Creation of 3D Human Avatar using Kinect

Kairat Aitpayev, Jaafar Gaber

***Abstract*— While the number of web and desktop 3D applications continue to increase, users would like to create their own 3D avatars which will look exactly like them. The Kinect[1] can help to users to reach this goal; after scanning human by Kinect's depth and RGB cameras it is possible to create almost the same 3D model including human faces and hairstyles that can be used for several algorithm in education, medicine, conferences, game industry, 3D social networks, etc.**

***Index Terms*— Kinect, Depth Camera, 3D Avatar, 3D Body Scanning.**

I. Introduction

NOWADAYS 3D technologies are becoming more and more popular, which by itself increases the amount of 3D applications in the social networks and game industries. There is the idea of using 3D avatars which look like their owners instead of models of heroes in games and predefined avatars in social networks. This was a difficult issue so far, since only well experienced developers with the necessary skills in 3D graphics were able to do it, and even today it takes a lot of time to do it manually.

Now after the release of Microsoft Kinect[1] and other depth cameras this idea has become more real. No need to know anything about 3D graphics to create one's own avatar.

The whole process takes a few minutes by doing the following steps. At the beginning a camera scans the whole body, afterwards it scans the face and makes snapshots of the head to reconstruct the hairstyle. At the last the user chooses clothes and other accessories which will be applied to the final model.

Today, Kinect is one of the most popular devices used in games in NI[2][3][4]. Kinect is used with Xbox 360 console and play with body movement. According to Microsoft, Kinect is designed for a revolutionary new way to play with no controller required [1].

Fig. 1. Depth camera data (left) and RGB camera data (right) visualized using Kinect with Microsoft SDK.

In this work, we are using Microsoft Kinect SDK[5]. This library is open-source and can be used for testing and implementations. Figure 1 illustrates depth image data and RGB data that is gotten from Kinect like interface and using Microsoft SDK to produce the data.

The data that is gotten from the Kinect's depth cameras is called depth data. As illustrated in Figure 1 (left) and differentiates depth with color (bright grey to dark gray). Therefore, depth data can be measured to calculate the distance from the camera to the object and can be manipulated to various things like body shape measurement, motion detection, distance measurement and others.

As the community of Kinect increases there will be more applications using depth camera. Today we already have a lot of online 3D games that were created using different technologies and recently the first 3D social network imvu[1] was released. In the near future virtual network, will be in 3 dimensions and depth cameras will be used like normal web-cameras today.

II. Creation of 3D Avatar

With Kinect scanning the whole human body, depth data has very pure information about the face, because the avatar creation process is divided into two independent parts:

A. Body creation

Instead of using a model which can be created by Kinect it is better to fit a predefined 3D model of human to the size of the user's body. Why this would be better? In Fact, when Kinect scans the environment in front of it, it creates the 3D model of weight 60,000 vertices. After removing unneeded background, weight of human model is approximately 14,000 vertices. Of course we can apply algorithms of smoothing and optimization, but in the end the model will be a little bit distorted. The best solution of this problem is to use a predefined low polygonal model of human with weight of 1,000 vertices.

The advantages of using a predefined model are:

1) Approximately three times less model weight.
2) Model looks more accurate.
3) Easy animation, because of already rigged model.

Kairat Aitpayev is with the Computer Science Department, University of Technology of Belfort-Montbeliard, 90000 Belforrt, France and Kazakh-British Technical University, 050000 Almaty, Kazakhstan (e-mail: kairat.aitpayev@utbm.fr, k.aitpayev@kbtu.kz).

Jaafar Gaber is with the Computer Science Department, University of Technology of Belfort-Montbeliard, (e-mail: gaber@utbm.fr).

[1] http://www.imvu.com

4) Ability to change size of avatar, type of hairstyle, color of eyes, etc.
5) Ability to applying different clothes and accessories.

There are three types of avatars, Figure 2: male, female, and kid. Here we meet with another problem of gender recognition, only kids can be recognized easily by the height of body. The choice between the other two avatars needs a more sophisticated solution. Kinect contains three devices for these purposes: RGB camera, microphone array, and depth camera. There is a lot of research work related to gender recognition using RGB camera and microphone [6][7]. In future for better recognition this algorithm might be updated by data of depth camera.

Each avatar has specific scalable regions by which it can be scaled as a user's body as illustrated Figure 2. Arms, legs, trunk, and neck have a tube shape, regardless of whether it is a muscular or a fat body so it's easy to the scale avatar by applying scalable rings on those regions was described above.

