

Interactive Personalized Human Avatar and Its Application to Work Incentive System



YUAN JINGYI

44181622-4

Master of Engineering

Supervisor: Prof. Jiro Tanaka

Graduate School of Information, Production and Systems

Waseda University

September 2020

Abstract

An avatar can be a graphical representation of the user which often resembles a real human. In recent years, avatar use in online gaming and collaborative system has grown tremendously. The more realistic an avatar is, the stronger the user's sense of substitution.

In this paper, we focus on generating a personalized human avatar and use it to perceive and understand user's emotion states in augmented reality environment. Based on it, we build a work incentive system which uses the avatar as the motivator. Our system allow user to have interactions with customized avatar and receive different feedback from avatar according to user's emotion.

Our system consists of three main parts:

1. Create a personalized human avatar. We build a realistic avatar with body shape, face and clothes based on a video in which a person is moving.
2. Human-avatar interaction. We use the user's voice and facial expressions as input continuously, and use this as a basis to design the interaction between human and avatar.
3. A work incentive system. We view avatar as a motivator and use it to help users improve their emotions.

We have invited some participants to test how similar our model is to real people and if the interaction is useful. We got a positive feedback through the preliminary user study.

Keywords: Personalized Avatar, Human-Avatar Interaction, Work Incentive System

Acknowledgements

First and foremost, I would like to express my great appreciation to my supervisor, Prof. Jiro TANAKA, for the continuous support and patient guidance through every stage of the research. He provided me professional guidance on HCI field and taught me a great deal about scientific research. He has inspired me a lot to become an independent researcher and helped me find out my way in personalized-avatar research. I consider myself very fortunate to have been his student for two years.

Besides, I am grateful for the assistance given by all my lab mates. They generously gave their time to offer me constructive suggestions toward improving my research. Thanks for your companionship and support for helping me execute important experiments.

Finally, I would like to thank my family, whose love and encouragement are with me in whatever I pursue. You give me financial and mental support throughout my master study life. Thank you for encouraging me when I feel down. You are always my strong shield.

Contents

List of figures	vii
List of tables	ix
1 Introduction	1
1.1 Introduction	1
1.2 Organization of the Thesis	3
2 Background	4
2.1 Personalized Human Avatar	4
2.1.1 3d Scanning	4
2.1.2 Adjust Parameters of Preset Model	5
2.1.3 Image-Based and Video-Based Method	5
2.2 Human-Avatar Interaction and Emotion Recognition	6
2.3 Work Incentive System	7
3 Research Goal and Approach	8
3.1 Goal	8
3.2 Approach	8
4 System Design	11
4.1 System Overview	11
4.2 Generated Personalized Human Avatar	12
4.2.1 Video Preprocessing	12
4.2.2 Generate Human Model	14
4.2.3 Texture	16
4.2.4 Surface Subdivision	17
4.3 Interact with Avatar	19
4.3.1 Add Movements and Voice to Avatar	19

4.3.2	Emotion Recognition	21
4.3.3	Speech Recognition	22
4.3.4	Place and Manipulate Avatar in AR Environment	23
4.4	Avatar-Based Work Incentive System	24
4.4.1	Emotion Analysis	25
4.4.2	Use Scene	26
5	System Implementation	29
5.1	Hardware	29
5.2	Development Environment	30
5.3	Framework	30
5.4	Avatar Generation	31
5.4.1	Model Generation	31
5.4.2	Texture Generation	33
5.4.3	Surface Subdivision	34
5.5	Interact with Avatar	34
5.5.1	Animation Controller	35
5.5.2	Add Voice to Avatar	36
5.5.3	Facial Expression Recognition	36
5.5.4	Voice Recognition	37
5.5.5	Multiple Aspects Emotion Recognition	37
5.5.6	Speech Recognition	39
5.6	Store Database	39
5.6.1	Entities	39
5.6.2	Connect Database and Store Data	40
5.6.3	Update Line Chart	41
5.7	Augmented Reality	42
6	Related Work	43
6.1	Related Work on Personalized Human Avatar	43
6.2	Related Work on Emotion Recognition and Analysis	45
7	Preliminary Evaluation	46
7.1	Participants	46
7.2	Method	46
7.3	Result	47

Contents	vi
<hr/>	
8 Conclusion and Future Work	51
8.1 Conclusion	51
8.2 Future Work	52
References	53

List of figures

1.1	The process of building a personalized human avatar	2
1.2	The workflow of work incentive system	3
2.1	Adjust parameters of preset model	5
2.2	SMPL models of various poses	6
3.1	Preprocess input video	9
4.1	System overall flow	12
4.2	The video frames of a person moving around	13
4.3	Body joints detected from frames	13
4.4	Extract human body silhouette and binarize the frame	14
4.5	Two intermediates	15
4.6	Naked body models with different shape parameter	15
4.7	The human body and the silhouette model	16
4.8	UV map between texture and human model	17
4.9	Different textures on the same avatar	17
4.10	Model before optimization and model after optimization	18
4.11	Human avatar from different views	18
4.12	Rig the avatar	19
4.13	Avatar Animation	20
4.14	Text-to-speech system	20
4.15	Four kinds of facial expression	21
4.16	The specified emotional trend was detected	22
4.17	The key words of user's audio clip were detected	22
4.18	Place the avatar into real world	23
4.19	Avatar Manipulation	24
4.20	A line chart used to indicate emotions	25
4.21	Trend line	26

4.22	Main interface	27
4.23	Encourage user	27
5.1	System framework	31
5.2	Model generation	32
5.3	OpenPose example	32
5.4	Extract body outline and binarize it	33
5.5	Texture eample	34
5.6	Surface subdivision	35
5.7	Animation controller	35
5.8	Convert text to speech	36
5.9	Facial expression recognition	37
5.10	Voice recognition	38
5.11	Combine two initial result	38
5.12	Speech recognition	39
5.13	Insert data to specified table	41
5.14	SQLite database	41
5.15	Read data	42
5.16	Configure plane in Unity	42
7.1	Questionnaire	48
7.2	Questionnaire results	49

List of tables

5.1	The information of PC	29
5.2	Facial expression table	40
5.3	Voice table	40
7.1	Investigative questions after using the system	47

Chapter 1

Introduction

1.1 Introduction

With the development of augmented reality (AR) technology, virtual human avatar has become a significant role in most virtual worlds [1]. Realistic human avatars with lifelike appearance and unique behavior can give user a sense of reality. People can see the avatar as the second incarnation of anyone they want. Avatar has many usage scenarios, and the most common one is to use it to represent a real person in social games. Nowadays, it is widely used in our life, for example, medical practitioners use it to treat mental diseases, fitness coaches use it to observe people's body changes and so on.

