Online Virtual Try-On System Based On Personalized Avatar

Yangyiqiao GAO
44171531-2

Master of Engineering

Supervisor: Prof. Jiro Tanaka
Graduate School of Information, Production and Systems
Waseda University

July 2019

Abstract

With the development of e-commerce, nowadays lots of people like to buy clothes online. To increase users' satisfaction degree and provide a better user experience, many companies and research institutions aim to develop better virtual try-on system. In such system, you can try on clothes before paying for them online. It shows you accurate fit and look.

We intend to improve the current Model-based system by giving the system the feature of 'sense of substitution' and 'support for movement', 'sense of substitution' means users feel that they are wearing the clothes themselves.

Our goal is to propose an automatic, dynamic and personalized virtual try-on system. Once we deploy the system to a server, users can use it through smartphones or computers. After uploading a video of self, the user can operate a personalized avatar to try on different clothes.

There are mainly 3 points of this paper: 1. Designed a simplified framework to implement virtual try-on system. 2. Improved the performance of current model-based try-on system, decreased the time and increased the quality of appearance(result). 3. First time to make an interactive model in virtual try-on system.

Keywords: Virtual try-on, Online, Personalized, Model-based, Interactive

Acknowledgements

At the beginning, I am very glad to express my sincere gratitude and appreciation to my supervisor Prof.Jiro Tanaka, a very respectable and responsible mentor with rigorous academic attitudes and creative ideas. He has given me lots of help and encouragement, and is always very patient for guiding me to solve problems. Thanks a lot to him for the many valuable and earnest advices during every stage of preparation of this thesis.

Besides, it is grateful to all my dear labmates. They discussed a lot with me and always give me lots of inspiration. Thanks so much for their many helps in experiments so that our experiments could be performed successfully and the good results could be get.

Last but not least, I would like to thank my parents and all love ones. For my parents, they give me financial support and warm mental support throughout my master study life. I will keep going on and making progress to return all those people.

Contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Introduction

Virtual garments try-on system is a system in which users can wear different garments without actually wearing the clothes. Such system could improve the shopping experience by assisting users to make purchase decisions.

If you search the benefits of virtual garments try-on system in google, below is the top-rated answers:

- You can try on clothes before buying(payment) them online. It shows you accurate fit and look.
- As the users already know how the clothes will look and fit there is very low chance of return.
- It keeps the customer engaged for a longer time, trying out different clothes on a look-a-like 3D avatar. This has more chances of conversion into a sale.
- It saves you from the long queues at the malls or departmental stores.
- It saves the retailer from damaged or soiled garments due to heavy try on by customers.

In general, it could increase users' satisfaction.

From the view of usage scenario, there are some kinds of virtual try-on system in the market, while this three are the most popular: image-base system, live AR and model-based system.

Image-based system is the most aged one among above systems. In such system, users can warp a clothes' image onto a user's photo to make it look like the user is wearing the clothes. An image warping technique is applied to map the apparel image onto the individualized body. While its result is 2D and unrealistic. Such techniques also do not allow to see the textured images from arbitrary viewing angles.

For better try-on result, live-AR system is published and becoming more and more popular in some big shopping mall or clothes shop. In such system, images captured by video cameras in a virtual fitting room are augmented with the images of garments in real-time. users can see the augmented result of themselves from a 'mirror'. 'Mirror' here is actually a screen full of sensors. While this kind of system requires a fitting room equipped with several 3D cameras, which affects the convenience of users. And mapping cloth textures on a moving model can still be a complicated task. In general, users can only use this in some big shopping mall.

To combine the advantages of above 2 systems, here comes the model-based system. In such system, users can check a 3D human-shape model dressed with 3D clothes model. Combining augmented reality technologies, users can get a good try-on experience. While for now, this kind of system is not mature enough. For example, current such systems in the market do not provide the support for personalized model, motions and some other features.Which means, you may not feel that it is you that who is trying-on the virtual clothes.

In this study, we give users a better experience by improving current model-based system. We provide a 3D personalized avatar of the user. User can control it by scaling, moving and rotating it. Also the avatar will be able to do some motions. To make it more realistic, user can put it into real environment, like bedroom, office or somewhere else with AR technologies.

Sometimes users may want to see what would it look like when it is put on some specific environment, like dinning hall or meeting room. To help user check the result in these environment, we also enable user to put the avatar into some predefined virtual environment.

To use our system, users just need to record a video of themselves and upload it to our server, and the user only need to do this once. User can check the result both in mobile phone or PC.

1.2 Organization of the thesis

The rest of the thesis is organized as follows: Chapter 2 introduces the background about the thesis and also the related works in this field. Chapter 3 will tell the research goal and also the approach will be told briefly. Chapter 4 is the system design part, where the design concept and ideas will be introduced and the algorithm design will also be told. Chapter 5 will be the system implementation part where the detailed environment and implementation will be talked. Chapter 6 will introduce the related work. Chapter 7 will be about the experiments, we will talk about the performance of our approach and the comparison of different approached will also be done. The last part, Chapter 8, will be conclusion and future work part, where we will conclude the previous content and talk about the future possibilities.

Chapter 2

Background

2.1 Virtual clothes try-on system

Virtual clothes try-on system mainly refers to a system in which shoppers can try on different clothes or pants virtually, being able to check the style, size, fit or some other features, without a physical one in hand. It can be understood as an equivalent of an in-store changing room.

Form the view of usage mode, nowadays there are mainly 2 kinds of virtual try-on applications: online and offline. Online applications mean users can access it online, by mobile phone or computer, without going to a real store. To use offline applications, users need to be in front of a big device, which contains at least a body scanner, a screen and some other sensors. This kind of device is often available in some big shopping mall.

From the very beginning, virtual try-on system is mainly implemented with image warping technique. Users can scale or rotate a clothes' image onto a photo to make it look like the user is wearing the clothes. While it is has many limitations, for example, users cannot check its fit. So here comes the 3D solutions. Nowadays, the mainstream method to create online 3D virtual try-on result is to scan a real human model wearing different clothes with 3D camera, then export it to the server. Users can adjust the model according to their own body figures. Replacing the face with user's own is still a challenge.

Compared to online applications, nowadays offline virtual try-on device can give a better result from the view of realistic degree. With a lot of sensors, a 3D body model can be built in real-time. Users can see themselves wearing different clothes from a screen, just like a mirror. While there are also some flaws, to make the system real-time, the clothes models in the screen are always 2D. So the fit is not that good.

2.2 3D model generation

As we all know, 3D models can show more information comparing to 2D images. 3D is an important concept in virtual try-on system. 2 ways are widely used to generate a 3D body. First is using the body scanner, like Kinect, to scan the user from 360 degrees. Second is building a 3D model before, then adjust it to a customized shape according to users' figures, like measurements and shape information. These figures are collected manually. Nowadays with the development of machine learning, there appears some more efficient methods to generate 3D models. We can generate 3D body from a picture [1]. Not only the body, we can also generate a high-poly 3D head from a head portrait [2]. To get a more accurate result, in our research, we used Matthew's work [3] here, which could generate a low-poly 3D avatar from a video.

2.3 Augmented reality

Augmented Reality It is an emerging technology developed on the basis of virtual reality, which can superimpose computer-generated virtual scenes on real scenes seen by users.

Augmented reality technology is a technology that increases the user's perception of the real world through the information provided by the computer system, and superimposes the computer-generated virtual object, scene or system prompt information into the real scene, thereby realizing the "enhancement" of the reality. It superimposes computer-generated virtual objects or non-geometric information about real objects onto real-world scenes, realizing enhancements to the real world. At the same time, the interaction is more natural because the connection to the real world is not cut off.

Chapter 3

Research Goal and Approach

3.1 Goal

The fundamental goal of our research is to **improve the current model-based system by assigning it the feature of interactive personalized avatar, which is to make an online virtual try-on system based on personalized avatar.**

Comparing with current model-based system, live AR system has two important features, immersion and support for movement. To combine the advantages of live AR system and model-based system, we need to assign our system such features.

We intent to build an automatic, dynamic and personalized virtual try-on system. Once we deploy the system to a server, users can use it through smartphones or computers. After uploading a video of self, the user can operate a personalized avatar to try on different clothes.

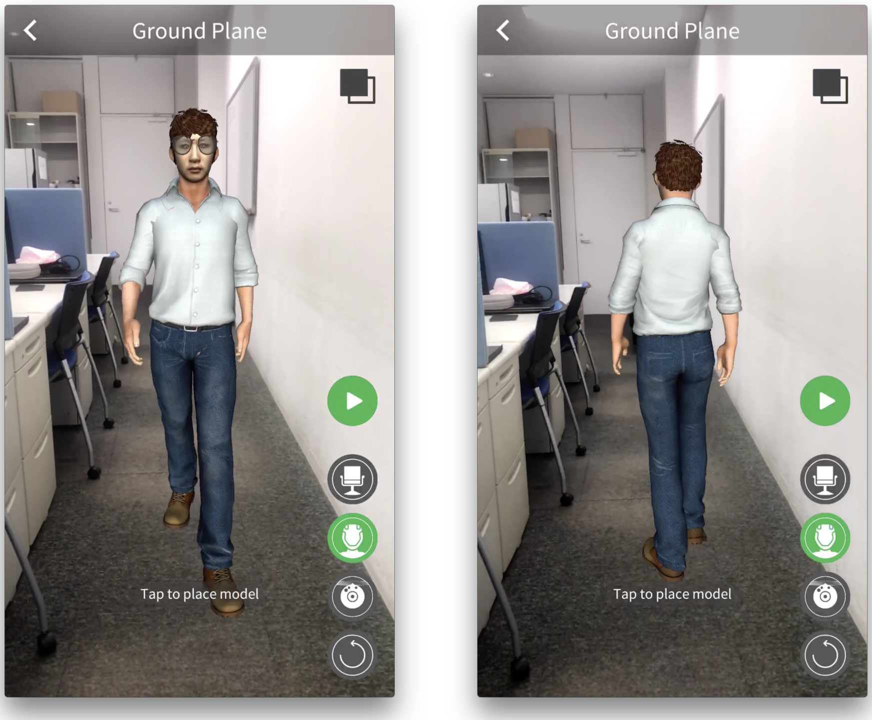
3.2 Use Case

3.2.1 Use Case 1

A user is browsing an e-commerce app in his mobile phone. He chose one and checked some images of it. Although the clothes looks good when it is on a model, he still want to try it on by himself to check the appearance.