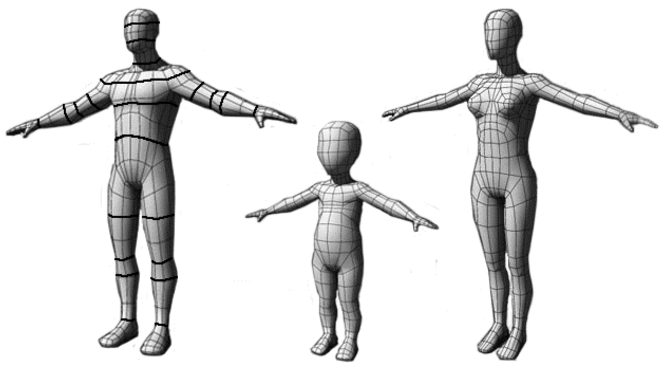

Fig. 2. Types of avatar: Male, Kid, and Female. On the male avatars showed scalable rings.

Instead of using information from anatomy to measure the user's body it's better to look to craft tailoring in the places where the seamstress measures the body for clothing [8]. The same method is used to take measurements from the body after it was detected and calibrated by Kinect.

B. *Head (Face) creation*

The hardest task of head creation is the reconstruction of facial geometry. Making the same head size and applying hairstyle is not so hard as the first task. Even if there are some casualties with hairstyle, the user can usually find any appropriate hairstyle in the database.

Researchers at the Computer Graphics Group in Erlangen have presented a system for generating 3D face avatars using the Kinect. A generic face model is fitted to the depth map obtained by the Kinect. The resulting 3D face model is finally textured using the captured RGB image. The proposed algorithm combines the advantages of robust non-rigid registration and fitting of a morphable face model. It is possible to obtain a high quality reconstruction of the facial geometry and texture along with one-to-one correspondences with generic face model as illustrated Figure 3. [9]

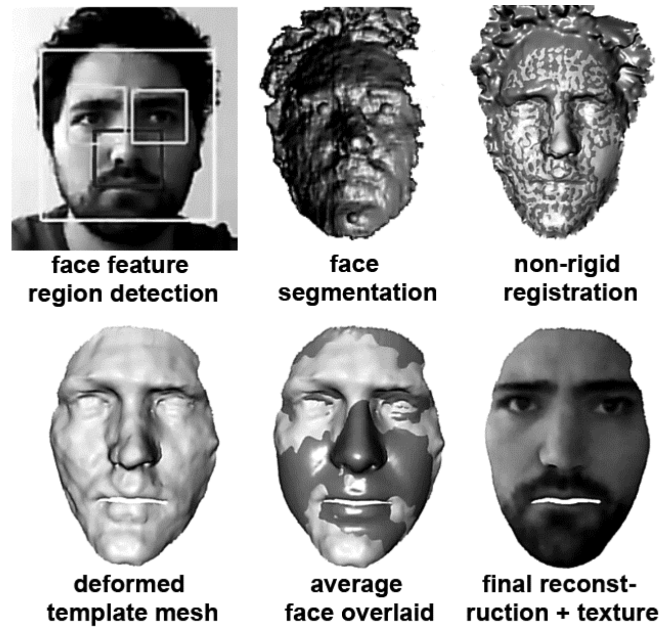

Fig. 3. Step by step pipeline of 3D face scanning.[9]

III. AVATAR ANIMATION

Games usually need predefined animation instead of real time manipulation of avatars and it will be better if the avatar and user moves alike. Microsoft Kinect SDK and OpenNI [11] support skeleton tracking so it's easy to create animations of body movements for the avatar.

When all users have their own avatars, they could be used in online conversations, instead of creating video streams between users like we do today in skype, ooVoo[2] and similar programs. People will send avatar once, so it will be saved on user's local machines with whom one usually talk and afterwards send only needed coordinates which are usually simple text or xml file. This kind of system has a big advantage of use in meetings, when there are more than two people making conversation with each other. Video windows of current systems will be replaced by virtual 3D rooms with real time animated avatars of participants. In addition such future systems could be updated with 3D sound to make them more realistic.

Like the creation of avatars, animation is also divided into two parts:

A. *Avatar's Body Movement*

There are several tools to create movement animation of avatars. The simplest way is to use programs like Brekel Kinect[3], which allow you to capture 3D objects and export them to disk for use in 3D packages. One can also do skeleton tracking which can be streamed into Autodesk's MotionBuilder in real time, or exported as BVH[4] files.

[2] ooVoo brings easy video chat to all platforms, free software, http://www.oovoo.com/home.aspx

[3] Brekel Kinect is an application using a Microsoft Kinect, and PrimeSense's OpenNI and NITE. http://www.brekel.com

[4] BVH is a standard format of biped (humanoid skeleton) animations which is used in all nowadays 3D redactors and game development libraries.