On this basis, the interaction between people and avatar has also developed. Avatar can have realistic appearance and behavior like human beings. Even, the communication with avatar is the same as that with people, so that avatar can also assume a certain social role like a human, such as a good friend who will comfort you when you are sad or a company manager who will encourage you when your work encounters bottlenecks.

Similarly, it can also become a motivator or a partner in our work and daily life. Therefore, after generating a realistic avatar, We consider it as a role of motivator and use it in the work incentive system. For companies today, incentive system is essentially a reward system which always uses base pay or annual salary as the motivator. However, numerous studies [2] have proved that some intrinsic motivators like the sense of achievement are

more effective than extrinsic reward in boosting user's motivation to work. So, avatar may bring a new opportunity to the traditional work incentive system that give user the intrinsic motivation.

In this research, we create an interactive personalized avatar and introduce it into work incentive system and use it to help user adjust their mood.

The Fig.1.1 shows the process of building a personalized human avatar. We generate avatar's model and texture according to the video user uploaded. The body shape and dress of the generated model will look as much like the person as possible. Then model will be rigged and animated to make the avatar.

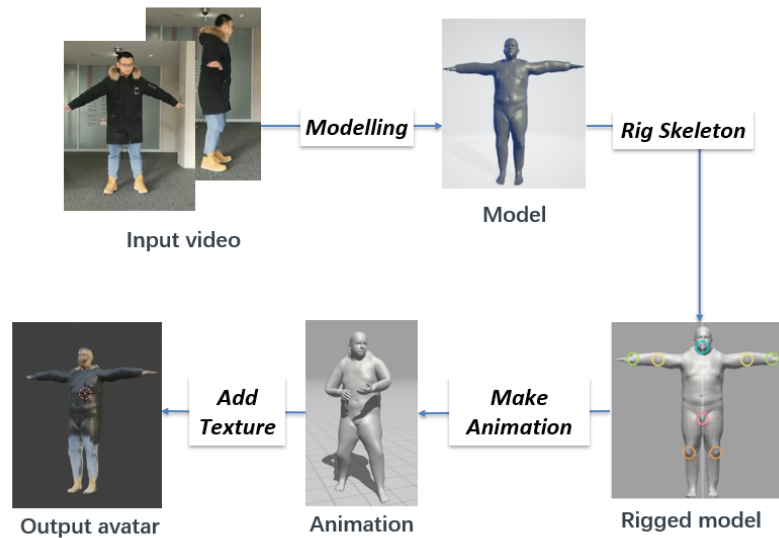


Fig. 1.1 The process of building a personalized human avatar

Fig.1.2 shows the workflow of avatar-based work incentive system. Our 3d avatar can be placed in augmented reality environment. In order to make avatar assume the role of motivator, the system will perceive the user's emotional fluctuations at work and feed it back to avatar in real time. After analyzing the user's emotional information, the analysis results will drive avatar to give correct feedback and help user adjust their working mood.

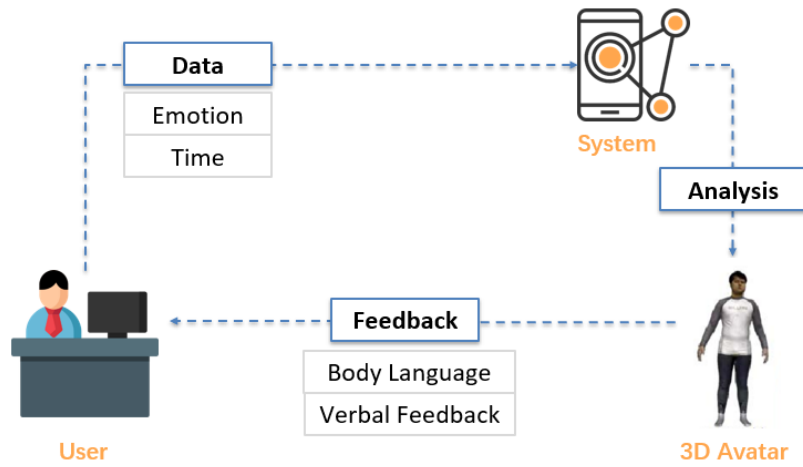


Fig. 1.2 The workflow of work incentive system

1.2 Organization of the Thesis

The rest of the thesis is organized as follows: Chapter 2 introduces the background of the thesis and some related research fields. Chapter 3 will tell the research goal and the approach briefly. Chapter 4 is the system design part, where the design concept and ideas will be introduced, and the mechanism 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 preliminary evaluation; we will talk about the usability of our system. Chapter 8 will be the conclusion and future work part, where we will conclude the previous content and talk about future possibilities.

Chapter 2

Background

2.1 Personalized Human Avatar

In HCI field, a personalized avatar usually can be considered as the graphical representation of a person [3] which is required to represent human features such as their body shape and appearance as accurately as possible. The advantage is that it can give users a sufficient sense of substitution.

There are several common ways to generate avatars: 3d scanning, adjust parameters of preset model, image-based method and video-based method.

2.1.1 3d Scanning

3d scanning technology can scan user from 360 degree and quickly obtain the full surface geometry of the human body.

However, this technique usually relies on external hardware devices such as Kinect to build the body model. Therefore, it is expensive and inconvenient for general user who has no relevant professional skills to create their virtual avatar. At the same time, this method is usually used to capture the shape of the human body, but some details of the human body cannot be captured well.

2.1.2 Adjust Parameters of Preset Model

As shown in Fig.2.1, this method allow user to manually modify the parameters of the human template model according to their own figures, such as their height and waist circumference.

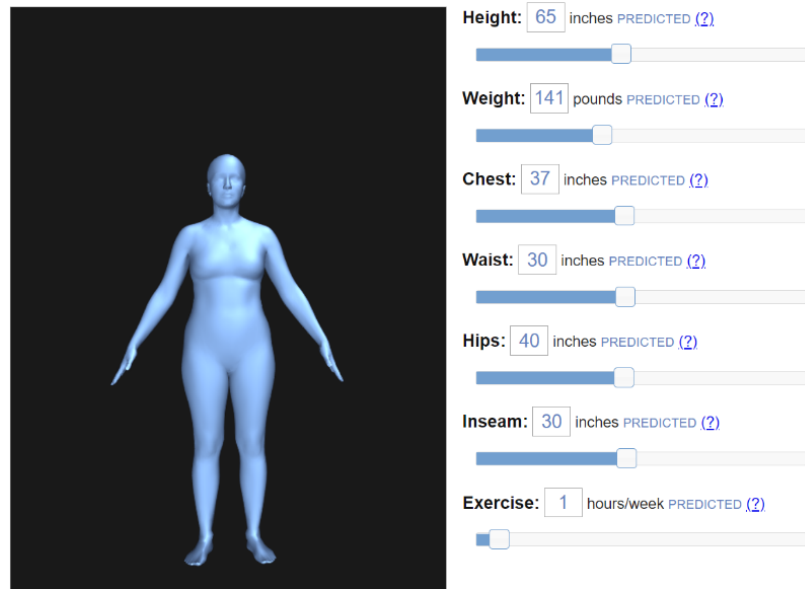


Fig. 2.1 Adjust parameters of preset model

The disadvantage of this method is obvious, that is, users need to measure their own body data in advance. Therefore, the modification is subjective and error prone. Also, this method is limited to body data so that it is difficult to change user's appearance like face and clothes.