Then he click the 'try-on' button, which is connected to our system, and here appears an personalized avatar which has the same body figures and the same face, he put the 3D virtual avatar onto ground of his lab, with augmented reality technology, and check it from different view.

It looks good and the size also fits well. Now he want to check what does it look like when the avatar is moving. So he activates its movement, and also check the result from different view.



(a) Walking view (b) Walking view from the back

Fig. 3.1 (a) is front view of the avatar when walking, (b) is the back view when walking.

3.2.2 Use Case 2

A user is browsing an e-commerce site in his laptop. He find a good one and want to try in on by himself. What is more, he want to check what does it look like when he is dressing it in different places. So he click the 'try-on' button and see a 3D avatar of himself. Then he click the ’Change Environment’ button, as the figure shows, he can check the situation of dinning hall.

Fig. 3.2 Use case 2

3.3 Approach

To achieve the goal, we built a double-skeleton based framework to implement it. In this framework, there are actually two process running asynchronously: one is processing the garments model, one is processing the body model.

In order to generate a personalized avatar, we firstly generate a customized body [3] based on users’ videos. While this 3D body is very low-poly, the face and the texture are very blurred. We cannot adopt this model directly. To make it more realistic and high poly, we firstly make the surface more smooth, then we re-attach the predefined skin texture to it. Now that we have got a high-poly body without a high-poly head. So we generate a 3D face [2] based on users’ head portraits. Current trained model in [2] does not support generating a hair model since hair model is harder to build, so we need to assign the 3D head a predefined hair model also. We designed some common hair model, when using, users can choose a best hair model for themselves. Then we attach the 3D head to the 3D body, so that we made a high-poly personalized avatar.

While we cannot control this 3D model yet. Now it is just a static model. It is very hard to dress a 3D clothes to a static 3D body model. In previous research [4], researchers do this manually. Since it is a very hard computer graphic problem that how to fit a garment model to a 3d body model.

To make the whole process automatically and faster, we proposed the double-skeleton based method. Its main idea is we generate two skeletons with the same structure separately for 3D clothes and 3D body, then we adjust the pose and the position of this two models with the help of the skeleton to achieve the goal of fitting. The skeleton is predefined and universal. It means that before fitting clothes to a 3D body, we fit the clothes to a skeleton first.

To make that firstly we use OpenPose[5] to detect the body model's key points. In our research we detected chin, neck, wrists, spine root and ankles. With the coordinates of these key points, we can align the predefined skeleton to the model easily. Just by moving the corresponding parts of the skeleton to the geometric location we can finish the step of rigging [1]. We do the same thing with the clothes. Firstly we detect the key points of clothes using Fashion Landmark [6], then we fit it to the skeleton with the same steps of what we have done with the 3D body.

So now we get a rigged clothes model and body model. With that, work becomes much easier. I just need to align these two models to the same pose and the same position. And this is the double-skeleton based method.

Fitting is the most important step. After fitting, we export it to clients, PC or smartphones. Since the personalized 3D model contains a lot of private information, so we will not save it in server. This avatar is dynamic and interactive. User can put it onto real ground through the camera of a mobile phone. We use Vuforia [1] to reach the augmented result. We also predefined some virtual scene. If users want to check the result in more scenarios, he/she can put the avatar to virtual scenarios. So that he/she does not need to go to a big hall while he/she can check the effect when he is dressed with such clothes.

[1] Rig is a term in 3D model building industry, which means generating a skeleton for the 3D model
[1] Vuforia is a platform built by PTC, Inc.

3.4 Novelty

The novelty of our research mainly reflects in these aspects:

1. Improved the performance of current model-based try-on system a lot, decreased the time and increased the quality of appearance;
2. Designed a simplified framework to implement virtual try-on system;
3. Made an interactive model in virtual try-on system for the first time.

Chapter 4

System Design

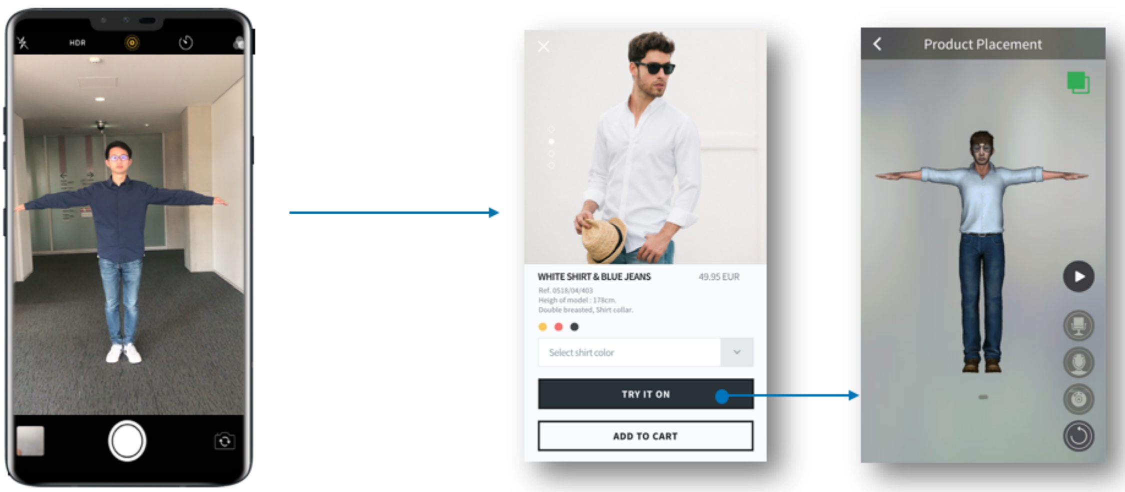


Fig. 4.1 Pipeline of our system

In this chapter, we will introduce our system design and explain each part. The Fig. 4.1 shows the overall pipeline of our system in mobile phone. Before using our system, users need to record a video from 360 degree with a webcam or mobile phone camera. User only needs to do this once, we will generate a 3D model for the user according to the video and save lit locally. Then just like using common e-commerce applications, user first choose a clothes and click the 'Try-On' button in the screen. Then he/she will enter the screen in which he/she can interact with the 3D avatar dressed with the clothes he/she just chose.

From the view of user, the system input is a video of user and the output is an interactive user's 3D model with garments on, user can operate it in smartphone or PC. We will explain our system design following the user flow, which is below five steps:

- Get a personalized avatar;
- Dress the avatar;
- Interact with the avatar;
- Put the avatar into real environment;
- Put the avatar into virtual environment.

4.1 Get Personalized Avatar

For user, user only to upload the video and then he/she will get a personalized avatar. While for us there are actually 3 steps happening in the back-end. To make it more clear, we will explain them one by one, which is generating body, head and hair. After all these steps, we will merge them into one and send it to the user. In order to protect user's privacy, We will save the avatar locally instead of saving it in our server.

4.1.1 Get 3D Body

To get a 3D body, user needs to take a video from 360 degrees first.

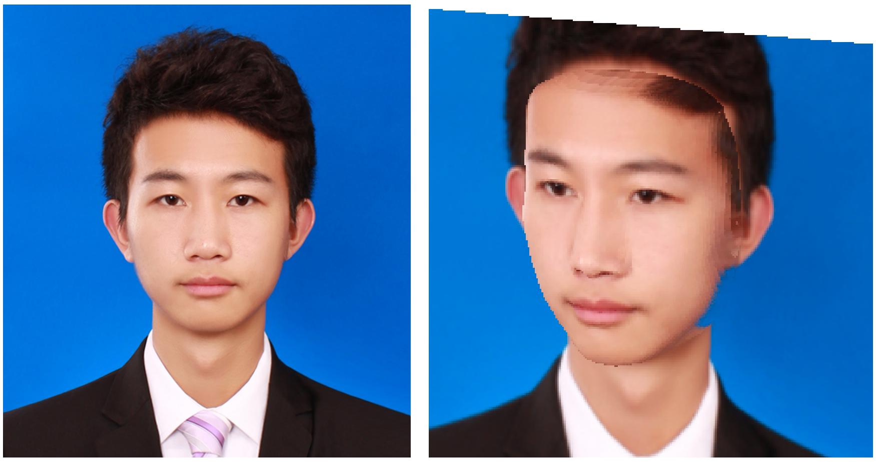
(a) Front view (b) Side View (c) Back view

Fig. 4.2 (a), (b) and (c) are captured when the user is recording a video.

After taking the video, the user needs to upload it to our server. We will build a personalized 3D body for the user. The reason why we need a video instead of pictures is that with the video we can get more detail of the user, which is helpful for us to build a more accurate model. It is worth mentioning that the user can take this video with any device. He/she can use the smartphone, or any webcam. While the higher resolution the better. Since we need to get a clear face from the video. The resolution of the video has a positive correlation with the final mode.

4.1.2 Get 3D Head

As Fig.4.2 shows, when user is recording the video, we will capture different screenshots from it. This will be helpful to build a 3D head. By segmenting the captures we can get different view of the head. Actually the front view is enough for us to build a 3D head. If the user is not satisfied with the resolution of the video, for a better result, he/she can also upload a high-resolution head portrait.



(a) High-resolution head portrait (b) 3D face

Fig. 4.3 (a) is an example of the uploaded image, (b) is the generated 3D face.

4.1.3 Get 3D Hair

The structure of hair is complex. It is hard to generate a good hair model in real-time. To balance the time the effect, we predefined some common hair models. When using the system, we will first assign the user a hair most similar to his/her current hair. If the user is not satisfied, he/she can also choose a hair model by self.

