2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics

Data transform for dance motion capture based on Kinect

Licheng Wu, Yu Yang, Xiaer Li
School of information engineering
Minzu University of China
Beijing, China
wulicheng@tsinghua.edu.cn

***Abstract*—this paper presents a data conversion method between Kinect raw motion data and standard BVH format file. To verify the feasibility of conversion and correctness of data, the motion data in BVH format is imported into MotionBuilder, 3D animation software, forming a demo animation. The results show that data conversion is workable and data are correct.**

Keywords-Kinect; motion capture; BVH; 3D animation

I. INTRODUCTION

National dance, as the carrier of human action and emotion, is an important part of intangible cultural heritage. Traditional record-keeping method is technically two-dimensional, including text, pictures and video etc. The above methods cannot give a detailed description of dancer's body movement and motion trajectory, not to mention an analytical research. With the help of digital information acquisition and processing technology, three-dimensional motion capture technology, which takes account of dance's characteristics, is studied and dance can be saved as a standard motion data file. Motion capture technology provides a new means for national dance's digital protection and has an important practical significance [1].

Motion capture technology can track, measure and record human's motion trajectory in 3D space. Human's motion is captured and digitized for future playback, analysis and utilization. Its basic principle is recording human's action and trajectory as movement data such as displacement and angle, then using these data to drive virtual character to restore human's motion. At present in principle, it can be divided into four types of machinery, acoustics, electromagnetism and optics [2]. Optical motion capture system is the most mainstream and almost all films that use motion capture technology use optics. As it works best, its disadvantage cannot be ignored, that is high-priced, sensitive to environment factors such as light, reflection and large amount of data post-processing [3].

Kinect is XBOX360 somatosensory peripherals developed by Microsoft. Its composition includes an RGB camera, an infrared emitter, an infrared depth camera, four microphone arrays and an angle control motor. The infrared depth camera receive reflected infrared signal that is emitted by infrared transmitter and convert it to depth information. Thus Kinect senses the distance with the object that reflect infrared signal. As Kinect's core technology, skeleton tracking function can track up to 6 people with their position information and provide 2 people's skeleton information, which include the three-dimensional coordinates of 20 points, relevant rotation matrix and other information. As the motion data can be provided at the rate of 30 frames per second, it meets the real-time interaction need. Researchers have developed different applications based on the Kinect skeleton information in terms of gesture recognition, human-computer interaction and intelligent control [4][5]. These efforts have achieved good results.

Reference [6] calculates Euler angles with Kinect's original coordinate data and save them as the standard BVH file. This paper studies data conversion technology and method between Kinect raw motion data and standard motion capture file and presents a new data conversion method using Kinect's quaternion and coordinate data. A main aim is correct conversion and storage of motion data. The program is achieved by WPF application in C# language. Finally, BVH file is imported into MotionBuilder software and successfully drive a 3D actor model to reappear the action captured. Thus the feasibility of data conversion method and correction of BVH data is proved.

II. DATA TRANSFORM

A. Kinect motion data and motion capture file

The Kinect sensor provides original data stream to SDK(software development kit), which also includes Kinect device drivers. Applications get access to these data via SDK and are developed on that basic. Besides, Kinect SDK encapsulates high-level APIs that include skeleton tracking which is called by our application. The architecture is shown in Figure 1.

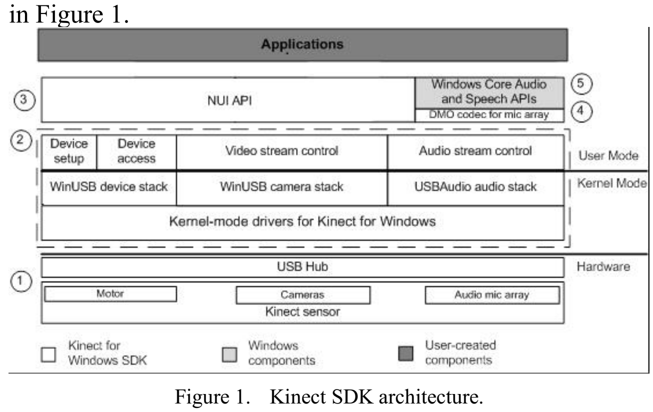

Figure 1. Kinect SDK architecture.