2.1.3 Image-Based and Video-Based Method

Compared with the above two methods, image-based and video-based methods have obvious advantages in reconstructing human avatar. After user uploading an image or video, the method takes advantage of machine learning to build more accurate avatars which contain facial portraits and clothing textures.

Foe video-based method, the real person is rendered from multiple input video frames. In contrast, human models generated based on pictures are not very accurate due to lack

of some key information like depth. Therefore, compared with image-based, video-based methods can extract more human body information from multiple angles. In our research, we use Alldieck's work [4] to get accurate 3d human body models from a single video sequence of the person moving in front of the camera. This method is based on Skinned Multi-Person Linear model (SMPL) [5] which is driven by ten human body shape parameter and pose parameter. As Fig.2.2 shows, this model can be well used in animator control. In this study, we used the video-based model building method to create the initial human model. Based on the original work, we optimized the surface details of the generated model and generated corresponding textures. After that, we add sound and movement to the avatar after the model has been rigged.

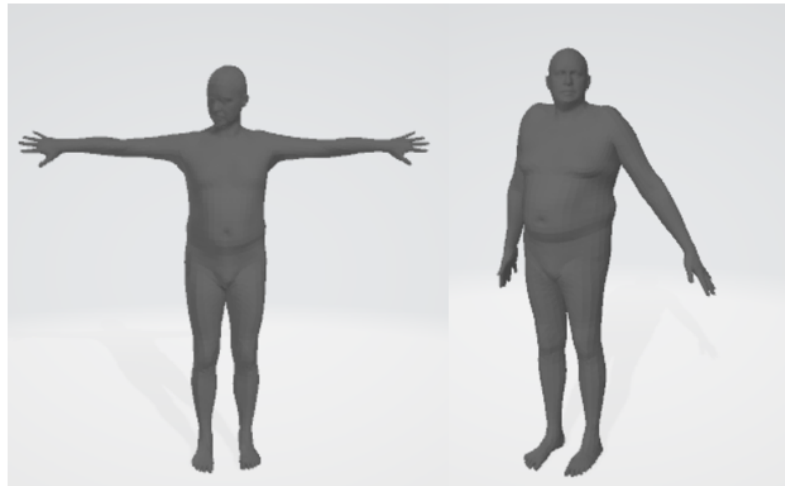


Fig. 2.2 SMPL models of various poses

2.2 Human-Avatar Interaction and Emotion Recognition

After building the personalized avatar, the design of the interaction between avatar and user is needed. The existing interactive technology can be divided into the following aspects: track user data, understand user intentions, understand user emotions and communicate with user. In our research, we mainly focus on tracking user emotion data, identifying user emotions and giving verbal and non-verbal feedback.

Emotion recognition has become a hot research topic in human-avatar interaction [6] with the development of machine learning. It is proved that interactive avatar with appropriate emotional feedback can effectively improve the user experience. In our research, the interaction allows the personalized avatar to track and understand user's emotion in real time and adjust its own behaviors as the feedback accordingly. 4 kinds of common emotion of user (calm, happy, sad and angry) can be tracked and recognized. Different emotion states will drive different avatar's behavior.

2.3 Work Incentive System

Incentive system is an aggregate of numerous research areas, including psychology, economics and computer science. Its main purpose is to boost people's motivation to do something. Organizations always use different tools such as rewards to enhance the performance of the employee [7]. However, it is worth noting that by improving the user's mood and thereby improving the user's work efficiency is also the method the company is trying to use. Application like Jibun Yohou [8] improves user's motivation to work by identify employees' mood from voice. User need to actively record their voice then system will give an analysis according to the input data. This makes it possible for company managers to detect employees' moods in time and help them improve work efficiency.

Chapter 3

Research Goal and Approach

3.1 Goal

Our research tries to build a realistic personalized human avatar based on a video with a person moving around and try to use this avatar as a motivator in work incentive system, which aim at boosting the work motivation of user. Our goal can be divided into the following points:

1. Generate an interactive personalized virtual avatar with body shape, face and clothes.
2. Add realistic movements and voice to avatar.
3. Make a work incentive system: Real-time emotional recording and analysis. Avatar is used as a motivator to give positive feedback.

3.2 Approach

To achieve our goal, we divided our work into three parts: one is generating an interactive realistic avatar and another is making a work incentive system.

To build the avatar, user need to upload a video with the person moving around as the input. In order to make the final result good enough, the video resolution needs to be as high as possible, and the character outline needs to be clear. We also need preprocess the

video user uploaded. Fig.3.1 shows the process of preprocessing. In order to get enough information from different views, we need to convert mp4 file to a sequence of JPG files. Then we detect the 18 body joints using OpenPose and human silhouette from these image files. The information of body keypoints will be stored in JSON files. Here we consider JSON files and mask files as preprocess result.

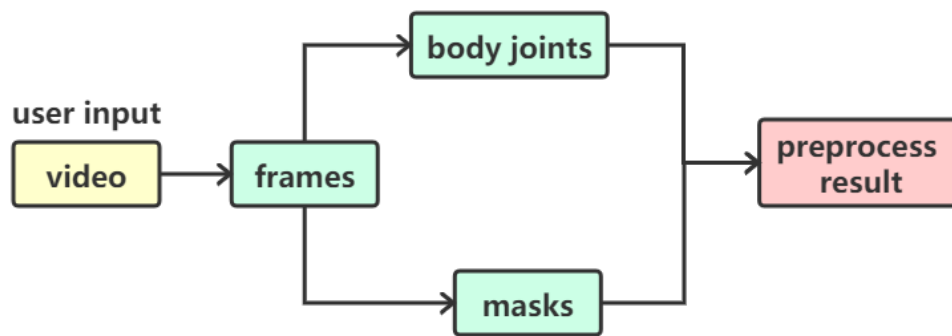


Fig. 3.1 Preprocess input video

We choose Alldieck's method to generate basic body model. By this ways, we can get the model with body shape and face. While this model cannot be adapted because of its low-poly. Optimization is needed to make it more realistic. We subdivide the surface by add vertices on the edges of the model to make avatar more smoother. Then, we can add the texture to make the avatar looks more like a real person.