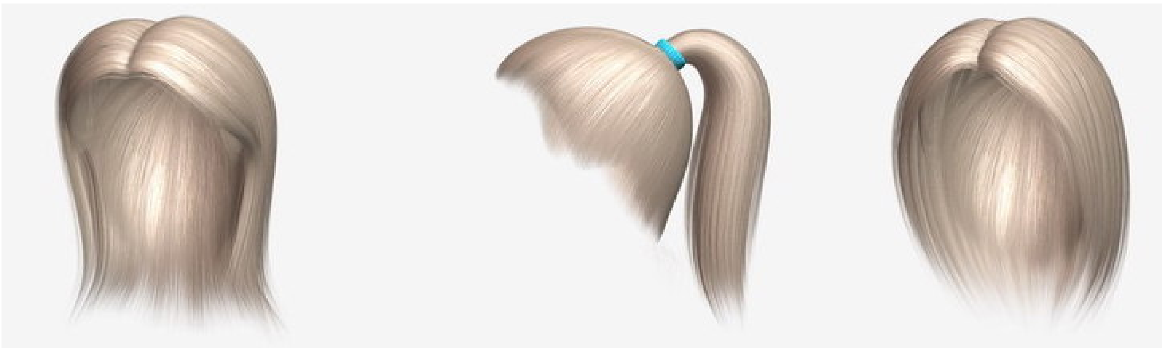


Fig. 4.4 Example of predefined hair models

4.2 Dress the Avatar

Once users have set up the avatar, they can dress the avatar with any clothes. The user experience is similar to common e-commerce applications. First users choose one pleased clothes by browsing clothes images. Then he just need to click the 'TRY IT ON' button before adding it to the cart, as Fig. 4.5 shows. The user will be able to see the dressed avatar after that.

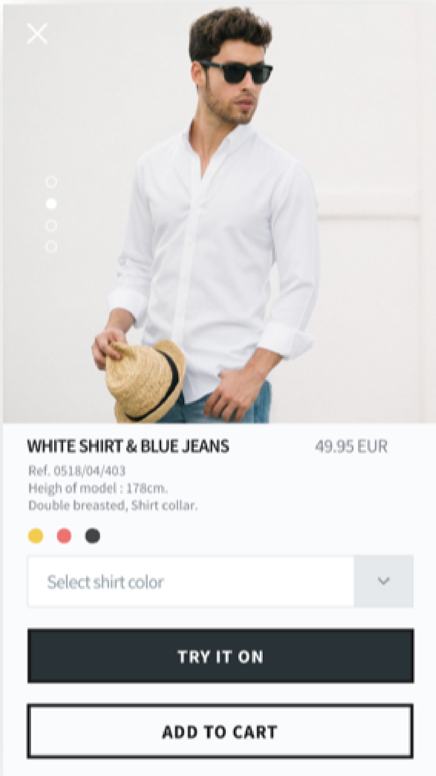
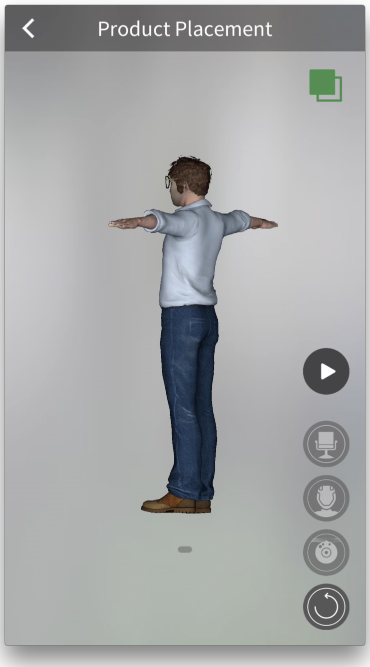


Fig. 4.5 Choose a clothes

4.3 Interact With the Avatar

After choosing a clothes, the user will be led to the avatar page and the user can interact with the avatar. Before putting the avatar to real environment or virtual environment, the user is able to rotate the avatar by sliding the screen. So that the user can have a rough view of the appearance from 360 degrees.

(a) Front view



(b) Side view

Fig. 4.6 (a) is front view of the avatar, (b) is the side view after rotating.

4.4 Put the Avatar into Real Environment

After having a rough view about it, the user can choose to put the avatar into the real environment. With the help of AR technologies, we can put the avatar onto ground through the camera. We will automatically detect the ground for users so that they can place the avatar correctly.

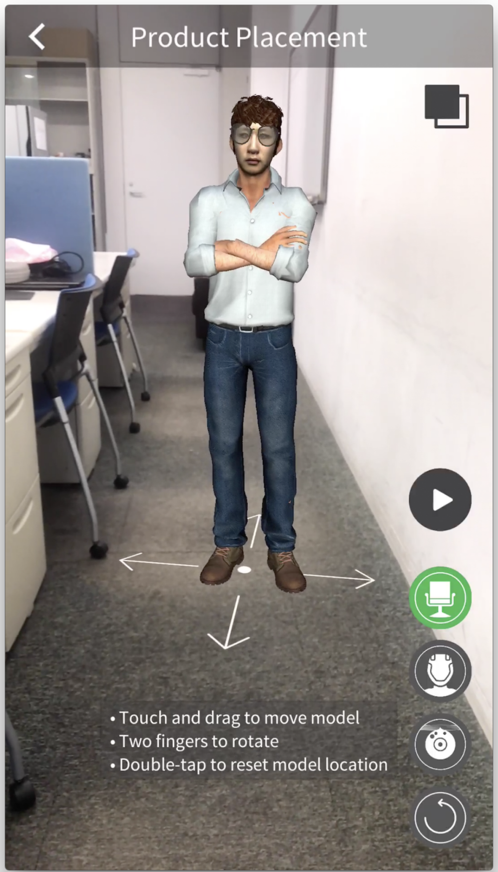


Fig. 4.7 Place the avatar into real environment

As we can in Fig. 4.7, the avatar is at one with the environment. After placing the avatar, the user is also able to interact with the user. As Fig. 4.7 shows, user can touch and drag to move avatar,rotate the avatar with two fingers and replace the model to anywhere by double tap.

Besides putting the static avatar, we can also control the avatar to do some movement. So that we can check the clothes when doing movements.

(a) Walking view

(b) Walking view from the back

Fig. 4.8 (a) is front view of the avatar when walking, (b) is the back view when walking.

As Fig. 4.8 shows, after animating the avatar, we can also rotate it to check from different view.

It is worth mentioning that after placing the avatar, we can get close to check the detail of the clothes both from front and back.

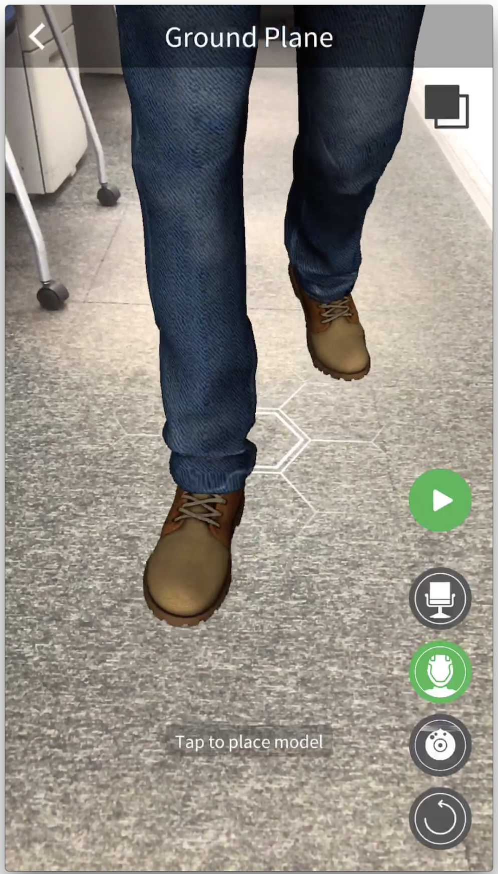


Fig. 4.9 Check detail of the jeans

Besides placing the avatar onto ground, the user can also scale the avatar and put it onto table surface.

Fig. 4.10 Put the avatar onto table

4.5 Put the Avatar into Virtual Environment

For these people who use our system in laptop, it may be different to place the avatar onto ground, since the camera in laptop is not as handy as the one in a smartphone. And also sometimes the user want to know the appearance when dressing the clothes in some specific scenes. Like although I am at home now, but I will dress the clothes only in formal occasions in the future. For these considerations, we also designed some virtual scenes for the user to place the avatar.

Fig. 4.11 Place the avatar in virtual environment

Chapter 5

System Implementation

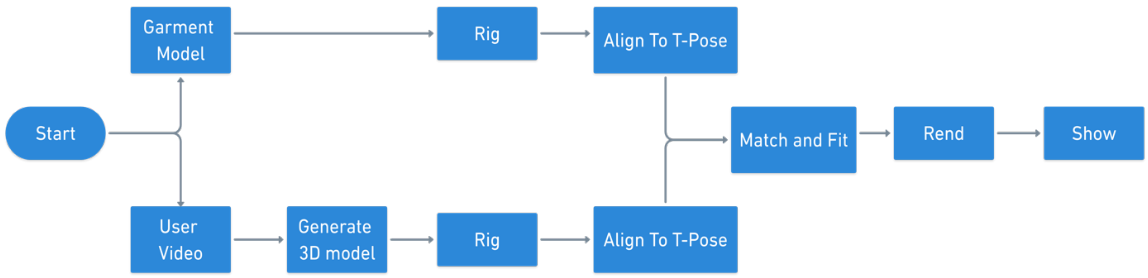


Fig. 5.1 Overall flow.

In general, there are two parts in my system. They are the garments-processing part and user-processing part. For the user part, user first record a video with mobile phone, then upload it to our server, we will generate a 3D model for him, and then we will do rig for it, Rig is a term in model building, it means make a skeleton for 3d model, which can control the model to do some movements or motions.After that we will align the model to a standard pose. And we do the same thing for garments. After that, we fit this two models and show it in client.

5.1 Software environment and Data Collection

In this pager, we use python and C++ as the main language. To make the implementation more fast, we used Blender, Unity and Vuforia.

Blender is an open source software for model building. It is free and has Python API. So that we can integrate our Python code. We import our generated 3D model to it and operate these models by Python inside Blender.

