A Clustering Compression Method for 3D Human Motion Capture Data

Kai Zhou†, and Feng Tian†

† School of Computer and Information Technology, Northeast Petroleum University, Daqing, China

Summary

Human motion capturing has become an important tool in fields such as sports sciences, biometrics, and particularly in computer animation, where large collections of motion material are accumulated in the production process. Efficient storage, retrieval and transmission methods are needed to fully exploit motion databases for reuse and for the synthesis of motions. In this paper, a compression method for 3D Human motion data is proposed. We represent and compress the motion data using the clustering method and primary component analysis. The compressed data is adapted to network transmission with shorter time in order to maximize the use of network bandwidth and computational performance of local machines. At the client, we decompress the motion chips and rebuild corresponding human motion. Experimental evaluation of the method showed that the proposed method has high compression rate and is effective.

Key words:

Human motion compressing, Motion Capture, Motion Animation

1. Introduction

Virtual human is an important component of virtual environment, and the research of the human animation based on motion capture data that has a high efficiency and more realistic is very important in the research field of computer graphics and widely used in a lot of virtual human application fields. The investigation of human motion has drawn a great attention from researches in computer vision or computer graphics recently. Fruitful results can be found in many applications such as visual surveillance [1-2], diagnosis and therapy for rehabilitation [3-4], athletic training [5-6], person identification [7-8], animation generation [9-10], user interface [11-12], and so on. Modern motion capture technology can easily produce realistic motion data with high precision and fine detail. Using motion capture data with motion synthesis method, we can get more new motion sequence that can satisfy the user's requirement.

As tools and systems for producing and disseminating motion data improve significantly, the amount of human motion data grows rapidly. Therefore, an efficient approach to compress human motion data is needed. Three-dimensional capture data is different from traditional 2D data. For example, the kick sequences shown in Figure 1 describe a kind of kick motion. Suppose that human motion data are composed of a consecutive sequence of frames as shown in Figure 1. In the sequence, each frame corresponds to a posture, which is structured information of human skeletal segments in the 3D space. These skeletal segments move simultaneously from one frame to the next forming multiple 3D motion trajectories as the sequence proceeds. The length of motion sequences are various depending on applications. In this study, we confine ourselves to motion data of relatively long sequence. For example, a twenty minutes Chinese martial art or aerobic performance has 72000 frames at 60Hz and each frame contains about twenty feature values associated to the posture. Therefore, the huge 3D human motion data is a challenge to current graphics hardware and storage space. On the other hand, the rapid development of Internet has made it possible to display digital collections online. However, the current network bandwidth limitation is seriously hampering such huge 3D media transmission. It is not sufficient to solve this problem by relying only on increased investment in hardware equipment. Therefore, the main contribution of this study is that we propose effective and efficient method to resolve the problem of three-dimensional animation compression for human motion data. The rest of this paper is organized as follows. Section II reviews related work. Section III presents how to represent virtual human. Sections IV and V detail human motion data information extracting and compression file design method, respectively. Section VI describes our experimental design and evaluates the retrieval results. Section VII concludes our study.

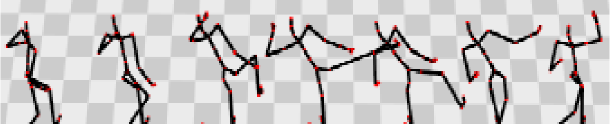

Fig.1. a side kick sequence.

2. Related Work

In the following, we investigate the background related to our work, including human motion capture and 3D data compression. Since we restrict our attention to the compressing in human motion storage, we omit some

Manuscript received August 5, 2014
Manuscript revised August 20, 2014

related issues such as human motion analysis, retrieval and reconstruction.

The most common recording technology for motion capture data uses an array of digital cameras to three-dimensionally track reflective markers attached to a live actor's body. We exemplarily discuss an optical marker-based technology, which yields very clean and detailed motion capture data. Here, the actor is equipped with a set of retro-reflective markers attached to a suit. These markers are tracked by an array of six to twelve calibrated high-resolution cameras, see Figure 2. The tracking data can then be post-processed to obtain a multi-stream of 3D trajectories corresponding to the joints of a fixed skeletal kinematics chain as indicated by Fig. 3. A full set of 3D coordinates describing the joint positions of a kinematics chain for a fixed point in time is also referred to as a pose. A motion capture data stream is thought of as a sequence of poses or frames, typically sampled at 30–600 Hz.