Then we rigged the avatar using a skeleton with 8 body joints, including wrists, groin, chin, elbows and knees. To make it interactive, We also generate some animation and audio clips for this avatar.

An application scenario is designed for the avatar. It is a work incentive system which allows our avatar to perceive user's emotion during work time and help user adjust their mood. This system can recognize four kinds of emotion including calm, happy, sad and angry which range from 0 to 50. In order to more accurate perceive these emotion from user, we combine facial expression recognition and vocal recognition together. After that, emotion data will be stored in the SQLite database. For analysing the trend of user's emotion, we use line chart to record and upload emotion data in real time.

In order to make the avatar work in the work incentive system. We firstly need to detect the plane and place the avatar in real world using ARCore and the rear camera of smartphone. Then the only thing users need to do is making sure their face is within range of the smartphone's front-facing camera. The system will detect user's emotion automatically and give feedback accordingly. For example, if our system detects that user keeps in a good mood, the avatar will do the animation of clapping hands and say "keep going!".

Chapter 4

System Design

In this chapter, we will introduce our system design and explain each part. At first, we will introduce the system overview and how the system works. Then the rest of this chapter will be divided into three parts to introduce the system design in detail.

4.1 System Overview

As we have mentioned above, our research contains three part. The first two parts are directly related to avatar: how to generate an personalized avatar and how to make the avatar interactive. The third part is an application of our avatar.

- Part 1 is about generating personalized human avatar. We explain how we build the 3d avatar from a single video and how we optimize the avatar.
- Part 2 is about interaction part. We add some body language and audio clips for avatar. And we explain the way of tracking user emotion data and how avatar understands speech and emotion of user.
- Part 3 will introduce the avatar-based work incentive system. We explain the way of using avatar in augmented reality and how the system analysis the trend of emotion and help user adjust their mood at work.

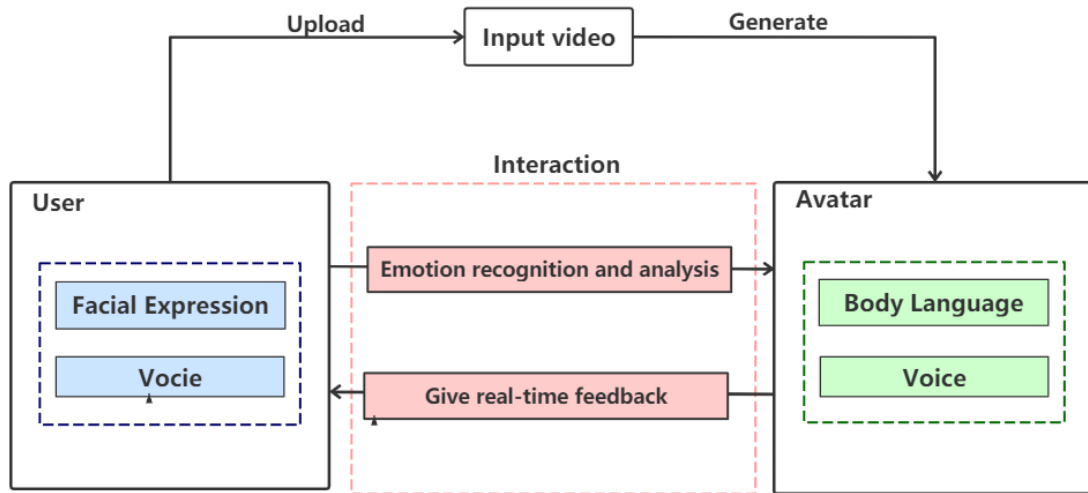


Fig. 4.1 Work incentive system overall flow

The Fig.4.1 shows the overview of the third part: an avatar-based work incentive system. First, we generate customized avatar based on the video user has uploaded. Emotion recognition technology is used to track user's data. At the same time, data of emotion and time will be uploaded to database in real time. At regular intervals, the system will analyze the data extracted from the database and guide avatar to give verbal and nonverbal feedback to the user.

4.2 Generated Personalized Human Avatar

4.2.1 Video Preprocessing

Before we generate avatars, we need to preprocess the videos uploaded by users. Data preprocessing is divided into three steps.

- Break the video into frames.
- Detect body joints from frame by frame.
- Extract human body silhouette and binarize the frame.

The advantage of using video as input rather than pictures is that the video contains more accurate details, such as depth information, which is helpful in building a more realistic human model. Breaking the video into frames is the basis for the rest of the work. As the Fig.4.2 shows, we generate avatar's model based on the video frames which a person is moving around.



Fig. 4.2 The video frames of a person moving around

The human body is made up of multiple joints. In order to make the generated model better match the human body, we need to extract the joints of the human body using openpose [9]. As the Fig.4.3 shows, eighteen body nodes, including the elbow, knee, neck, ankle, wrist and so on, will be detected from each frame.

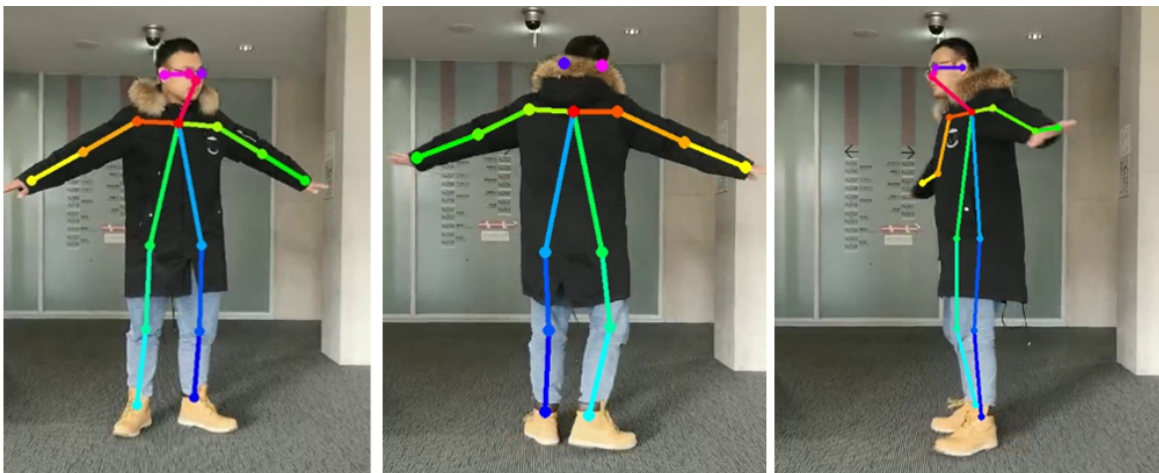


Fig. 4.3 Body joints detected from frames

The detected results will be saved in JSON files for later use.