Unity is a software for 3D game building. We import our suited model to it and define the interaction with the help of Unity. We define the UI and complex interactions like gesture control with C++ inside Unity. Also we can use Vuforia in Unity, as a plugin.

As we said before, to use our system, the sellers need to give us 3D clothes models first.This is the only support we need from sellers. Considering the whole process of garment production, most manufacturer should have the garment 3D model. Since it is the first step to produce a kind of clothes.

5.2 Pre-processing

5.2.1 Hair model

As we said before, the current methods can hardly generate a high-poly hair model. Since hair's structure is more complex. It is soft and has countless number. To make our result more realistic and to make the whole process fast and automated, we predefined some hair models, also collected some free models from the Internet. When in the implementation, we will choose a predefined model according to the user's hair.

Fig. 5.2 Predefined hair model.

5.2.2 Skeleton Building

To make the whole process fast and coherent, we need to define a human skeleton model first. It contains the main skeleton of a real human, we will align our generated body model and collected clothes model to it.

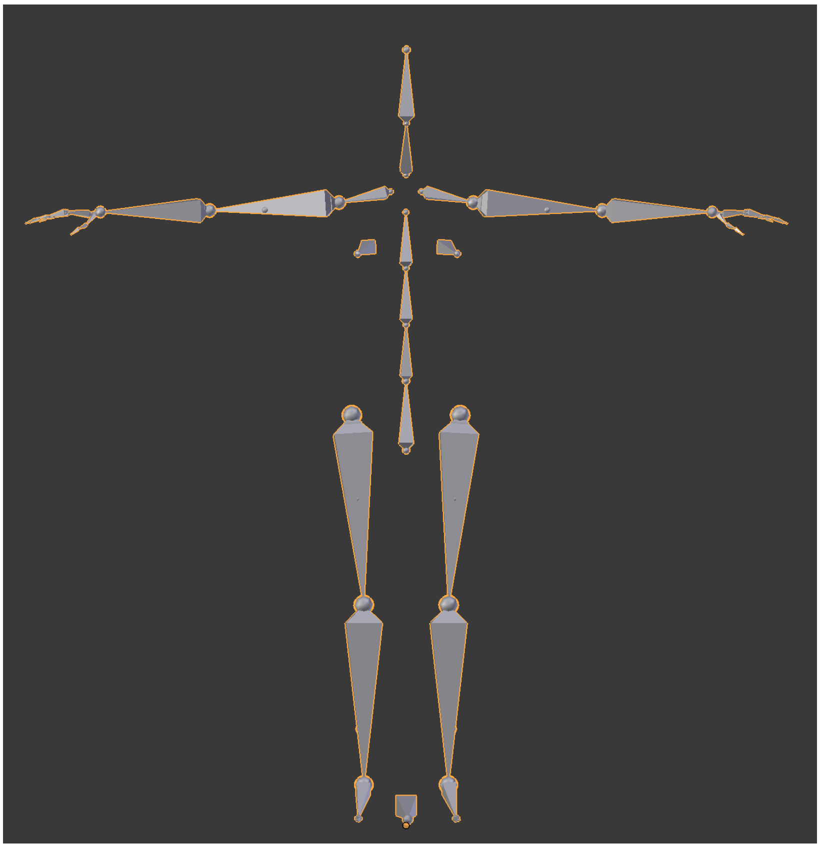


Fig. 5.3 Human skeleton model.

5.3 3D Body Generation

5.3.1 Low-poly Body Generation

As we mention before, when generate the 3D body, we used the method proposed by Alldieck et al. [1]. The user needs to record a video first. After we get this video, we can generate a low-poly body model for the user. While it is not enough. To get a high-poly body, we need to re-texture it. Here we first make the 3D model more smooth, then we texture the body with color of the user's skin. So that it is more realistic. While now we still need to get a high-poly 3D face and 3D hair.

Fig. 5.4 Re-textured body model.

5.3.2 3D Head Generation

Since above model does not has a high-poly 3D face and 3D hair. So we need to re-generate this two components. Here we used the work of Jackson et al.[2]. To use this method, we first need the user to upload a high-resolution head portrait. According to this image, we are able to generate a high-poly 3D face.

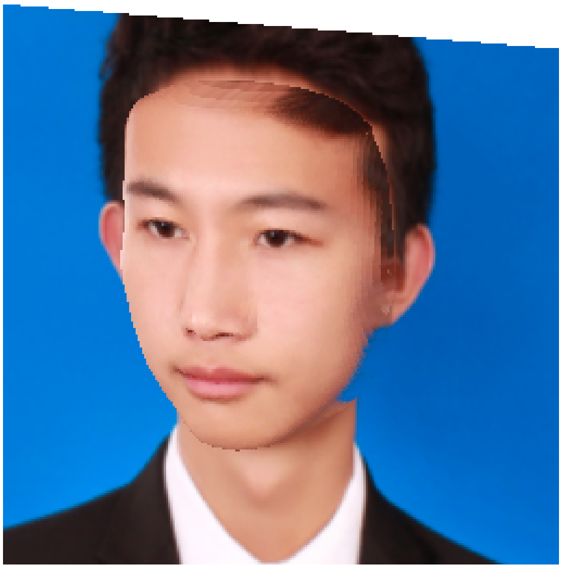


Fig. 5.5 High-poly 3D face.

After generating 3D face, we will choose a 3D hair which is the most similar to the user's with Google Vision API, which is able to calculate the similarity of different images. To this step, actually we have got a high-poly 3D face, high-poly 3D hair and a high-poly 3D model. So next is just to merge them together. We use Pyhton and Blender to merge this three parts into one, by simply aligning the geometric center. By that we can get a high-poly 3D avatar.

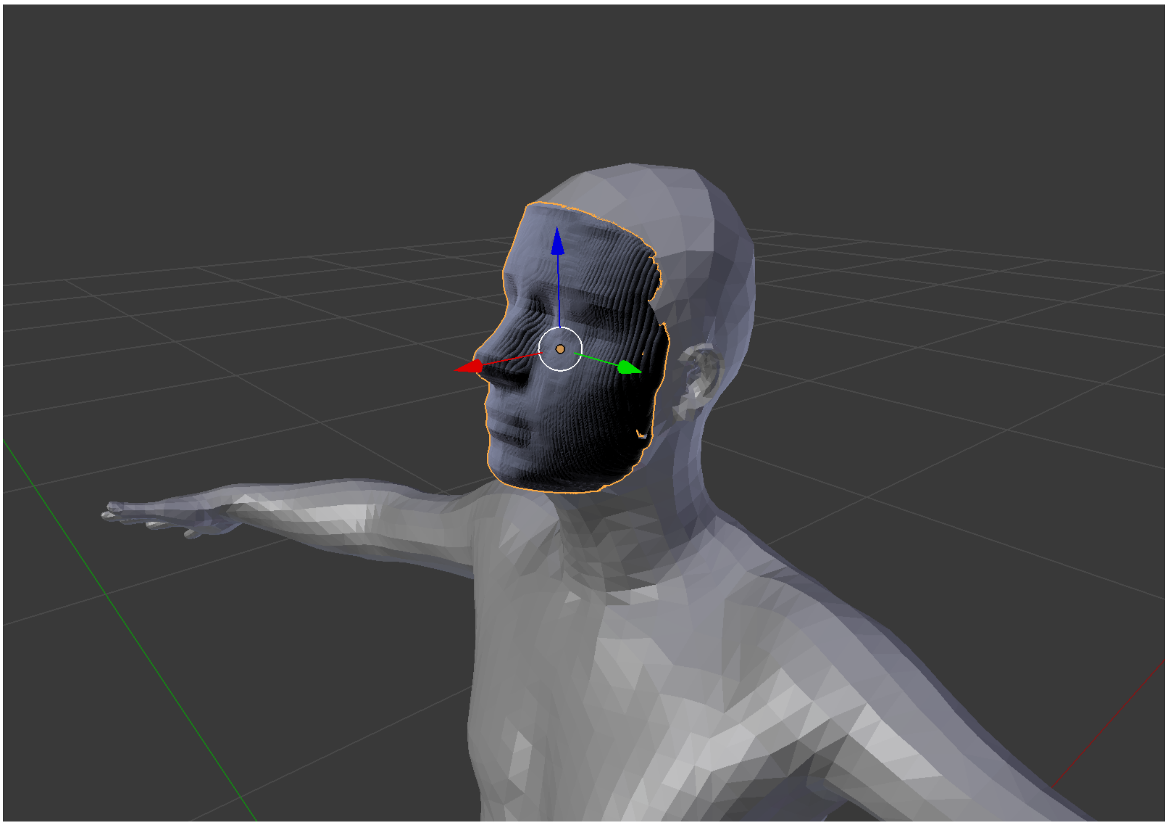


Fig. 5.6 Align models.

5.3.3 Rig

While now we just get a static model. It cannot move and it is quite hard for us to dress up a static model. So we need to Rig it. Rig means assigning the static 3D model a skeleton. We use every part of the skeleton to control corresponding parts of the model.

To do that, we first need to detect the body's key points[5]. Here we detected the chin, neck, wrists, spine root and ankles.

