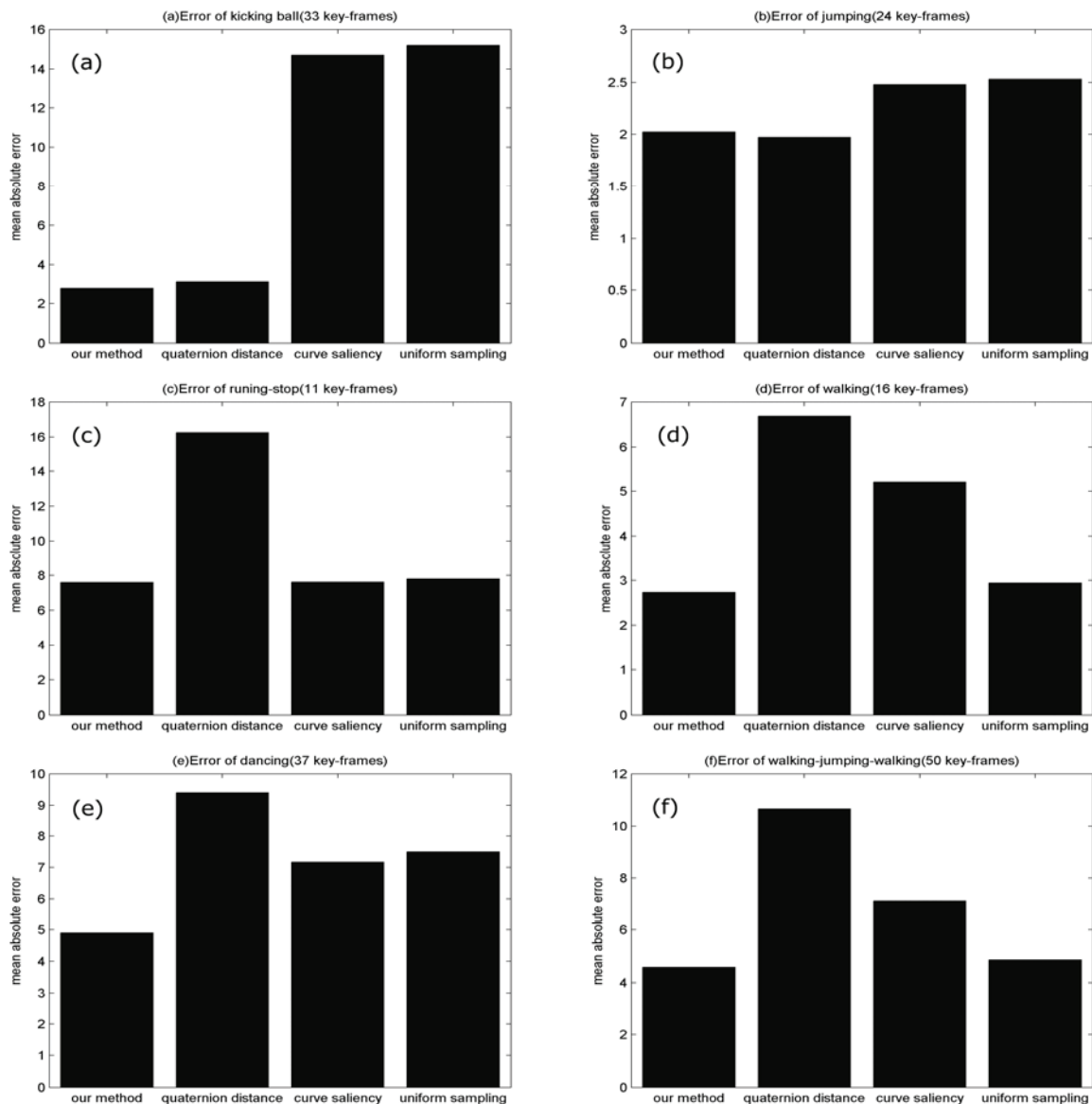


Fig.6. Comparison of mean absolute error by four methods of six sampling motion (a) Error of kicking ball (b) Error of jumping (c) Error of running-stop (d) Error of walking (e) Error of dancing (f) Error of walking-jumping-walking

We use four kinds of methods to get the same number of key frames from the same motion type. Then we reconstruct the motion sequence by the linear interpolation. As showed in Fig.6 (a)-(f), we can found that mean absolute errors with our method are less than with only quaternion and curve saliency[9] and uniform sampling under the same compression ratio. That is expected result, because it extracts the optimum key-frames which can be achieved the minimum error. We also show that our approach takes more obvious

advantage in the non-regular motion as showed in Fig.6 (a), (b), (e), and has a good performance in regular motion in Fig.6 (c), (d), (f). In conclusion, our method will spend less time and work well in the movement sequence by testing large amounts of data

## 5. Conclusion

The main contribution of this paper has two aspects. First, extract a feature curve by selecting a group of new distance characteristics, which can be better reflects the movement characteristics. Second, we obtain the final

key-frames set combine with uniform sampling based on amplitude of the curve separation automatically. Our method can be better adaptive motion, by testing large amount of motion sequences, which extract the less number of key-frames in the slow motion and extract more in the intense movement. Our key-frames extraction method can be used for multiple applications including motion summarization for browsing, motion retrieval, motion compression and reconstruction and so on.

In the future work, we will try to find a new better evaluation mechanism to evaluate the key-frames extracted stand or fall and to get the key frame collection with proper key frame number automatically. Meanwhile, we will select a more suitable interpolation algorithm to reconstruct the motion with the minimum reconstruction error for the decrease of the number of key frames.