Before we can formally generate the model, we also need to extract the outline of the human body. As Fig.4.4 shows, by using Baidu API extracting human from image, we can separate the background from the person and binary the segmented image.

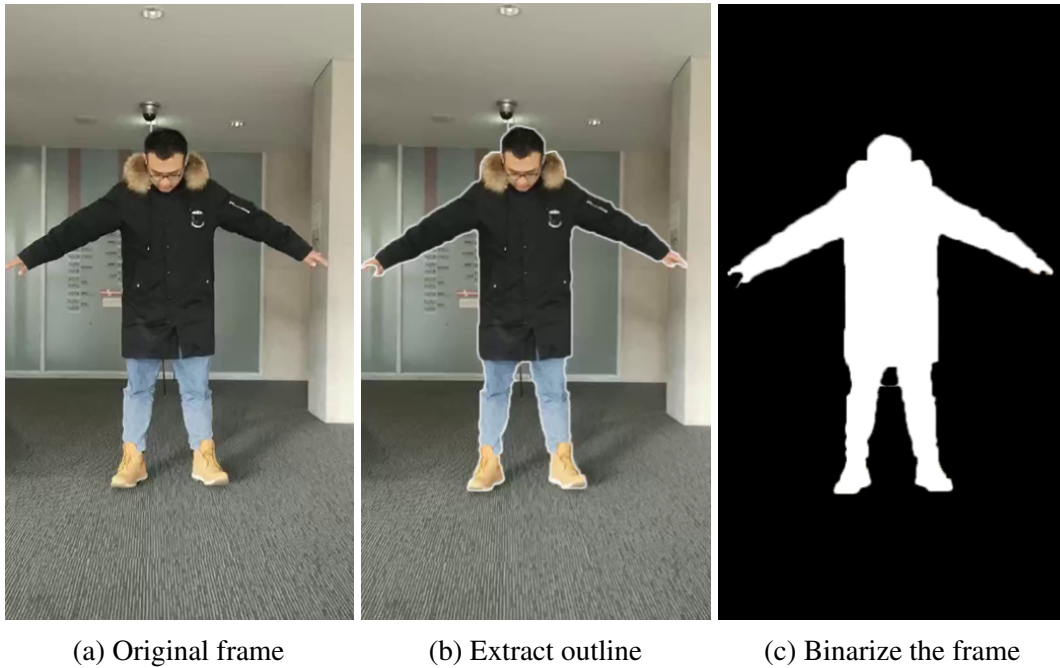


Fig. 4.4 Extract human body silhouette and binarize the frame

4.2.2 Generate Human Model

We use Alldieck's work and SMPL body model to generate the human avatar, including user's hair, body, face and clothes.

Before producing the final full human avatar, we will get two intermediates using the preprocessed data we have generated: the naked body model and the human silhouette model (shown as Fig.4.5). To get naked body model from the video, we adjust SMPL standard model's shape parameters using the calculated results from the video sequence user has uploaded. After that, we can construct human silhouette model based on the naked body model.

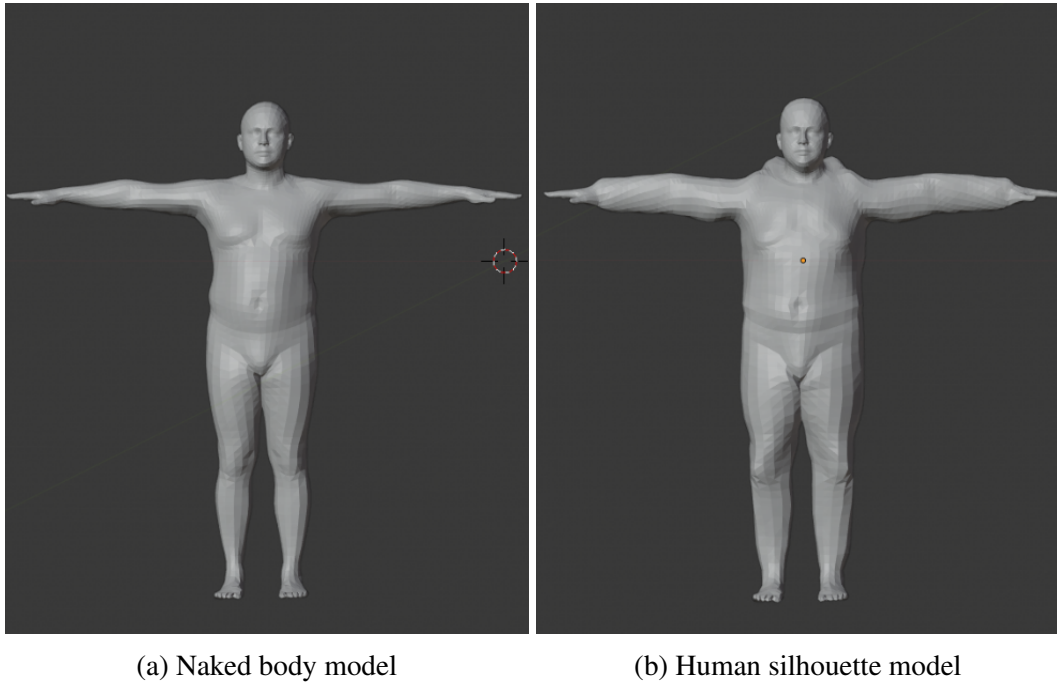


Fig. 4.5 Two intermediates

In order to generate suitable naked body model, we consider SMPL model as the starting point. SMPL is a kind of parameterized human body model which is controlled by 10 shape parameters and 72 pose parameters. Since we are not concerned with the posture of the model, we only consider how to adjust ten body shape parameters.

