Real-time 3D Virtual Dressing Based on Users' Skeletons

Ting Liu, LingZhi Li, XiWen Zhang*
Digital Media Department
Beijing Language and Culture University
*Corresponding Author
Beijing, China

Abstract—**3D virtual dressing allows users to try-on clothes' models without physically putting on and off as well as to attain 3D virtual fitting superior to 2D virtual dressing. It includes three components in a virtual 3D space: models of users and clothes, and their fitting. This paper presents real-time 3D virtual dressing based on users' skeletons. Users' skeletons are extracted and tracked in real-time to drive transformation and fitting of clothes' models. After clothes' models are rigged and weighted based on users' skeletons, skinned clothes' models are transformed and fitted to users with dynamic physical effects. They match users' bodies and actions, resulting in synchronous transformation: scale, rotation, and deformation. A software prototype is developed in Unity 5.2 and Microsoft Kinect for Windows V2. The prototype allows users to select different clothes' models and background images using motion-sensing gestures to view various collocations after background removal. Some users and clothes are tested, and satisfactory results are attained.**

Keywords-3D virtual dressing; real-time; motion-sensing; users' skeletons; Microsoft Kinect V2

I. INTRODUCTION

It is a common behavior for people to try clothes before shopping in life. Several ways can perform this. In physical stores, customers can try physical clothes and engage in real interaction and fitting, but have to put on and off them, thus wasting valuable time and effort. It is difficult and even impossible for some people to try some special clothes. In online stores, customers can explore freely every clothes in digital images or 3D models, but cannot attain effects of trying on themselves. In 2D virtual dressing, customers can try freely every clothes' images, but not feel 3D fitting because of digital images from customers and clothes. 3D virtual dressing can afford 3D fitting between models of customers and clothes because both are 3D models. Customers in this way can freely try every clothes' models in 3D spaces, even set size, color and texture they preferred.

There is few work on real-time 3D virtual dressing because of difficulty in real time modeling and fitting between human bodies and clothes in 3D spaces [1]. In fact, there are two ways of fitting clothes' models to users' models in 3D space. One is based on collision detection between surface models from users and clothes. It achieves higher fitting accuracy leading to a more realistic effect. However, it requires larger computation and is quite limited by computer performance. The other is based on skeletons of users. It applies real-time skeletons to fitting. To fitting in real time and usual computation, this paper uses users' skeletons to transform and fit clothes' 3D models.

The rest of the paper is organized as follows. In section 2, related work is surveyed. Section 3 introduces users' skeletons are extracted and tracked in real-time. Section 4 is to create 3D clothes' models with skinned mesh based on users' skeletons. Section 5 is to fit 3D skinned clothes' models to users' skeletons. Section 6 is to develop a software prototype presenting fitting, background switching, and height estimation. Section 7 gives some test of the prototype. Finally, conclusion and future work is discussed.

II. RELATED WORK

Existing virtual dressing approaches can be categorized into 2D, partial 3D, and full 3D solutions. A typical 2D version enables users to try on clothes in digital image. Users can check color and style virtually, but cannot feel real fitting. Partial 3D solution involves clothes' 3D models and 2D images of users' bodies. It is more efficient in computing, practical in realistic visualization [2]. Full 3D solution superimposes clothes on users both in 3D models. In this way, users can check real fitting in 3D spaces, besides color and style from different views such as front, back and side.

The main techniques involved in 3D virtual dressing technology are modeling and fitting of users and clothes, and size recommendation [3]. It is usual to model users by employing scanning devices, such as webcams, phone cameras, and depth camera. Microsoft Kinect V2 can detect users and access color, depth, and skeleton stream data [4]. Clothes with physics properties can be created using Marvelous Designer, Clo3D, Autodesk 3ds Max and Maya [5]. Approaches based on database use existing garments to recommend size. Others use measurements taken by customers or detected by body scanners, or combination of both.

This paper presents the real-time fitting between users' 3D skeletons detected by Kinect V2 and clothes' skinned 3D models. Clothes' 3D models are rigged to users' skeletons and allocated with different weights for realistic physical effects. Clothes' skinned 3D models are transformed and fitted in real time because users' skeletons are used to drive clothes' models.

978-1-5386-1107-4/17/$31.00 ©2017 IEEE

III. EXTRACTING OF USERS' SKELETONS

Kinect V2 can detect and track users' bodies more quickly and accurately. Thus, users' skeletons are obtained by Kinect V2, and are used to drive the transforming and fitting of clothes' 3D models.