978-1-4799-8646-0/15 $31.00 © 2015 IEEE
DOI 10.1109/IHMSC.2015.137

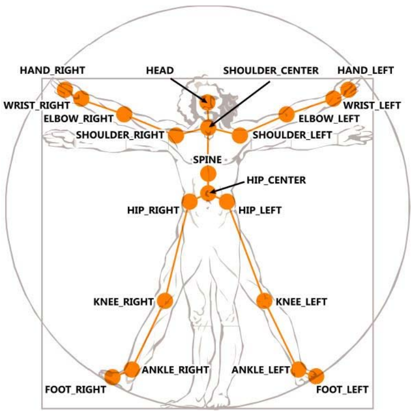

Figure 2. 20 joints of Kinect human skeleton.

When human is tracked, there is a matching skeleton that includes 20 skeleton joints, as shown in. Each joint also has a tracking state. If the tracking state is 'tracked', there is three-dimensional coordinate information that can be used in motion capture. In addition to providing tracked joint's position information, Kinect SDK can compute the rotation information of bone which is defined as BoneOrientation class. This class contains the following data members: StartJoint on behalf of the start joint of a bone; EndJoin on behalf of the end joint of a bone; HierarchicalRotation is the rotation information between the start joint's coordinate and the end joint's coordinate in a bone and AbsoluteRotation is the rotation information between the coordinate of a bone' end joint and Kinect basic coordinate.

In the field of motion capture, there are no uniform criterion and technical standard about data storage format to follow. As for commercial applications, many 3D computer graphics software have their own standard of 3D data. In this paper, from two aspects of data format's generality and data storage's efficiency, motion data is stored in BVH format is chose as the motion capture. BVH file consists of two parts: HIERARCHY and MOTION [7]. The first part describes a human skeleton as tree structure. The skeleton hierarchy is defined through offset of the child node relative to the parent node, thus the skeleton initial attitude is determined. Node represents skeleton joint while the line between parent and child node represents a bone. Each node has three rotation parameters (Euler angles) to describe the motion information of a bone, which is constructed by this node and its child node. Root node has 3 position parameters besides the above rotation parameters to describe translation in three axes. The second part records motion data, which is every joint's DOF measurement value in the order defined in the first part. Frame number and time interval between two frames is recorded too.

B. Data conversion

Due to the difference of skeleton hierarchy, coordinate systems between Kinect motion data and BVH data, there must be a data conversion process. This paper presents an efficient method of converting Kinect motion data to BVH format file.

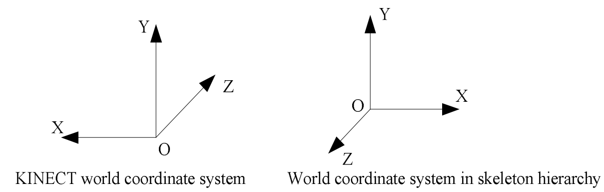

Figure 3. World coordinate system is different.

1) Skeleton data conversion: Skeleton data conversion describes how to obtain BVH's HIERARCHY section. The skeleton hierarchy has been built following H-Anim standard [8][9]. The size of skeleton is determined by Kinect's motion data of skeleton joint's three-dimensional coordinate. As the world coordinate system in Kinect is different from the one defined in skeleton hierarchy as shown in Figure 3. . coordinate data conversion is needed. To be specific, turn value in X and Z axis to opposite number and value in Y axis remain unchanged.

Take node Spine as an example, the offset data in HIERARCHY section is calculated as below:

$$\begin{aligned}&[xoffset \quad yoffset \quad zoffset]\\&=[-(Sp.x-HC.x) \quad Sp.y-HC.y \quad -(Sp.z-HC.z)]\end{aligned} \tag{1}$$

$xoffset$, $yoffset$ and $zoffset$ are node Spine's offset relative to its parent node Hips in X , Y and Z axis respectively in HIERARCHY section. $(Sp.x \quad Sp.y \quad Sp.z)$ and $(HC.x \quad HC.y \quad HC.z)$ represent 3D coordinate of Joint Spine and Joint HipCenter in Kinect coordinate system.