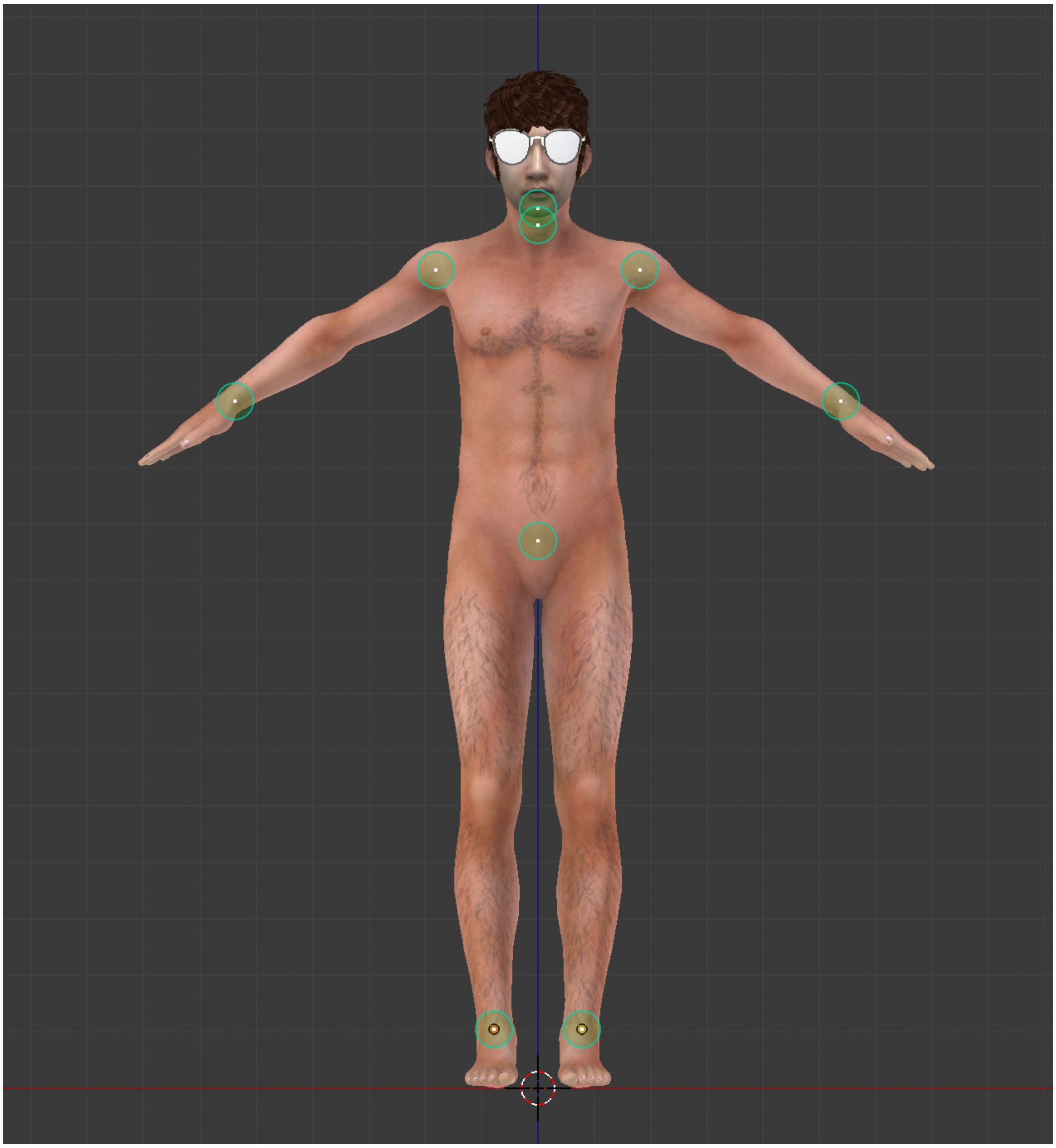


Fig. 5.7 Detect key points.

After this, we just align these detected points to the skeleton, then we can get a rigged 3D avatar.

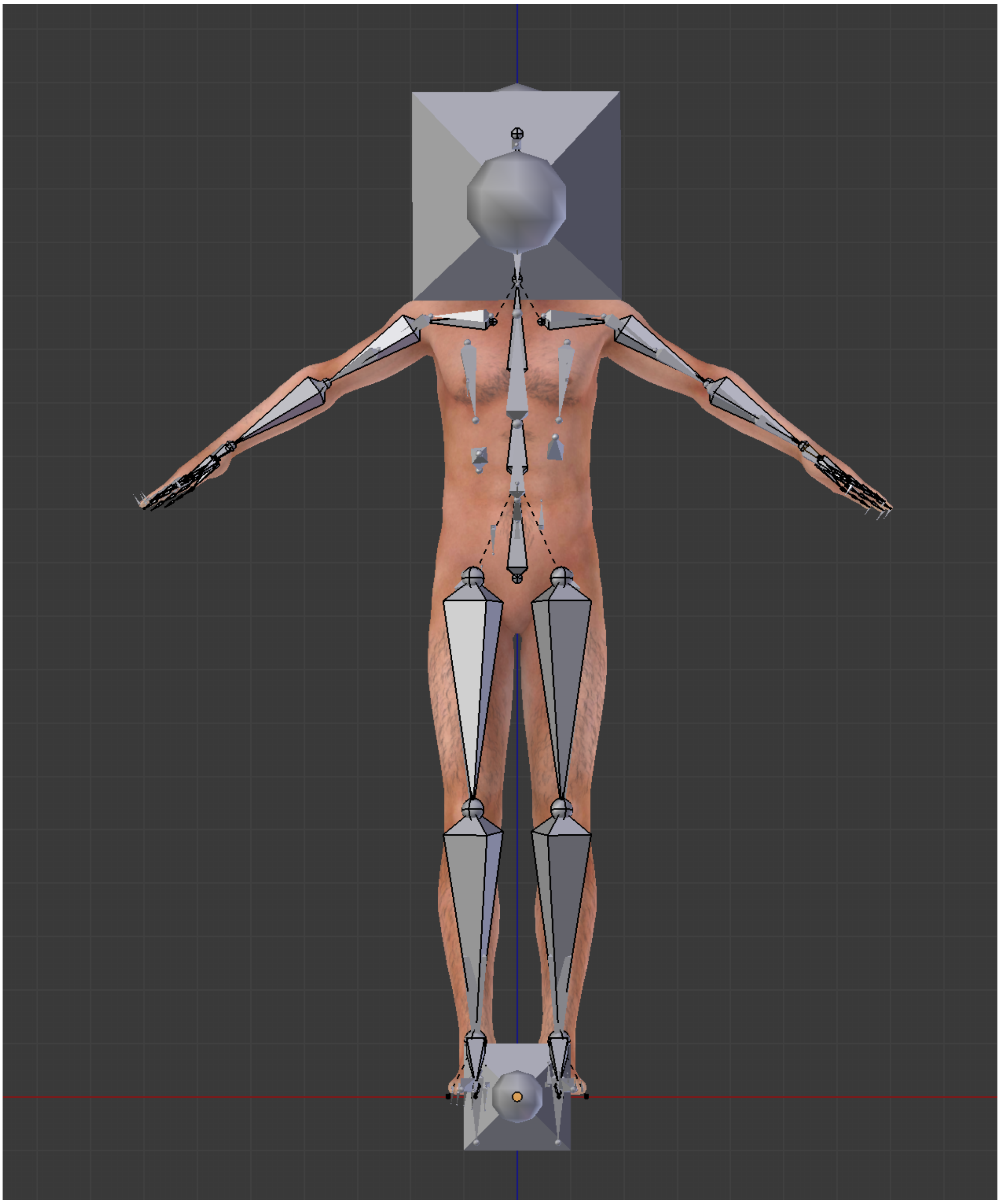


Fig. 5.8 Rigged model.

We can now assign the model some movements.

5.4 Clothes Processing

The process of processing the clothes model is almost the same with the body model. We do not fit the clothes to the body directly. We first detect its key points[6], then align these points to the skeleton.

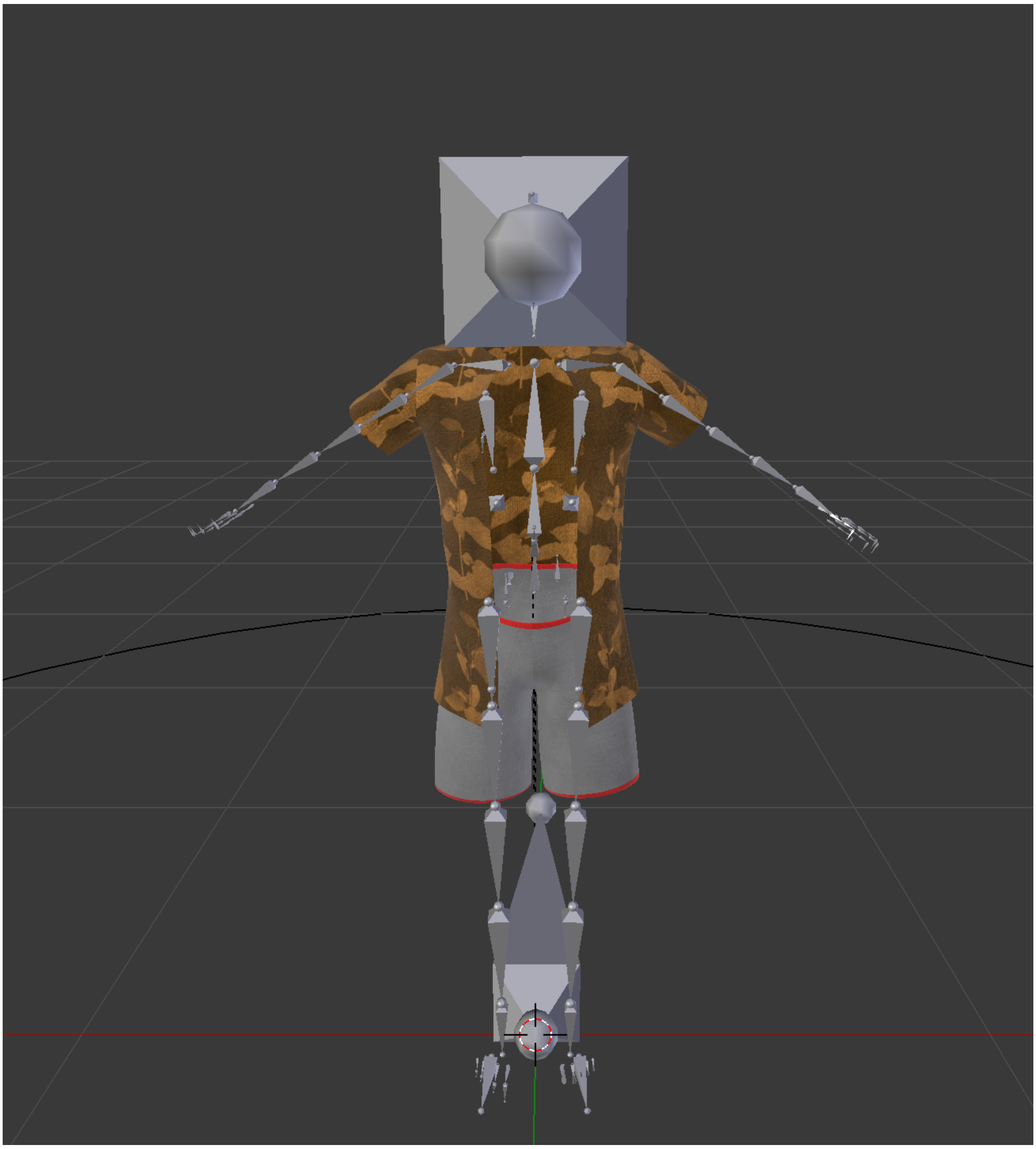


Fig. 5.9 Rigged clothes model.