Kinect V2 can acquire both depth and color information and process them in real time. Using infrared camera, Kinect can recognize up to six users and track up to two of them in the field of view of the sensor. By using image segmentation technology [6], it can extract human bodies from complex environment. Then it labels possible image regions as bodies. Thus, depth information is transferred into skeleton information that generates 25 skeleton joints, as shown in Fig. 1. These joint positions are the data consumed as position and pose, and they are used for many purposes, such as gesture detection and driving avatar models to move.

The skeletal tracking algorithm of Kinect V2 is created using a large training data set with over 500k frames distributed over many video arrays [7]. The built-in skeletal tracker returns the 3D coordinates of 25 body joints in terms of pixel locations for x and y coordinates and meters for the depth [1], allowing Kinect to recognize users and follow their actions.

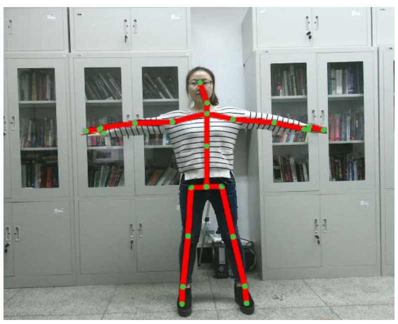


Figure 1. User's skeleton from Kinect V2.

IV. RIGGING AND WEIGHTING OF CLOTHES' 3D MODELS

Clothes' 3D simulation holds the key to virtual try-on experience, especially 3D models. Real-time clothes physical effects and perfect fitting to users all require well-functioned modeling techniques. A good model is supposed to have realistic clothes physical effects and be controlled by skeleton and body properly. To fitting in real time, clothes' 3D models are skinned by rigging skeletons and allocating weights.

A. Rigging skeletons of clothes' 3D models

Users' skeletons are used to control the movement and deformation of vertexes in skinned meshes of clothes. Skeletons can be understood as coordinate spaces which is arranged in a hierarchical structure. A body joint is the origin as well as the center of scale and rotation of a bone' s coordinate space. With a hierarchical structure of bones, the movement of bones can be controlled through their parent bones.

A skeleton is supposed to have at least 17 bones to produce a rather good effect. Clothes' skeletons are constructed carefully and named in a way that reflects the body parts they represent. The skeleton in Fig. 2 is constructed with head, neck, shoulder, arm, hip, leg joints etc. And the hip center joint is the root of skeleton.

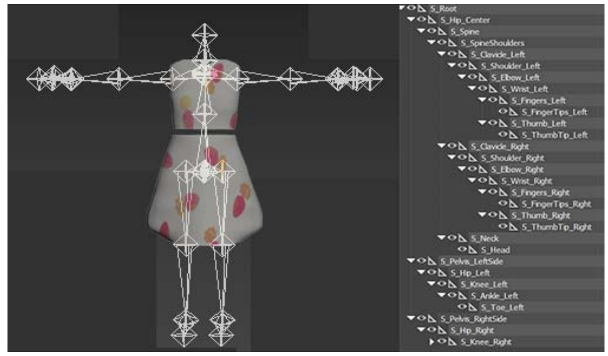


Figure 2. Clothes' skeletons are rigged.

B. Weight allocation of clothes' skeletons

Weight allocation is used to bind bones with vertexes of model mesh. Each vertex movement will be affected by a set of bones to different degrees. Each bone movement will affect a series of vertexes in different weight. The allocation of different weights is based on two principles primarily. First, the vertex weights are allocated by joint-to-mesh distance. The closer vertexes of mesh get to the key bone point, the more they are affected by the motion. For example, while the knee joint moves, the vertexes in vicinity of knee joints changes more dramatically compared with other vertexes which is closer to upper or lower leg. Second, a more flexible bone controls a wider range of vertexes and these vertexes are allocated with more weights. A representative instance is that the number of vertexes controlled by knee joints is larger than that of shoulder joints and that the allocated weight ratio of vertexes controlled by knee joints is higher. As the movement of a skirt while lifting legs is more strenuous than that of a shoulder while carrying arms. Therefore, the weights of different vertexes are allocated appropriately according to human physical activities, after which the overall clothes' physical effect are achieved. The weight distribution for a skirt is shown as Fig. 3. After finishing above steps, a skinned 3D model is created and then mapped UV textures to add more details of clothes.