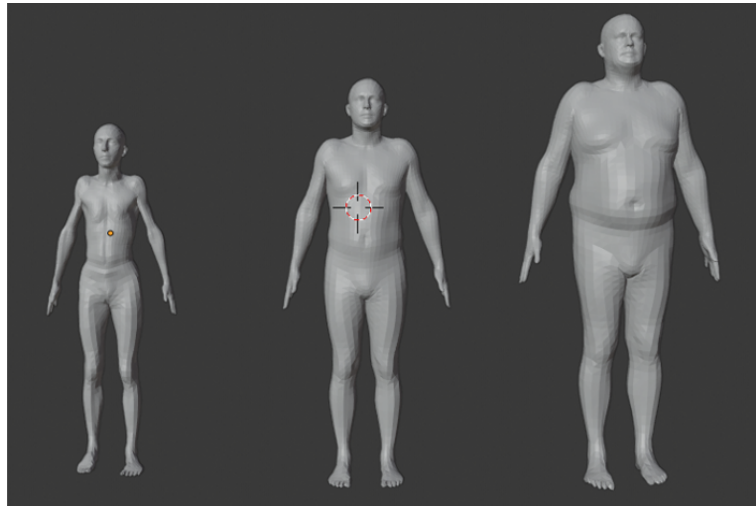
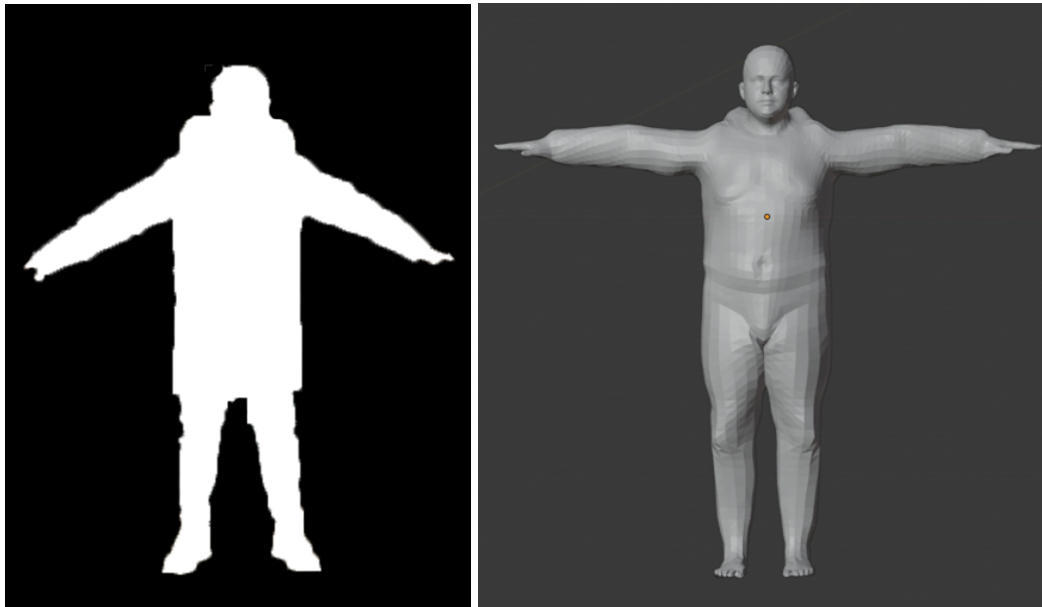


Fig. 4.6 Naked body models with different shape parameter

As the Fig.4.6 shows, 10 different parameters affect different properties of body shape. In this case, we can modify the height, fat and body ratio of the mannequin by changing 10 relatively independent body shape parameters.

Here, we use SMPLify [10] to calculate the value of parameters for each video sequences to fit the body joints detected in the frame. During this frame-by-frame period, 10 parameters are continuously fixed until the final result is generated.

Unlike naked body model, which only cares about shape parameter, silhouette model reconstructs naked body model according to the silhouette of video frames (shown as Fig.4.7). In other words, it modifies not only body shape, but also appearance and clothing. Therefore, 3d shape including face and clothes will be used to reconstruct human model.



(a) The silhouette of a human

(b) Silhouette model

Fig. 4.7 The human body and the silhouette model

4.2.3 Texture

Texture is the crucial part to make the model more realistic. To generate the texture map, the color of the image is projected to the vertices of the body model. After that, UV map between texture and human model will be defined (shown as Fig.4.8). In this way we can get a complete avatar when we specify a texture for our model. However, sometimes because the

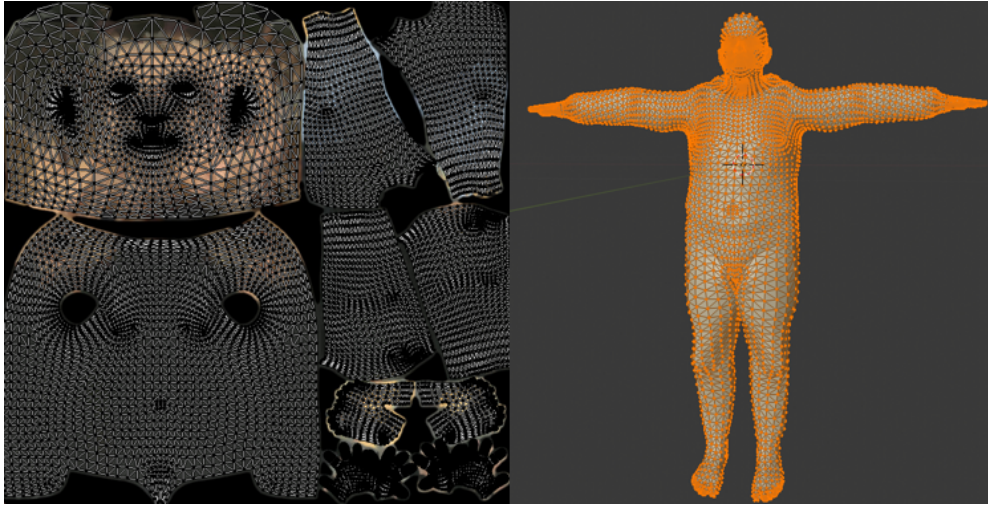


Fig. 4.8 UV map between texture and human model

resolution of the video is not high enough and the user's head accounts for a relatively small proportion of the entire picture, the generated texture is easily blurred. At this point we need to get the texture from the clear face portrait and integrate it into the whole texture. Fig 4.9 shows blurred texture and clear texture on avatar.