For real time and advanced animations Microsoft Kinect SDK and OpenNI libraries could be used. Both these libraries can track 20 joins of key points of human's skeleton as

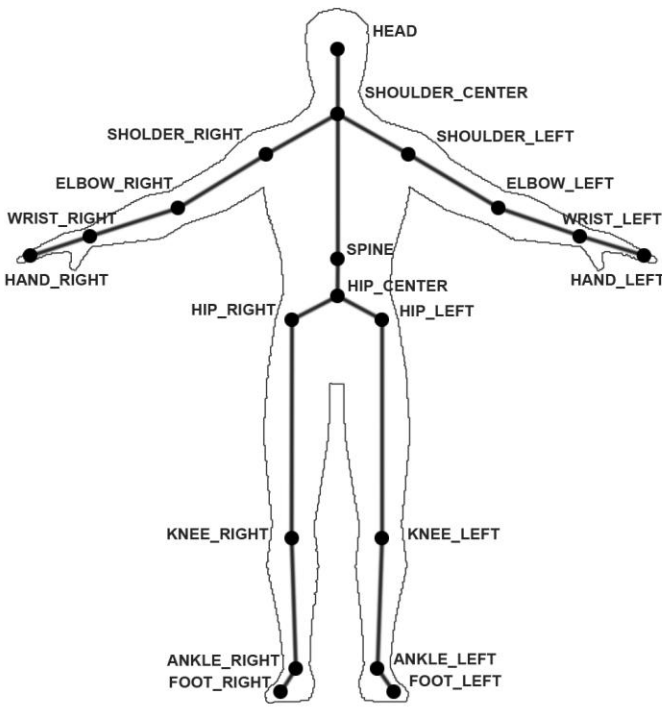

Fig. 4. Skeleton structure with 20 joints. Same as in Microsoft Kinect SDK as in OpenNI.

illustrated Figure 4.

B. Facial Animation

Two types of facial animation are needed. The first is for predefined animation which will be used afterwards in applications like games. This type of animation should have a special format that can be saved in a database. The second is for real time animations, the problem in this type is the performance of local machines, in other words it should be performance-based.

Such types of animation already exist. Hao Li and his team presented realtime performance-based facial animation [12]. This system enables any user to control the facial expressions of a digital avatar in realtime. The user is recorded in a natural environment using a non-intrusive, commercially available 3D sensor Kinect. To effectively map low-quality 2D images and 3D depth maps to realistic facial expressions, they introduce a novel face tracking algorithm that combines geometry and texture registration with pre-recorded animation priors in a single optimization. Formulated as a maximum a posteriori estimation in a reduced parameter space, this method implicitly exploits temporal coherence to stabilize the tracking. They demonstrate that accurate 3D facial dynamics can be reconstructed in realtime without the use of face markers, intrusive lighting, or complex scanning hardware.

IV. Conclusion

In conclusion we can say that the ability of making and animating 3D avatars will become in the near future necessary. With the increasing technological capabilities, the field of 3D avatar usage will be unlimited. In education – distance learning, where students will be able to see their teacher in real time, in three dimensions as a real person in class and vice versa. In market – virtual fitting rooms which could be used in home and shops, allowing customers to try on clothes. Online communications – usage of systems with 3D facial avatars instead of video streams will reduce the load on the Internet connection. Games, 3D social networks and other web and desktop applications – the virtual world will become more realistic with 3D avatars of real humans. This paper gives the guidelines to create 3D human avatar using Kinect.

Acknowledgment

This work is supplied by the EU Erasmus Mundus Action 2 TARGET program (EM A2) for fundamental exchange research.

References

[1] "Kinect for Xbox 360 (Kinect)." from http://www.xbox.com/en-US/Kinect.
[2] Valli, A. (2005). "Notes on Natural Interaction."
[3] Calle, J., P. Martínez, et al. (2009). "Towards the Achievement of Natural Interaction." Engineering the User Interface, Springer London: 1-19.
[4] Del Bimbo, A. (2008). "Special issue on natural interaction." Multimedia Tools and Applications 38(3): 293-294.
[5] "Microsoft Kinect SDK." from http://kinectforwindows.org.
[6] Jisha C.T., Suman M.I. "Language independent voice-based gender indentification system", A2CWiC 2010, September 16-17, 2010, India
[7] Liangliang C., Mert D., Yun F. Thomas S.H. (2008)"Gender recognition from body" MM'08, October 26-31, Vancouver, British Columbia, Canada.
[8] "How to get our body measurements" http://www.dummies.com/how-to/content/how-to-get-your-body-measurements.html
[9] "3D face scanning using the kinect" http://kinecthacks.net/3d-face-scanning-using-the-kinect
[10] "Realtime Performance-Based Facial Animation" http://kinecthacks.net/realtime-performance-based-facial-animation
[11] "OpenNI" http://www.primesense.com
[12] Thibaut W., Sofien B., Hao L., Mark P. (2011) ACM Transactions on Graphics, Proceedings of the 38th ACM SIGGRAPH Conference and Exhibition.