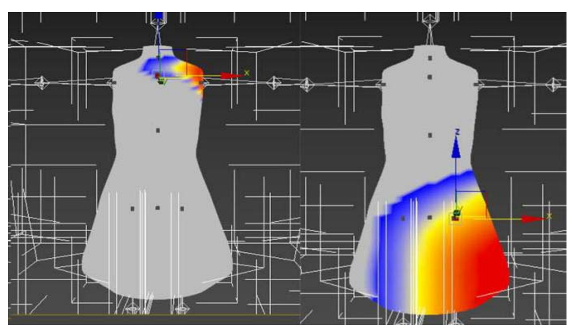


Figure 3. Weight Allocation based on a skeleton of a user.

V. FITTING OF CLOTHES' SKINNED 3D MODELS

Clothes' skinned 3D models are transformed for movement, scale, and rotation to fit to users. The transformation of clothes' models is driven by users' skeletons detected by Kinect. Positions, rotations, and scales of several pivotal body bones, such as shoulders, hips, and legs, are assigned to corresponding bones in the bone structure of clothes' models. And movements of users' skeletons detected by Kinect are mapped onto the movements of clothes' models in a certain proportion. Thus, clothes' models bound to the bones of users are to actualize the effect that clothes' models fit to and respond users' skeleton.

A. Scaling of clothes models

First of all, several coordinates of main body joints are recorded, and the height and width of users are calculated.

Height = ShoulderCenter.position - posHipCenter.position
Width = rightUpperArm.position - leftUpperArm.position

Then, the height and width of a user's body are used as userHeight and userWidth to compute scale factors on 3 dimensions.

heightScale = userHeight / modelHeight
widthScale = userWidth / modelWidth
depthScale = (heightScale + widthScale) / 2

Finally, the local Scale variable on each connected section between two bone joints is applied, and the positions of body joints which affects the scales of clothes' models directly are adjusted. At this point, clothes' models can fit different users' body shapes. And also, they can scale as the user moves forward or backward, which is regulated by depth scale.

localScale = (modelBodyScale.x * widthScale,
modelBodyScale.y * heightScale,
modelBodyScale.z * depthScale)

B. Rotation of clothes models

When Kinect detects that the rotation of a user's body part occurs, the corresponding bone structure rotates the same angle. Such rotation directly results in the whole clothes models' synchronous rotating effect. The detailed implementation is shown in the following Fig. 4.

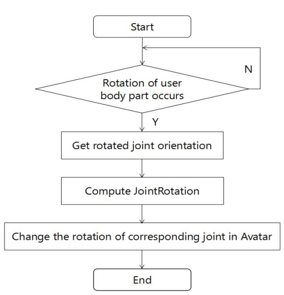


Figure 4. Rotation of clothes' models

Especially when a user turns his back, skeletal positions of two sides of the user' s body joints are exchanged. Being able to detect the rotation of 180 degrees becomes an advantage, making it possible to view the back of clothes' models.

VI. SOFTWARE DEVELOPMENT

Kinect V2 has better performance on motion-sensing than one in V1. It is only a consumer-level depth camera which is low cost and easy to use. It can also be used as an input device to capture users' figure information in Unity. Fitting is developed in Unity, which is a popular game engine. An effective and efficient software prototype for real-time 3D virtual dressing is developed in Unity 5.2 with Kinect V2.

With the help of the Avatar System function in Unity, the prototype realizes the real-time automatic matching of models from clothes and users which includes synchronous scale, rotation and deformation. In addition, the prototype brings about an interactive mode using motion-sensing gestures to choose different clothes' models and background images. It also provides a rough measurement of users' height and vital statistics such as bust, waist, and hip.

A. Avatar system in Unity

Avatar system in Unity can serve as a medium for matching a user's skeletons with clothes' bones. After a clothes' model file with skeleton data is imported into Unity, Unity's Mecanim animation system attempts to match up existing bone structure of a clothes' model to the Avatar bone structure [8]. In general, the model-avatar can be created from a clothes' model automatically by analyzing the connections between these bones of the two. Clothes' bone transformations should be reset manually in avatar configuration, while sometimes the skeleton of clothes cannot be set to a user' s body joints correctly.

And then, when a clothes' model is selected and loaded in Unity, it is dynamically added with the Avatar Controller component, which obtains the movements including positions as well as rotations of a user's body bones in Kinect and utilizes them to control itself. Similarly, the loaded clothes' model is added with the Avatar Scaler component, which however obtains the scale of a user's body bones.