With the skeleton, we can change the pose of the clothes and assign it movements.

5.5 Fitting

Even we have body model and clothes model, how to fit them into one is still a problem. It's a very hard computer graphic problem that how to fit two 3D models into 1. It's also the reason why in previous research, researchers choose to do this manually. To make it automated, we proposed this framework. We call it as double-skeleton based method.

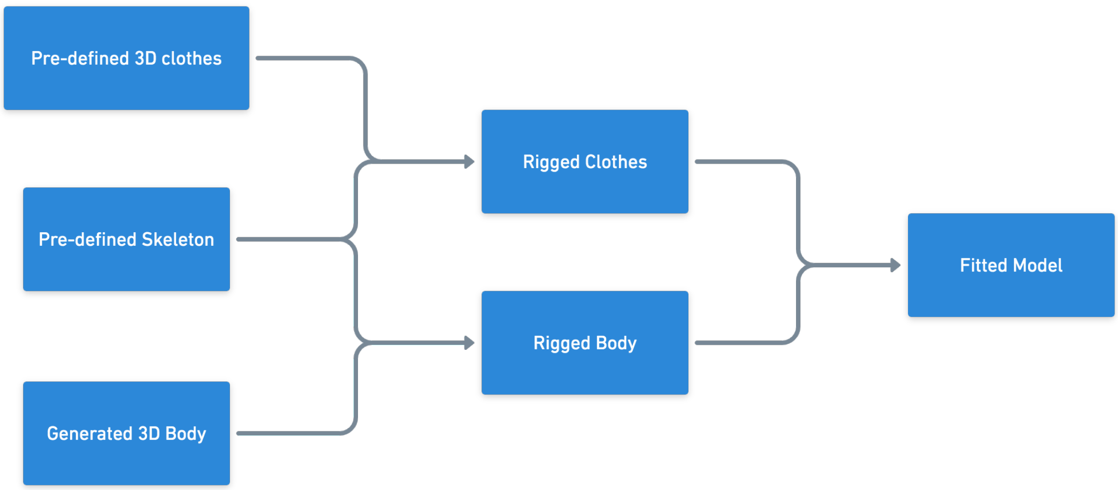


Fig. 5.10 Fitting flow.

We choose to fit a skeleton with the clothes and body first, instead of fitting them directly. Using this framework not only solved the problems, also saved a lot of time. Since the clothes models and body models are generated synchronously.

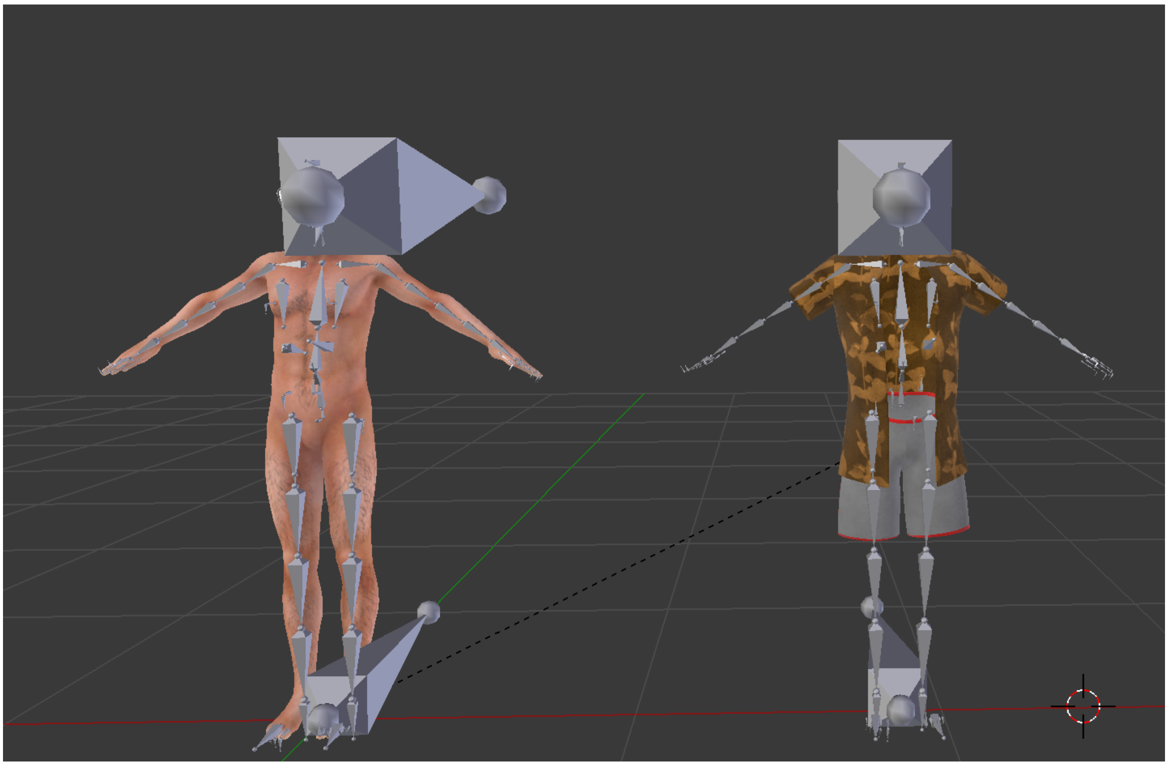


Fig. 5.11 Rigged two models.

Now what we do is to first control them to the same pose. As figure. 5.11 shows, they are both aligned to T pose. Because this two skeleton are with the same structure, next we just overlap this two skeleton to a same position. and after that we get a fitted model.

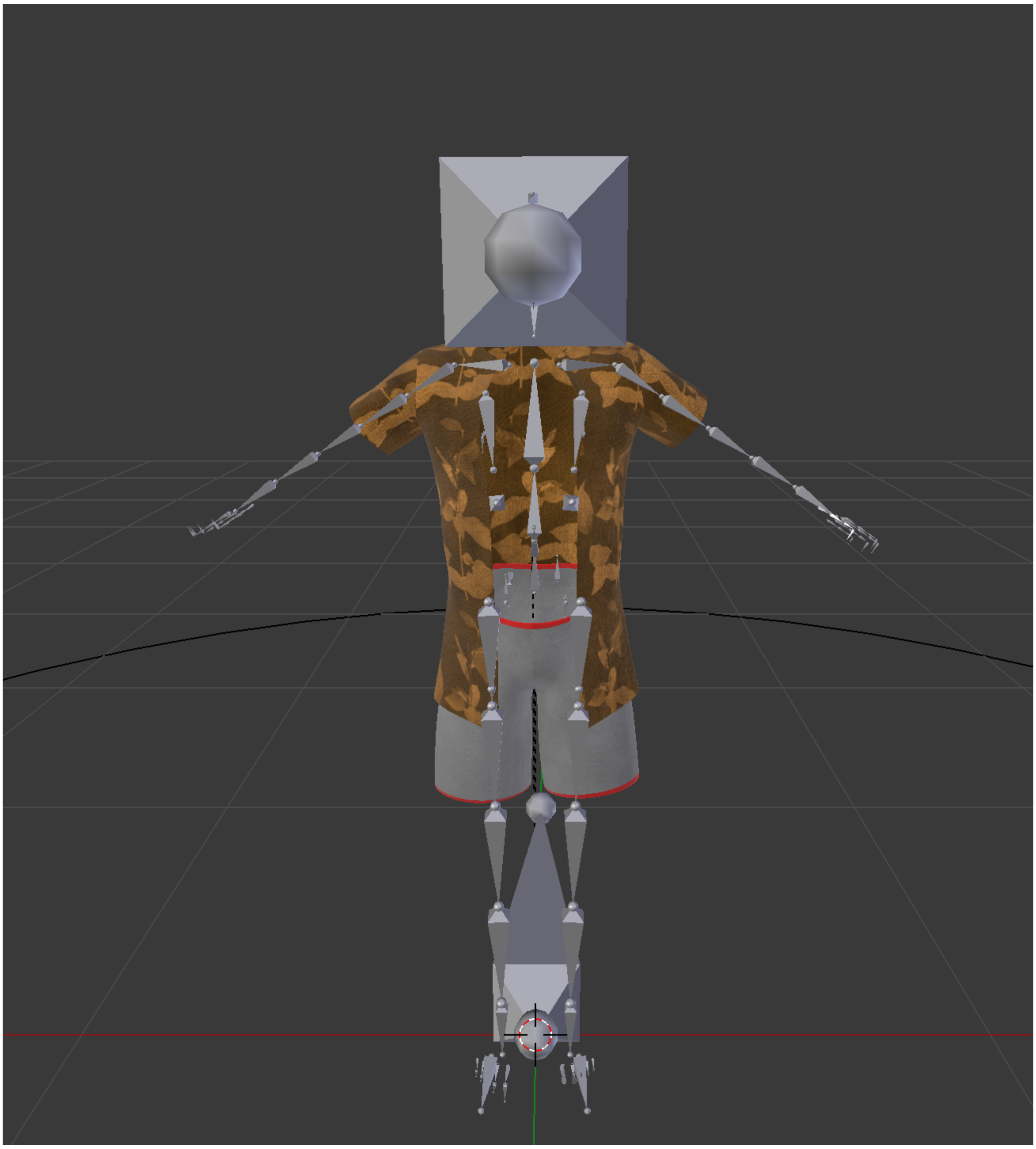


Fig. 5.12 Fitted model.