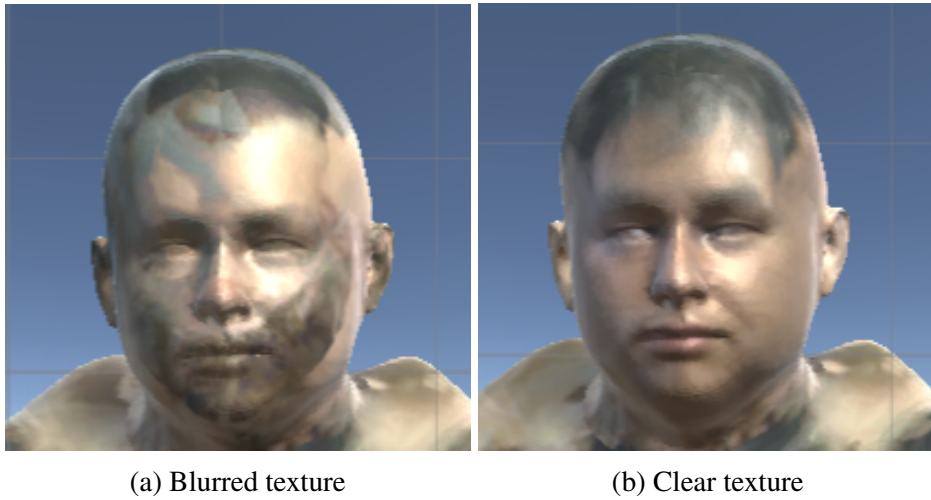


Fig. 4.9 Different textures on the same avatar

4.2.4 Surface Subdivision

After getting the human model and texture, we can optimize the results. As we can see, SMPL is a parameterized model consisting of triangle mesh with 6890 vertices. Too few

vertices make the human body appear too rough and have sharp edges and corners visible to the naked eye. So, we try to refine the surface of the model and increase the number of vertices to make the model smoother. As the Fig.4.10 shows, after surface subdivision, the model is more delicate and realistic in texture.

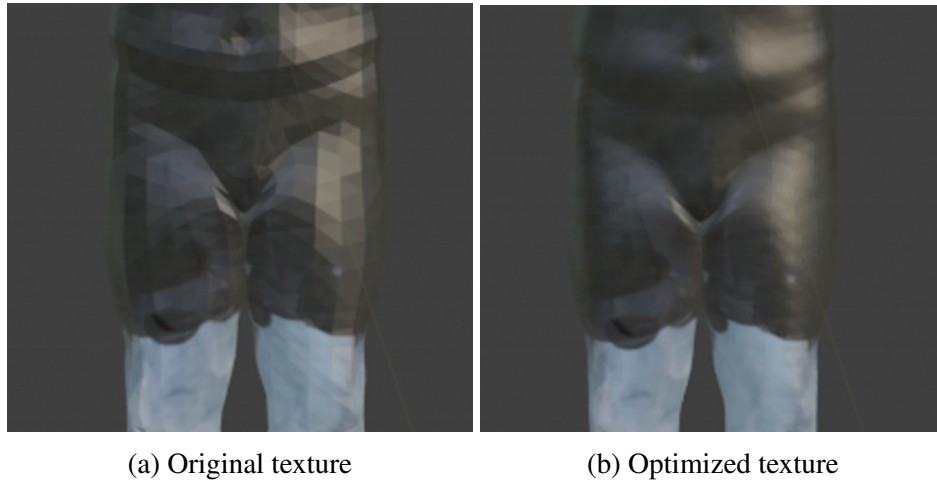


Fig. 4.10 Model before optimization and model after optimization

With the steps mentioned above, we finally got the complete human avatar including clothes, face and hair. The final result is as shown in Fig.4.11.

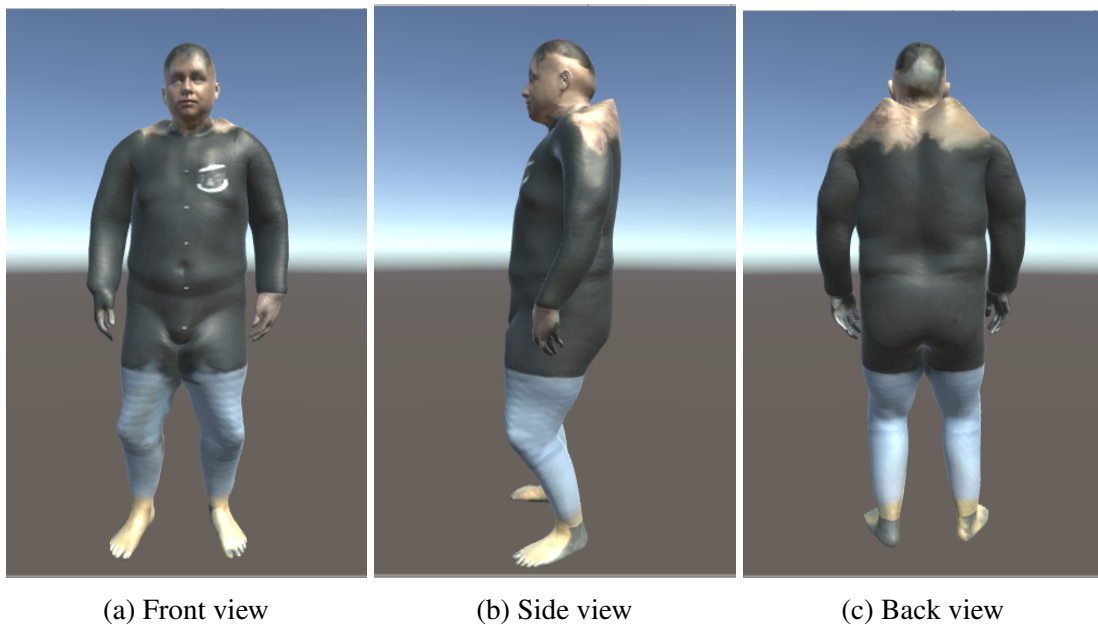


Fig. 4.11 Human avatar from different views

4.3 Interact with Avatar

How to make our avatar interactive or what kind of interaction we need is another question we need to consider. To interact with avatar, first we need to add some body languages and voice to avatar. After that, our human-avatar interaction will be completed by two methods: emotion recognition and speech recognition.

4.3.1 Add Movements and Voice to Avatar

The feedback provided by avatar to users is divided into two types: body language feedback and voice feedback.

For adding body language to the avatar, we need to add the skeleton to the avatar model and make corresponding animation for it.

As the Fig.4.12 shows, we rig the model with 8 key body joints, including groin, wrists, elbows, knees and chin.