The length of a bone can be calculated according to segment length formula:

$$\|(x \quad y \quad z)\| = \sqrt{x^2 + y^2 + z^2}. \tag{2}$$

2) Motion data conversion: Motion data conversion describes how to obtain MOTION data in BVH file from Kinect motion data. i.e. DOF(degree of freedom). Considering one frame data, it is the translation and rotation information that make initial attitude become the current frame attitude.

Calculation of root node's DOFs:

Due to the defined offset of root node in HIERARCHY is $(0 \quad 0 \quad 0)$, root node's three translational DOF in MOTION data is actually the corresponding Kinect joint's 3D coordinate in world coordinate system. That is:

$$\begin{aligned}&(xoffset \quad yoffset \quad zoffset)\\&=(-HC.x \quad HC.y \quad -HC.z)\end{aligned} \tag{3}$$

Root node's three rotation DOF is Euler angles in the order of $X-Y-Z$, that is rotation matrix is combined by three basic rotation matrix around X, Y and Z axis. Assuming α is the first rotation around X axis, β is the second rotation around Y axis, and γ is the third rotation around Z axis, the rotation matrix is established as follows:

$$R_{zyx}=R(Z,\gamma)R(Y,\beta)R(X,\alpha)$$
$$=\begin{bmatrix} \cos\beta\cdot\cos\gamma & \cos\gamma\cdot\sin\alpha\cdot\sin\beta-\cos\alpha\cdot\sin\gamma & \sin\alpha\cdot\sin\gamma+\cos\alpha\cdot\cos\gamma\cdot\sin\beta \\ \cos\beta\cdot\sin\beta & \sin\alpha\cdot\sin\beta 2+\cos\alpha\cdot\cos\beta & \cos\alpha\cdot\sin\beta 2-\cos\beta\cdot\sin\alpha \\ -\sin\beta & \cos\beta\cdot\sin\alpha & \cos\alpha\cdot\cos\beta \end{bmatrix} . \quad (4)$$

HipCenter's absolute quaternion $AbsoluteRotation.Quaternion$ in Kinect is used to calculate the root node's Euler angle:

$$\begin{cases} \alpha = A\tan 2\left(2\left(wx+yz\right),1-2\left(x^2+y^2\right)\right) \\ \beta = -A\sin\left(2wy-2zx\right) \\ \gamma = A\tan\left(2\left(wz+xy\right),1-2\left(y^2+z^2\right)\right) \end{cases} . \quad (5)$$

As local coordinate system and referred world coordinate system are different, the ultimate rotation DOF of root node is $\left(-\alpha \quad \beta \quad -\gamma\right)$.

Calculation of node HipRight's rotation DOF is divided into two steps: construct quaternion and compute Euler angles. As Euler angles are calculated as the same as root node, quaternion construction is the focus.

Quaternion construction: form a rotation matrix $R_y(\beta)$ that expresses rotation about the Y axis by angle β where β is the father's second Euler angle.

$$R_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \quad (6)$$

The bone that node Hip Right represents is expressed in terms of vector $\overrightarrow{KH}$:

$$\overrightarrow{KH} = \left(Kn.x-HR.x \quad Kn.y-HR.y \quad Kn.z-HR.z\right) \quad (7)$$

In its father's coordinate system, bone vector is $\overrightarrow{KH}'$:

$$\overrightarrow{KH}' = \overrightarrow{KH}\cdot R\left(Y,\beta\right) \quad (8)$$

Then construct quaternion with angle between $\overrightarrow{KH}'$ and $\left(0 \quad -1 \quad 0\right)$ and rotation axis $\left(0 \quad -1 \quad 0\right)$.

In conclusion, the quaternion used to calculate rotation DOF is not provided by Kinect directly for some node. In this case, it is necessary to construct quaternion using bone vector and parent node's rotation information. Pay special attention to coordinate system conversion and joint correspondence.

III. 3D ANIMATION

Through data conversion, the Kinect original motion data is stored as BVH motion capture file. The feasibility of conversion and the correctness of data need to be verified. Besides, it is hard to get an intuitive understanding about the dance by reading the bone length or Euler angles in BVH.

MotionBuilder is an excellent 3D animation software using for games, video and multimedia production. It can read BVH format file and edit-modify motion data. When a 3D character model is created and characterized, it can be drove to move. It is possible to view the motion effect from different direction in the form of animation. The process is as Figure 4.