5.6 AR and VR

Till now we have got a fitted model. Next is to put it into real environment or virtual environment. With the help of Unity, we can define the interaction. For mobile use, we enable user to put the avatar into ground through camera. The user can control the avatar with different gestures in the screen.

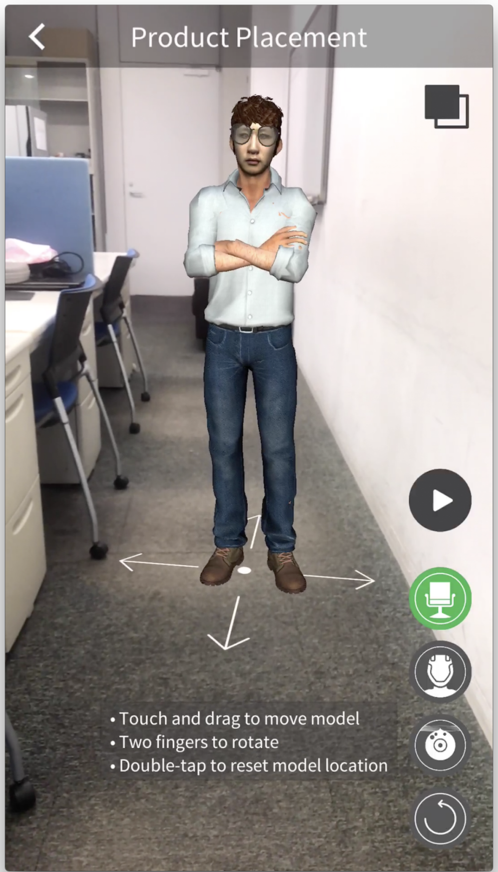


Fig. 5.13 Place the avatar into real environment

As it shows in screen, user can touch and drag to move model, they can rotate the the model with two fingers, they can also scale the model with two fingers.

For laptop use, the user is able to put the avatar into virtual environment. They can use the mouse to rotate the avatar. They can also change the background from these predefined by us.

Fig. 5.14 Place the avatar in virtual environment

Chapter 6

Related Work

The related work will be introduced in this part. There will be two kinds of related work, the first is the work about customized model building and the second is about virtual try-on.

6.1 Related Work about Customized Model Generation

To build a model-based virtual try-on system, a customized 3D human is an important issue. At the very beginning, most researchers prefer to build it with multiple videos or 3D camera. To now, here comes many effective ways to build it.

Ahmed et al. [7] chose to generate a personalized 3D avatar from multiple video data. They recorded a moving person from different view, and then extract the texture from all the video data. By doing this, they are able to generate a realistic 3D avatar.

Looper et al. [8] proposed a Skinned Multi-Person Linear model (SMPL) to generate different body models. This model is trained by large number of aligned 3D meshes of different people in different poses. In this paper, they used many parameters to represent the body model, so the final result is easy to customize by adjusting these parameters.

To generate a whole avatar, hair is also an important part. While the geometry and the structure of hair is complex. So it is much more time-consuming. Most people chose to build it manually before. With the development of deep learning here comes many methods that has good effect. Chai et al. [9] chose to train a neural network to generate the hair model automatically. Though it is time-consuming, while it can generate a 3D hair model from a single image.

The same with the body, with the help of neural network, we are able to generate a 3D body according to a single image. Angjoo Kanazawa et al. [1] proposed a method that we can get a 3D body from a single image. In this paper, they also train a network to map the iamge to the 3D model.

While to increase the accuracy, we chose the method proposed by T. Alldieck et al. [3]. In this paper, they also used the SMPL model proposed by Looper et al. [8]. Compared to the work proposed by Angjoo Kanazawa et al. [1], they record a video for the user from 360 degrees. So that they get more 3D information about the 3D body. And with the help of this we can get a customized 3D body.

When build a customized 3D avatar, the requirement for the face is high. Since it is the biggest feature to differentiate different people. Jackson et al. [2] proposed a method which can generate a high-poly 3D face. They also trained a network here. With their method, we are able to get a realistic 3D face according to a single head portrait.

When generating the 3D models, the garments model is also a difficult part. Since the structure of garment model is more complex. Pons-Moll et al. [10] proposed a method to generate dynamic clothes model. They first capture a clothed person in motion, then they segment the clothes and the body. So that they get a separate clothes model with motion. At last the put this clothes to any body shapes.

R. Danerek et al. [11] trained a convolutional neural network to estimate 3D vertex displacements from a template mesh with a specialized loss function. Their work is able to generate a 3D clothes from a single image. If more images of a clothes is provided, their result will be more accurate. Their work also support for the video input, which is much more accurate.

Yang et al. [12] also proposed a method to generate the 3D clothes with a deep learning method. It is worth mentioning that their work can generate a 3D body and a 3D clothes at the same time. And the input for the system is only an image. Their work cannot generate high-poly 3D body, but they can generate a detailed 3D clothes. With their method, it is also possible to implement garment transfer.

Zhou et al. [13] proposed another method to generate 3D garment model from a single image. They designed an all-pose garment outline interpretation and a shading-based detail modeling algorithms. It is worth mentioning that their work can keep the details well for the clothes, like the wrinkles and folds.

There are a lot of work focusing on how to build a customized avatar. It is a hot area since it is important in the industry of movie-making, gaming, and virtual try-on. While as we can see, thanks to the development of machine learning, now we can generate it automatically and in a very fast way.

6.2 Related Work about Virtual Try-On

As we talked in the ealier chapters of this paper, there are mainly 3 kinds of virutal try-on systems on the market now, they are separately image-based system, live AR and model-based system. Considering the convenience and the satisfaction of the system, we chose to build our system based on model-based system. Here are some related researches about virtual try-on system.

Yu-I Yang et al. [14] proposed a two-stage object tracking algorithm to help user try-on shoes virtually. To use this system, users also need to put some markers on foot first. This system is real-time and have a good effect. The user is able to the result from a screen, like a laptop.

L. Wang et al. [15] proposed an augmented reality system based on Kinect for online handbag shopping. They use only 1 Kinect to implement the magic mirror. In their system the user is able to check the handbag from different view.

Szu-Hao Huang et al. [16] proposed a method to enable the user try-on different glasses frame. They use machine learning algorithms to track the user's face, then they put the virtual glasses frame onto the corresponding place. They also mentioned that their system is mainly used in designing industry.

Pachoulakis et al. [17] proposed a method to build a magic mirror which can relect the reuslt of virtual try-on different clothes. In this research they used 3D scanners to build a 3D model for users first. Then they use graphic algorithms to put the 3D clothes model onto the body model. It is worth mentioning that this system also support for glasses virtual try-on.

Kanamori et al. [18] proposed a method for image-based virtual try-on system. They first cut the image of clothes, then reshape it to the user's photo. Also they did the color correction to make the result more realistic.

Xintong Han et al. [19] proposed a model called VITON. They also trained a network to implement image-based virtual try on. It is worth mentioning that they do it with a coarse-to-fine strategy, which makes the result more realistic.

Giovanni et al. [20] proposed a live virtual try-on system. In this paper, they deployed Kinect and 3D camera to track the user, and then show the result on a interactive screen. Which makes the whole experience is like seeing a mirror.

Yuan et al. [21] also proposed a system giving a experience like looking into a mirror. They also used Kinect. The different thing is that they proposed 3 different scenarios for their system: virtual clothes on the avatar, virtual clothes on the actual user's image, and virtual clothes on the avatar blended with the user's face image.

Sekhavat [22] proposed a system that could be used in mobile phone. In this system, he first build 3D body dressed with target garments in advance, then adjust the model's shape according to the user, last attach the user's face to the 3D body and then show it in an augmented reality way.

Wang et al. [23] proposed a method to implement the virtual try-on in a 2D way, in which the input and the output are both 2D. They called their network as Characteristic-Preserving Virtual Try-On Network. By training this network, they are able to generate realistic dressed images.

When designing the virtual try-on system, the fitting is a big problem, which is how to fit a 3D clothes model onto a 3D body model. J. Li et al. [24] proposed a method to implement this step. They use skeleton to adjust the pose of the clothes model. So that they can get the clothes and the body aligned. We also used this idea in our research.

There are generally two important issues to make a virtual try on system. First is how to generate a realistic 3D avatar, second is how to fit the clothes onto the avatar. To build the system, There are two methods common used in previous research to generate a customized avatar. first is generate a 3D model by scanning the user, with the help of 3D camera. Second is pre-define some 3D models and then adjust them according to the user's personal figures. About the fitting, the most common used method is making it manually.

Chapter 7

Experiment and Result

In order to evaluate our system, we conducted an experiment referring to 10 users. They were asked to use our system and after that they are required to answer some questions. Before users use our product, we will tell them how to use our system. To make the experience more smooth, we insert our system to a normal e-commerce app demo. The only difference with other e-commerce apps is that we provided the function to virtual try-on in our demo app.When designing the experiment we are mainly aiming at testing users' satisfaction. We divided this problem to 3 sub-problems:

1. Are they willing to use this function;

2. When using our system, is there any experience that the user does not like;

3. Is this function helpful for users to make purchase decisions.

To get the answers of above 3 questions, we design the experiment according to following ideas:

1. For the user it is not required that whether they have to user the function. We just observe that how many users will use it and when will the user use it. After the whole test will interview the user according to their behaviors in the experiment.

2. Degree of satisfaction is hard to calculate.In order to get that, we asked them some questions related to Net Promoter Score (NPS), which is widely used by many companies.

7.1 Participants

To evaluate our system, we invited 10 users to join our experiment. They are all asked to follow the process we provided. All the users have the experience of buying clothes online and used at least 1 e-commerce product before. They are all master students and from different research area. Their age is between 20 and 26. They have the skill of English reading and speaking and is familiar with operating the smart phone and laptop.

7.2 Method

As we talked before, to make the experience more smooth, we insert our system to a normal e-commerce app demo. The only difference with other e-commerce apps is that we provided the function to virtual try-on in our demo app. In the experiment, The users are required to follow below steps:

1. Use the demo app to choose the clothes that they like, and 'buy'(enter the process of payment) at least 1 clothes. We will only tell the user how to use our system before the test. The user does not have to use the function of virtual try-on. We will observe their behavior when they are using the demo app and we will not interrupt them or talk to them when they are using the demo app.