Fig.2.Optical motion capture system based on retro-reflective markers attached to the actor's body.

At present, compression of topology is close to the theoretical limit, while geometry coding is becoming a hot topic. Early researches on 3D mesh compression focused on single-rate compression techniques to save the bandwidth between the CPU and the graphics card. With the increasing popularity of networked applications, progressive compression and transmission of 3D meshes have been studied intensively. At the beginning, the connectivity coding drives the geometry coding. However, since the geometry data occupy most of the storage space in a model file, the connectivity-centric algorithms restrict the efficiency of geometry compression. Thus geometry-centric algorithms have emerged in recent years such as the kd-tree mesh codec, the octree mesh codec, the spectral coding geometry codec and wavelet-based mesh codec.

Recently most research has focused on the compression of grid data where 3D animation compression and reconstruction can be achieved according to the continuity characteristics of the 3D grid topology vertices in the trajectory of the each frame. In [13], a space-time compression method of the 3d animations of triangle meshes with fixed connectivity is described. Guskov and Andrei describe another method of wavelet compression of parametrically coherent mesh sequences [14]. In papers [15-17] the rigid portions are compressed and implement efficient rendering animation. Sattler compressed grid animation based on primary component analysis clustering [18]. Park adopts quaternion to represent high-dimensional data [19]. Wolf extract key frame by motion analysis [20]. Zordan picked up the key frames by defining the similarity of adjacent frames [21].

In this paper, we propose a new compression method for 3D human motion capture data, which meets the current data storage and network transmission requirements. First, the motion information is extracted by clustering analysis. Secondly, the dimension is reduced by the method of principal component analysis in each cluster. Finally, the structure of compressed data is designed to store compressed information.

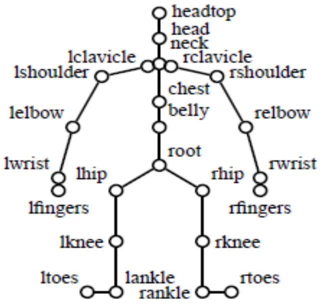

Fig.3. Skeletal model labeled with joint name

3. Virtual Human Motion Representation

The human skeleton can be approximately represented as an articulated figure, which is a tree-like hierarchy consisting of a set of joints and bones. Joints and bones respectively make the inner nodes and edges of the tree-like hierarchy (Fig. 3). In the hierarchy, each joint has exactly one parent except for the root. The skeleton is assumed to have rigid bones, which means bones are length-fixed straight lines. Note that we only consider the whole body movement here, and do not consider the minor movement of fingers, toes and face muscle. So, the skeleton model in Fig. 3 does not have the detailed structure of these parts.

It is a quite straightforward way to use the 3D coordinates of these joints to describe body motions. But this kind of motion representation will bring us much inconvenience in later data manipulating work, because it describes the movement of each joint independently and can not treat the skeleton as a whole body. So, we use the hierarchical motion representation method: the movement of a human body is comprised of translation and rotation. The root can both translate and rotate. The translation and rotation of the root respectively change the position and orientation of the body. All other joints can only rotate around their parents to change the orientation of corresponding bones.

The translation and rotation of the root take place in the world coordinate system, and the rotation of other joints is relative to their parent. The root has 6 DOFs — 3 DOFs for translation and 3 DOFs for rotation. Other joints have 1-3 DOF(s) for rotation. For example, wrist has 3 DOFs and elbow has only 1 DOF. The hierarchical motion representation implicitly hinges the adjacent bones together by removing the translation DOFs from all joints except the root. So, we do not need to maintain explicit constraints to hold all bones together. The motion data represented by the hierarchical method typically includes the hierarchy definition, the sampling data sequences for the position and orientation of the root and all joints. The position and orientation of the root and the orientation of joints make up body pose parameters. A set of values of body parameters specifies a fixed body pose. A motion can be represented by a function which describes the relationship between time and body parameters:

$$F_i = (p_0(t), q_0(t), q_1(t), q_2(t)...., q_n(t)) \qquad (1)$$

where $p_0(t)$ and $q_0(t)$ represent respectively the position and the orientation of the root node. Arikan introduces a kind of motion data representation based on virtual flag [22]. First, the motion data are divided into N sections where each section composing by Fr frames data. The location and orientation of the body center bone are recoded by mark; Secondly, denote the rigid coordinates of animation clips' transform as virtual flag (a, b, c), since human bone is a chain structure and the joint Euler angle space is not suitable for space compression; Thirdly, Extract three directions virtual flag by multiplication of displacement matrix and rotation matrix

$$T^k * R^k * \begin{bmatrix} 1001 \\ 0101 \\ 0001 \end{bmatrix}^T = \left[a \mid b \mid c \right] \qquad (2)$$

where T^k and R^k respectively represents the transformation matrix and rotation matrix of the bone (k). Bone in each frame is represented by flag ($[a \mid b \mid c]$), rather than the three rotation angle. We use least squares approach to fitting 3D Cubic Bezier curve as

$$P(t) = \left[t^3 \quad t^2 \quad t^1 \quad 1 \right] \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} \qquad (3)$$

4. Information Extraction

Since each animation clip is composed of Vector (X_i). The entire virtual human animation can be described by vector space $R(X_1, X_2, \cdots, X_n)$. First, we cluster these vectors by K-Means. Secondly, the dimension is reduced by principal component analysis in each cluster. Then the average vector m_c and the fundamental matrix P_c are computed, where P_c is the eigenvectors of covariance matrix row vector. Vector X_i is re-expressed as $\tilde{x}_i = P_c(x_i - m_c)$. Similarly, X_i can be computed by $x_i = P_c^T \tilde{x}_i + m_c$. We can draw conclusions that the front p rows of P_c represent the principal component.

5. Data Compression File Design

The file structure includes following content. First, the average vector of each cluster is needed; Secondly, the matrix of each cluster and the row number; Thirdly, the index of clip in cluster and new coordinates of each clip; At last, the first frame's absolute position and orientation. The compressed file and decompressed as follows: First, we select animation clips to decompress by the absolute position and orientation of the first frame; Secondly, we obtain the clip average vector and get position and orientation information according to the equation.

6. Experiments

We use the motion capture data from the CMU Graphics Lab Motion Capture Database (http://mocap.cs.cmu.edu) as our testing collection to show the effectiveness of the proposed method. Our method is implemented in Java using Java 3D SDK. Table I shows the motion data compression ratio.

Table Ⅰ: motion data compression results and compression ratio

Original data size	**Compressed datasize**	**Compression ratio**
1.67M	245K	1:7.5
4.72M	651K	1:7.8
8.71M	1.2M	1:8.0
14.18M	1.78M	1:8.2
23.41M	2.34M	1:12.5

7. Conclusion

In this paper, we propose a new compression method for 3D human motion capture data, which meets the current data storage and network transmission requirements. The main contribution of this study is that we propose effective and efficient method to resolve the problem of three-dimensional animation compression for human motion data.

Acknowledgment

This work is supported by Youth Foundation of Northeast Petroleum University (NO: 2013NQ120, 2012QN117), Scientific Research Fund of Heilongjiang Provincial Education Department (NO: 12521055). We also would like to thank the anonymous reviewers for their helpful comments and suggestions.

References

[1] I. Haritaoglu, D. Harwood, L. S. Davis, W4."real-time surveillance of people and their activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 2000, pp. 809-830.

[2] N. M. Oliver, B. Rosario, A. P. Pentland."A Bayesian computer vision system for modeling human interactions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 2000, pp. 831-843.

[3] M. Kőhle, D. Merkl, J. Kastner, "Clinical gait analysis by neural networks: issues and experiences", IEEE Symposium on Computer-Based Medical Systems, 1997,pp. 138-143.