In Figure 5. , the first row shows four frame attitudes in motion capture video, and the second row is corresponding skeleton in captured BVH file. The third row shows the animation effect of BVH motion data driving 3D model. Experimental results show that the data conversion method is workable and BVH file is correct.

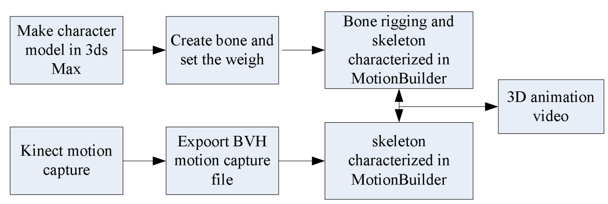

Figure 4. The process of forming animation using motion capture file.

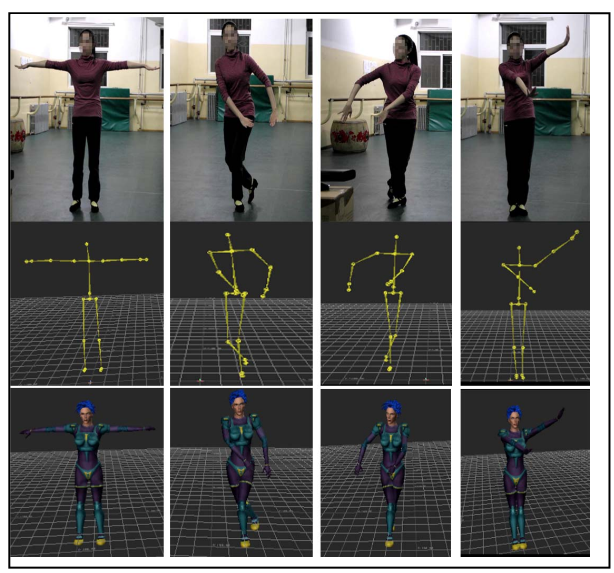

Figure 5. Four attitudes from motion capture to animation video.

IV. CONCLUSION

Motion capture technology provides a new approach to national dance's digital protection. This paper studies a motion capture system based on Kinect and discusses data conversion at length. The main goal of data conversion and storage in motion capture is achieved, and it is low-cost and easy to operate. But there are still some shortages needed to be improved, such as data missing or inaccurate problem. Data optimization and utilization are the future job.

ACKNOWLEDGMENT

This work was supported in part by the transition funds of the first-class university and the first-class discipline construction, the 2013 Beijing university youth talent plan No. YETP1294, the NSFC projects No. 51375504, and the Program for New Century Excellent Talents in University.

REFERENCES

[1] QIU Wand-biao, LI Chao. Research of protection methods about the dances artistic of national minorities based on motion capture technology. journal of university of electronic science and technology(social sciences edition) [J]. 11(4): 101-104(2009).

[2] XIANG Ze-rui, ZHI Jin-yi, XU Bo-chu. Survey on motion capture technique and its applications. Application research of computer[J],.30(8): 2241-2244(2013).

[3] WU Xiao-yu, ZHANG Yi-chun, SHEN Ying-hua. National dace 3D digitizing protection method based on motion capture technology. Computer and modernization[J]. 1: 112-114(2013).

[4] LIU Yang, SHANG Zhao-wei. Traffic gesture recognition based on Kinect skeleton data. computer engineering and applications[J]. 3: 157-161(2015).

[5] ZHU Guo-gang, CAO Lin. Human motion recognition based on skeletal information of Kinect sensor. Computer simulation[J]. 31(12): 329-333(2014).

[6] Tong X, Xu P, Yan X. Research on Skeleton Animation Motion Data Based on Kinect[C]. //International Symposium on Computational Intelligence & Design. IEEE, 2012:347 - 350.

[7] M.Meredith S.Maddock. Motion Capture File Formats Explained

http://www.dcs.shef.ac.uk/intranet/research/public/resmes/CS0111.pdf.

[8] ZHOU Qian-xiang, ZHAN Yu-hong, NIU Hai-yan. Modeling and application of virtual human technology efficiency[M]. Beijing: National Defence Industry Press.pp.151(2012).

[9] http://17de.com/x3d/hanim.html