## Acknowledgement

This work is supported by the Program for Changjiang Scholars and Innovative Research Team in University (No.IRT1109), the National Natural Science Foundation of China (Nos. 61370141, 61300015) the Program for Liaoning Science and Technology Research in University (Nos.LS2010008,LS2010179), the Program for Liaoning Innovative Research Team in University (No.LT2011018), Natural Science Foundation of Liaoning Province (Nos.201102008, 2013020007), the Key Project of Chinese Ministry of Education (No.211036), the Program for Liaoning Excellent Talents in University (No.LR201003, LJQ2013132), the Program for Liaoning Key Lab of Intelligent Information Processing and Network Technology in University and by "Liaoning BaiQianWan Talents Program (2010921010, 2011921009)".

## 6. References

1. X. P. Wei, R. Liu, Q. Zhang. Review of techniques for motion capture data processing. *Computer Aided Drafting, Design and Manufacturing*. 1(22),2012,pp.1-11
2. I. S. Lim and 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, Turkey, 2001), pp.1167-1169
3. H. Ogawa, and M. Okuda. Position-based keyframe selection for human motion animation. *Proceedings of the 11th International Conference on Parallel and Distributed Systems Workshops*, (Fukuoka. 2005), pp. 182-185
4. S. Y. Li, M. Okuda, S. Takahashi. Embedded Key-Frames extraction for CG Animation by Frame Decimation. *Proceedings of IEEE International Conference on Multimedia and Expo*. (Amsterdam. 2005), pp. 1404-1407
5. J. Assa, Y. caspi, and D. Cohen. Action synopsis: pose selection and illustration. *ACM Transactions on Graphics*. 24(3) (2005)667-676
6. L. Zhao, W. Qi, and S. Z. Li. Key-frame extraction and shot retrieval using nearest feature line (NFL). *Proceedings of ACM Workshops on Multimedia, Los Angeles, CA.2000*, pp. 217-220
7. J. Xiao, Y. T. Zhuang, T. Yang, and F. Wu. An efficient keyframe extraction from motion capture data. *24th Computer Graphics international conference*. 4035,(2006), pp. 494-501
8. J. Xiao, Y. T. Zhuang, F. Wu, T. Q. Guo, and Z. Liang. A group of novel approaches and a toolkit for motion capture data reusing, *Multimedia Tools and Applications*. 47(30), 2010, pp. 379-408
9. C. H. Lee, A. Varshney, and D. W. Jacob. Mesh saliency. *ACM Transactions on Graphics, Proceedings of ACM SIGGRAPH 2005*,24, 2005,pp. 659-666
10. E. Bulut, and T. Capin. Key Frame Extraction from Motion Capture Data by Curve Saliency.*Proceedings of the 20th International Conference on Computer Animation and Social Agents*, (Hasselt, 2007), pp. 182-185
11. Y. T. Zhuang, Y. Rui, T. S. Huan. Adaptive Key Frame Extraction Using Unsupervised Clustering. *1998 International Conference*, (Chicago, USA.1998), pp. 866-870
12. F. Liu, Y. T. Zhuang, F. Wu, and Y. H. Pan. Motion Retrieval with Motion Index Tree. *Computer Vision and Image Understanding*, 92(2/3)(2003)265-284.
13. M. J. Park, and S. Y. Shin. Example-based motion cloning. *Computer Animation and Virtual Worlds*. 15(3/4)(2004)245-257
14. J. M. Zhang, H. Y. Liu. Key frame extraction based on Artificial Fish Swarm Algorithm and k-means. *Transportation, Mechanical, and Electrical Engineering* (2011 International Conference on, 2011), pp. 1650 - 1653
15. L. M. Xia, K. J. Deng. Key frames extraction based on self-splitting competitive learning. *Computer Engineering and Applications*. 47(2) (2011)146-148.
16. K. S. Huang, C.F. Chang, Y. Y. Hsu, and S. N. Yang. Key Probe: a technique for animation keyframe extraction. *The Visual Computer*. 21(8/10)( 2005)532-541
17. Clifford. K. F. So, G. Baciuc. Entropy-Based Motion Extraction for Motion Capture Animation. *Computer Animation and Virtual Worlds*. 16,(2005)225-235
18. Z. Y. Qu, W. Q. Niu, C. X. Niu. Yanqiu Cao .Key frame extraction based on information entropy and edge



- matching rate. *Future Computer and Communication, (2010 2th International Conference on. 2010)* pp.91-94
19. N. Cuntoor, and R. Chellappa. Key frame-based activity representation using antieigenvalues. *Proceedings of Asian Conference on Computer Vision*, (Hyderabad, India. 2006),pp. 499-508
  20. X. M. Liu, A. M. Hao, and D. Zhao. Optimization-based key frame extraction for motion capture animation. *The visual computer*. 2012
  21. Y. S. Avrithis, A. D. Doulamis, N. D. Doulamis, S. D. Kollias. A Stochastic Framework for Optimal KeyFrame Extraction from MPEG Video Databases. *Computer Vision and Image Understanding*. 1-2 (75) (1999)3–24
  22. J. Barbic, A. Safonova, J. Pan, C. Faloutsos, J. Hodgins, and N. Pollard. Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004*, (Canadian Human-Computer Communications Society School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada. 2004), pp.185-194
  23. CMU Mocap Database. <http://mocap.cs.cmu.edu/>.