[4] D. Meyer, J. Denzler, H. Niemann, "Model based extraction of articulated objects in image sequences for gait analysis", IEEE International Conference on Image Processing, 1997, pp. 78-81.

[5] J. W. Davis, A. F. Bobick, "Virtual PAT: a virtual personal aerobics trainer", Workshop on Perceptual User Interfaces, 1998, pp. 13-18.

[6] P. T. Chua, R. Crivella, etl, "Training for physical tasks in virtual environments: Tai Chi", IEEE International Conference on Virtual Reality", Mar. 22-26, 2003.

[7] C. BenAbelkader, R. Cutler, L. Davis, "Person identification using automatic height and stride estimation", IEEE International Conference on Pattern

[8] A. F. Bobick, A. Johnson, "Gait recognition using static activity-specific parameters", IEEE Computer Vision and Pattern Recognition, Hawaii, Dec. 8-14, 2001.

[9] F. Multon, L. France, M.-P. Cani-Gascuel, "Computer animation of human walking: a survey", The Journal of Visualization and Computer Animation, 10(1), 1999, pp. 39-54.

[10] O. Arikan, D. A. Forsyth, J. F. O’Brien, "Motion synthesis from annotations", ACM Transactions on Graphics, 22(3), 2003, pp. 402-408.

[11] W. T. Freeman, P. A. Beardsley, H. Kage, "Computer vision for computer interaction", ACM SIGGRAPH Computer Graphics, 33(4), 1999, pp. 65-68.

[12] J. Lee, J. Chai, J. K. Hodgins, P. S. A. Reitsma, "Interactive control of avatars animated with human motion data", ACM Transactions on Graphics, 21(3), 2002, pp. 491-500.

[13] IBARRIA, L., AND ROSSIGNAC, J. 2003. "Dynapack: space-time compression of the 3d animations of triangle meshes with fixed connectivity". In Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation, 126–135.

[14] GUSKOV, I., AND KHODAKOVSKY, A. 2004. "Wavelet compression of parametrically coherent mesh sequences". In Proceedings of the 2004 ACM SIGGRAPH/ Eurographics symposium on Computer animation, 183–192.

[15] JAMES, D. L., AND TWIGG, C. D. 2005. Skinning mesh animations. Proceedings of SIGGRAPH 2005 24, 3, 399–407.

[16] WANG, X. C., AND PHILLIPS, C. 2002. "Multi-weight enveloping: least-squares approximation techniques for skin animation". In Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation, 129–138.

[17] MOHR, A., AND GLEICHER, M. 2003. "Building efficient, accurate character skins from examples". Proceedings of SIGGRAPH 2003 22, 3, 562–568.

[18] SATTLER, M., SARLETTE, R., AND KLEIN, R. 2005. Simple and efficient compression of animation sequences. In Proceedings of the 2005 ACM SIGGRAPH/ Eurographics symposium on Computer animation, ACM Press, 209–217.

[19] Park, M.J., Shin, S. Y. "Example-based motion cloning, "Comput. Animat. Virtual Worlds,2004,15(3-4):245-257

[20] W. Wolf. "Key frame selection by motion analysis." In Proc. IEEE Int. Conf. Acoust, Speech and Signal Proc.,1996

[21] V. B. Zordan, A. Majkowska, B. Chiu, M. Fast. "Dynamic Response for Motion Capture Animation." In Proceedings of SIGGRAPH 2005, ACM SIGGRAPH.

[22] O. Arikan, "Compression of Motion Capture Database," Proc. ACM Trans. Graphics (ACM TOG), vol. 25, no. 3, pp. 890-897, 2006

Kai Zhou received the M.E. degrees from Northeast Petroleum University in 2006. She is research assistant since 2006 and is working as a lecturer from 2009 in the Department of Computer and Information Technology, Northeast Petroleum University. Her research interest includes pattern recognition, virtual reality.

Feng Tian born in 1980. PhD and associate professor. Received his PhD degree in computer application technologies from the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University in 2014. His main research interests include image annotation, image tagging, cross media analysis, multimedia mining, 3D model retrieval, virtual reality and pattern recognition