Therefore, when each of clothes' bones in avatar bone structure corresponds to the body bones in Kinect, 3D models of clothes can be driven by a human body, achieving synchronous movement between human body actions and clothes models as shown in Fig. 5. And also, clothes' models can be displayed in 3 views: front, side and back, which is presented in Fig. 6.

Figure 5. Different kinds of clothes models fit users

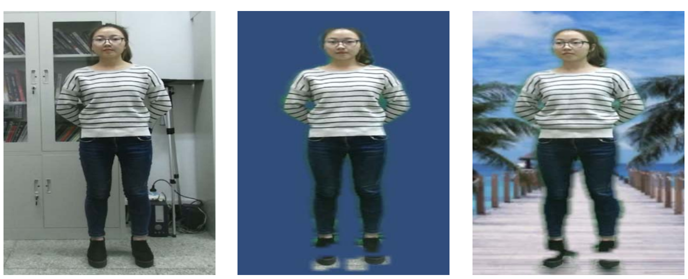
Figure 6. Cloth model displays in front, side and back views

B. *Extracting of users' regions*

It is crucial to extract users' regions from live background such as digital video for better virtual try-on user experience. Background removal helps to avoid interference of environment, and contributes to change different background scenes. A basic idea of it is to use depth data delivered by Kinect to filter background pixels and isolate users' regions. However, background removal based on video matting is not real time and not stable enough for practical applications [9]. A method to solve this problem is to convert video frames into GUI textures as the second background layer. The detailed procedures are given as follow.

Depth data and color data are first accessed, and then depth frames are mapped into color space. Different Alpha value is set for foreground and background pixels. Alpha values of background pixels are set as 0 while foreground as 255. Lastly, various image processing techniques are used to improve the stability and accuracy of user mask [10]. In this way, background is removed. However, the removal performance is poor when it comes to distinguish a user' s feet and ground, for the depth difference of them is small, as shown in Fig. 7.b.

Furthermore, background removal has many possible uses, one of which is to composite users' regions into other scenes. Digital images of different try-on scenes are applied to provide immersive environments. Users could check if their apparels are suit for certain occasions, as shown in Fig. 7.c.



(a) Video stream with a user. (b) A user' s region is extracted. (c) A user' s region with another background.

Figure 7. A user' s region is extracted.

C. *Switching clothes' models using gestures*

Motion-sensing gestures help users to get rid of contact devices such as the mouse and keyboard and to choose the clothes' models and background images conveniently to view the collocation effect. The Kinect SDK identifies whether a gesture is completed by tracking the location and time duration of body joints, the accuracy is quite high in Unity 5.2. The Kinect V2 defines T-pose, Swipe, Raise hand, Wave, Squat and other 20 gestures. This prototype mainly uses Swipe and Raise Hand gestures to change clothes' models and hats' textures. Several gestures used in our virtual dressing prototype are described as below.

1) Swipe Left gesture. If a user uses right hand to swipe left, SwipeLeft gesture is detected. Then the prototype loads next clothes' model and displays it.

2) Swipe Right gesture. Comparatively, swiping right with left hand will be defined as SwipeRight gesture, which correspond to back to previous clothes' model.

3) Swipe Up gesture. Using left or right hand to swipe up is SwipeUp gesture. SwipeUp gesture can control the change of the background board textures, providing different occasions for users to decide whether the clothes are suitable.

4) Raise Right Hand gesture. Raising right hand over head is RaiseRightHand gesture, the prototype will get into the hat model category. And under this category, SwipeLeft or SwipeRight gestures are used for choosing next or previous texture of the hat model.

5) Raise Left Hand gesture. When the prototype comes back to clothes category, using SwipeLeft or SwipeRight gestures can reload and change clothes' models in the same way.

D. *Height estimator*

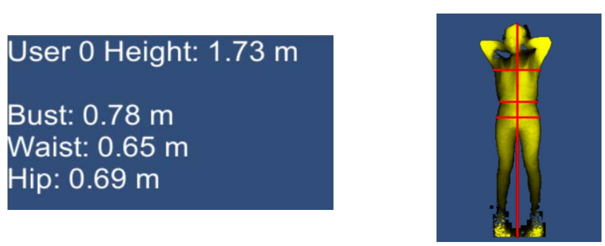
Automatic detection of 3D vital statics of users in this prototype can be used as a reference to select the size of clothes. A user of the depth data and skeleton data delivered by Kinect are to estimate height and other body measurements, including bust, waist, and hip, as shown in Fig. 8. Users could utilize these estimated data to make size decision. The algorithm of height estimator is explained as following.