2. After using the app, each participant will be interviewed and fill a questionnaire. Last they are asked to give a score from 1 to 10 to describe that how much they would like to recommend this function to their friends.

When the user is using the app, we are mainly observing below behaviors:

1. Will the user use the function;

2. When will the user use the function, before adding the clothes to cart or after adding them to chart;

3.

After using the demo app, we will interview them below questions;The answer is grading from 1 to 5 (1 = very negative, 5 = very positive).

Question	1	2	3	4	5
Do you think it is helpful for you to make purchase decisions?					
Do you think it is easy to use?					
Do you think it can decrease the chance of return?					
Do you think it is better than only pictures?					
Do you think this system is attractive?					

Table 7.1 Investigative questions after using the system

Besides the questionnaire, we also ask the user some open-ended questions:

1. Do you have any comments for the improvement of this system?

After that, we will let the user give the NPS:

Question	Score
How much are you willing to recommend the system to your friends?	

Table 7.2 Investigative question for NPS

We will classify the users into 3 categories according to the score:

1. Recommender (scores between 9-10): People with fanatical loyalty who continue to buy and introduce to others;

2. Passive (score between 7-8): Overall satisfaction but not fanatical, will consider other competitors' products;

3. Detractor (score between 0-6): Unsatisfactory use or no loyalty to the system.

The final score is the ratio between the difference of Rocommender and Detractor and the total number of the participants. The logic of the NPS calculation formula is that the recommender will continue to buy and recommend it to others to speed up your growth, while critics can undermine your reputation and allow you to stop growing in negative word of mouth. NPS scores above 50 percent are considered good. If the NPS score is between 70-80 percent, it proves that your company has a group of high-loyal customers. The survey shows that most companies' NPS values are still between 5-10 percent.

7.3 Result

In our evaluation, all the participants gave the positive feedback. Besides we also got some useful suggests for the improvement of the system. First let me show the behavior we observed when the user was using the system. We separate the behavior into 4 categories: Using the function before adding to cart, Using the function after adding to cart, Using the function in both above situations, Never use. The below table shows the result, Number means the number of participant that has such behavior:

Question	Number
Using the function after adding the cart	6
Using the function after adding to cart	3
Using the function in both above situations	1
Never use	0

Table 7.3 Result of users' behavior observing

As table 7.3 shows, all users used the function.Since the second and the third behavior are a little bit unexpected, so we asked some of the participants below questions according to the behavior:

1. Why do you use the function after adding them to the cart?

2. Why do you use the function both before and after adding them to cart?

And we mainly get two answers corresponding to above questions:

1. When choosing the clothes, checking some pictures are enough for me, while when it comes to the payment I want to confirm that I will not return it. So I want to check again.

2. Before I wanted to check the detail of the clothes, after I want to make sure that I will not return them back.

Now I would like to show the result of the questionnaire.

Question	1	2	3	4	5
Do you think it is helpful for you to make purchase decisions?					10
Do you think it is easy to use?				2	8
Do you think it can decrease the chance of return?					10
Do you think it is better than only pictures?					10
Do you think this system is attractive?					10

Table 7.4 Result of investigative questions after using the system

As table 7.4 shows, all users gave us a positive feedback. While there are 2 participants did not give us the full score for question 2, which means we need to improve the interactions design. And last comes to the NPS.

	1	2	3	4	5	6	7	8	9	10
Score									4	6

Table 7.5 Investigative questions after using the system

Luckily, according to the classification before, all users are Recommender in this experiment, which means our system can increase the satisfaction when the user is buying clothes online. In the experiment we got some good suggests:

1. I want to also check the size when trying-on, hope that you can make it available soon.

2. The most attractive function is putting the avatar onto the ground, maybe you can strengthen this function in your UI design, since all the function are in a same priority now.

3. Hope that I can also use the system to try on shoes and some other accessories, like watch.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

In this paper we proposed an automatic, dynamic and personalized virtual try-on system. Once we deploy the system to a server, users can use it through smartphones or computers. After uploading a video of self, the user can operate a personalized avatar to try on different clothes. By integrating multiple model building methods and proposing a double-skeleton based framework(generate skeleton both for the clothes model and body model), we enable user-like personalized avatars to try on different garments, which can please users a lot if they want to buy clothes online.

In general, we designed a dynamic and personalized virtual garments try-on system by providing personalized users' 3D avatar. We also designed a fast and simple framework to implement the virtual try on system and a double-skeleton based method to fit the 3D garments and 3D bodies.

Compare to previous system, our system mainly has below two advantages:

1. Enhance users' sense of substitution(more engaged).

2. Dynamic and interactive;

Compared to previous framework, the method we use has below advantages:

1. Full automated by integrating different latest 3D model building methods. Previous system needs to fit garments model onto body model manually.

2. Simple and easy to deploy. To use our system, we only need to deploy the code to server.

There are mainly 3 points for this paper:

1. We designed a simplified framework to implement virtual try-on system.

2. We improved the performance of current model-based try-on system a lot, decreased the time and increased the quality of appearance.

3. We made an interactive model in virtual try-on system for the first time.

8.2 Future Work

According to the feedback we got from the evaluation, Below are the directions we are going to work on:

1. Make the system more realistic so that the user can check the size also

2. Make the system more easier to use, give a better user experience.

3. Build the 3D hair model according to the user's picture also.

As we said before, now the 3D hair generation is finished by us in advance. In the future we hope that we can also automate this step. So that all the 3D body is generated automatically. To make that user can also check the size of the clothes when it is on the avatar, we need to pre-process the clothes model from the clothes shop.

References

[1] A. Kanazawa, M. Black, D. Jacobs, and J. Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Regognition (CVPR)*, 2018.

[2] J. Aaron S, B. Adrian, A. Vasileios, and T. Georgios. Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. *International Conference on Computer Vision*, 2017.

[3] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll. Video based reconstruction of 3d people models. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8387–8397, June 2018.

[4] Y. Sekhavat. Privacy preserving cloth try-on using mobile augmented reality. *IEEE Transactions on Multimedia*, 19(5):1041–1049, May 2017.

[5] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[6] P. Luo X. Wang Z. Liu, S. Yan and X. Tang. Fashion landmark detection in the wild. In *European Conference on Computer Vision (ECCV)*, October 2016.

[7] A. Naveed, d. Edilson, T. Christian, M. Marcus, and S. Hans-Peter. Automatic generation of personalized human avatars from multi-view video. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '05, pages 257–260, New York, NY, USA, 2005. ACM.

[8] L. Matthew, M. Naureen, R. Javier, P. Gerard, and B. Michael J. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, October 2015.

[9] C. Menglei, S. Tianjia, W. Hongzhi, W. Yanlin, and Z. Kun. Autohair: Fully automatic hair modeling from a single image. *ACM Trans. Graph.*, 35(4):116:1–116:12, July 2016.

[10] P. Gerard, P. Sergi, H. Sonny, and B. Michael. Clothcap: Seamless 4d clothing capture and retargeting. *ACM Transactions on Graphics, (Proc. SIGGRAPH)*, 36(4), 2017. Two first authors contributed equally.

[11] R. Danerek, E. Dibra, A. Cengiz Öztireli, R. Ziegler, and M. Gross. Deepgarment : 3d garment shape estimation from a single image. *Comput. Graph. Forum*, 36:269–280, 2017.

[12] Y. Shan, A. Tanya, P. Zherong, W. Ke, Y. Licheng, B. Tamara, and L. Ming. Detailed garment recovery from a single-view image. 08 2016.

[13] Z. Bin, C. Xiaowu, F. Qiang, G. Kan, and T. Ping. Garment modeling from a single image. *Computer Graphics Forum*, 32(7):85–91, 2013.

[14] Y. Yu-I, Y. Chih-Kai, and C. Chih-Hsing. A virtual try-on system in augmented reality using rgb-d cameras for footwear personalization. *Journal of Manufacturing Systems*, 33(4):690–698, 2014.

[15] L. Wang, R. Villamil, S. Samarasekera, and R. Kumar. Magic mirror: A virtual handbag shopping system. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–24, June 2012.

[16] S. Huang, Y. Yang, and C. Chu. Human-centric design personalization of 3d glasses frame in markerless augmented reality. *Advanced Engineering Informatics*, 26(1):35 – 45, 2012. Network and Supply Chain System Integration for Mass Customization and Sustainable Behavior.

[17] P. Ioannis and K. Kostas. Augmented reality platforms for virtual fitting rooms. *The International journal of Multimedia Its Applications*, 4, 08 2012.

[18] K. Yoshihiro, Y. Hiroki, H. Masaki, M. Jun, and F. Yukio. *Image-Based Virtual Try-On System with Garment Reshaping and Color Correction*, pages 1–16. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[19] X. Han, Z. Wu, Z. Wu, R. Yu, and L. Davis. Viton: An image-based virtual try-on network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7543–7552, 2018.

[20] G. Stevie, C. Chul, H. Jay, K. Tat, and Y. KangKang. Virtual try-on using kinect and hd camera. In Marcelo Kallmann and Kostas Bekris, editors, *Motion in Games*, pages 55–65, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[21] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M. Foo. A mixed reality virtual clothes try-on system. *IEEE Transactions on Multimedia*, 15(8):1958–1968, Dec 2013.

[22] Y. A. Sekhavat. Privacy preserving cloth try-on using mobile augmented reality. *IEEE Transactions on Multimedia*, 19(5):1041–1049, May 2017.

[23] W. Bochao, Z. Huabin, L. Xiaodan, C. Yimin, L. Liang, and Y. Meng. Toward characteristic-preserving image-based virtual try-on network. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[24] J. Li, J. Ye, Y. Wang, L. Bai, and G. Lu. Fitting 3d garment models onto individual human models. *Computers Graphics*, 34(6):742 – 755, 2010. Graphics for Serious Games Computer Graphics in Spain: a Selection of Papers from CEIG 2009 Selected Papers from the SIGGRAPH Asia Education Program.