Step1. Get spine joint position in World Coordinate and map it into Depth Coordinate.

Step2. Find the top of head position and the bottom position of body in depth image by traversal.

Step3. Map above two vectors into skeleton joints in World Coordinate

Step4. Height equals to the subtraction of two joints position.



(a)Estimated data (b)Depth data

Figure 8. Height Estimator

Other body slice measurements including bust, waist and hip are calculated in another way because the original measurement is based on width of body slice shown in the depth image. A user of 3D model can be treated as a simple cylinder, thus the original width of body slice measured have to be doubled and then added different empirical values. Such as for hip and bust, the empirical value is 0.1 meter while for waist, the empirical value is 0.05 meter. These values are tested based on a group of statistical samples. The measurements are shown in Fig. 8.

VII. EXPERIMENT RESULTS

The virtual dressing prototype has been basically accomplished, users can achieve motion-sensing control of choosing different kinds of clothes' models, hats and background images through Swipe Left or Right, Raise Left or Right Hand, Swipe up gestures.

(a) Red dress suit in a stage.

(b) Sitting with camouflage shirt on a road.

(c) Colorful stripe dress suit in front of the triumphal arch.

Figure 9. Visual dressing is tested.

The user can view the real-time collocation effect with the change of hats' textures and clothes' models. And with the synchronous movement between a user and clothes' models, the user can rotate different angles to check the displaying of clothes' models. At the same time, the software's interface provides some more detailed information such as body height, bust, waist, hip and gesture detection prompts. The integral dressing effect is shown in the Fig. 9.

VIII. CONCLUSION AND FUTURE WORK

This work uses user extraction from Kinect video stream and avatar system for skeletal tracking to align the clothes' models with users. And a virtual dressing software prototype is developed allowing clothes' 3D models to overlay users and were convenient to view in front, side and back perspectives. The clothes' model can be relatively fit for a user' s body shape immediately. And optional background images provide users with immersive environments for more practical uses.

There are still room for improvement to make such prototype widely applied in reality. It is possible to improve the accuracy of estimated user data with further processing, providing more valuable references. Furthermore, improving a clothes modeling approaches that achieves rapid reconstruction based on real clothes is also of great use.

ACKNOWLEDGMENT

This work is supported in part by Young and Middle-aged Academic Backbone Support Program of Beijing Language and Culture University and WuTong innovation platform of Beijing Language and Culture University (supported by "the Fundamental Research Funds for the Central Universities") (No.16PT04).

REFERENCES

[1] Isikdogan Furkan, Gokcehan Kara. A real time virtual dressing room application using Kinect[J]. CMPE537 Computer Vision Course Project, 2012(1):1-4.

[2] CHEN Guilin. Digital clothing design: three-dimensional simulation of clothing technology[M]. Beijing: China Textile&Apparel Press, 2012. (In Chinese)

[3] Chen Qingqing. The Research and Relization of Three-dimensional Virtual Garment Stitching and Cloth Simulation[D]. Xiamen: Xiamen University, 2009. (In Chinese)

[4] Kinect v2 Tips, Tricks and Examples. https://rfilkov.com/2015/01/25/Kinect-v2-tips-tricks-examples/.

[5] Mo Y I, Lee S G, Chang W. Method for converting 2d image into pseudo 3d image and user-adapted total coordination method in use artificial intelligence, and service business method[P]. U.S. Patent Application:10/583, 160,2005-12-5.

[6] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. 2011. Real-time human pose recognition in parts from single depth images. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, Washington, DC, USA, 1297-1304. DOI=http://dx.doi.org/10.1109/CVPR.2011.5995316.

[7] HE Bei, WANG Guijin, LIN Xinggang. Quick matting for videos based on depth images of the Kinect[J]. Journal of Tsinghua University (Science and Technology), 2012(04):561-565, 570. (In Chinese)

[8] Unity official website. https://unity3d.com/cn/.

[9] Giovanni S, Choi Y C, Huang J, et al. Virtual try-on using kinect and HD camera[C]. International Conference on Motion in Games, Springer, Berlin, Heidelberg. 2012: 55-65.

[10] HU Wanyue, LI Yanmei, WANG Yingmei, WU Xiaona. Development Status and Prospects of Virtual Fitting[J]. Journal of Shanghai University of Engineering Science, 2014(02):162-165. (In Chinese).