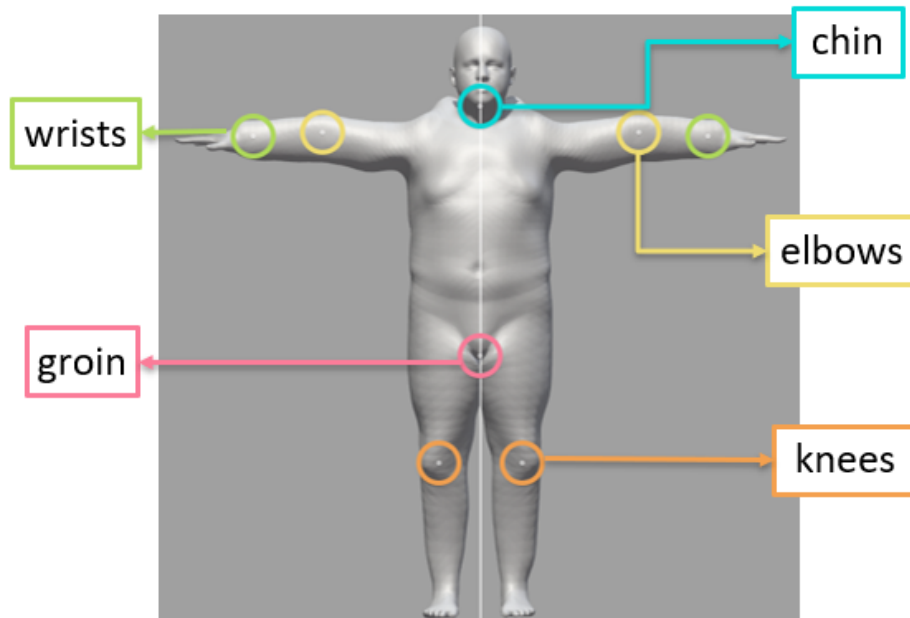


Fig. 4.12 Rig the avatar

Then we can animate avatar using these body joints. Some of the animations are shown as Fig.4.13.

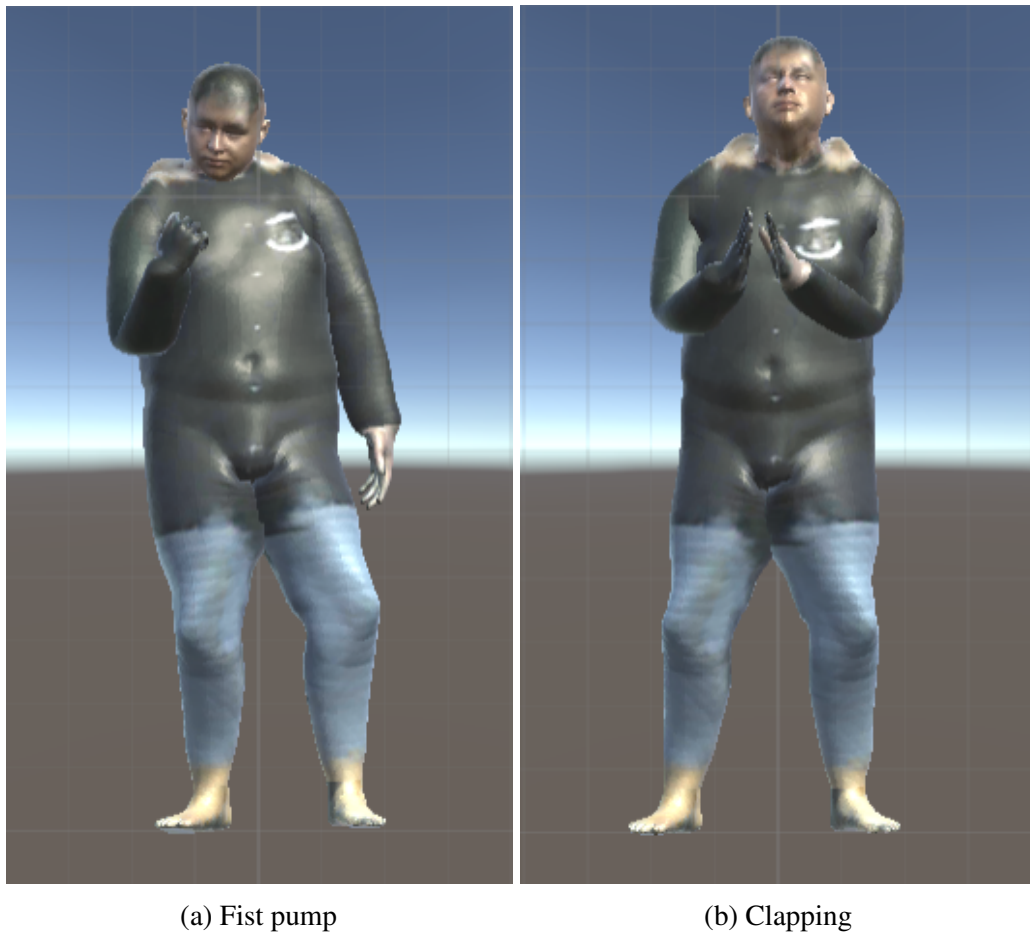


Fig. 4.13 Avatar Animation

In order to add speech to the avatar, text-to-speech system is used to obtain more natural human voice (shown as Fig.4.14). The system input is the text we want to convert, and the system output is some audio clips stored in wav format.

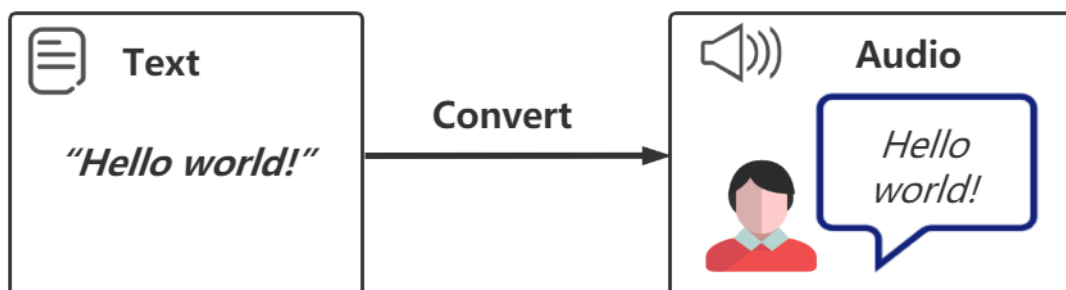


Fig. 4.14 Text-to-speech system

We will coordinate with the use of voice and animation for user feedback.

4.3.2 Emotion Recognition

People's emotions can be expressed in a variety of ways, such as the accent and tone of voice, facial expressions and body movements, etc. If we want to identify and analyze emotions more accurately, we need to consider more than one aspect. Here we choose the user's facial expression and voice as the input of the recognition system. We identified four emotions, calm, happy, sad and angry. In addition, we give each emotion a value ranging from 0 to 50 to express its emotional strength.

To recognize facial expression, we use facial expression recognition API to detect feature points in uploaded face pictures and judge the degree of various emotions based on it. The things user needs to do is to make sure their faces are visible from the phone's camera. As Fig.4.15 shows, four kinds of facial expression will be recognized.

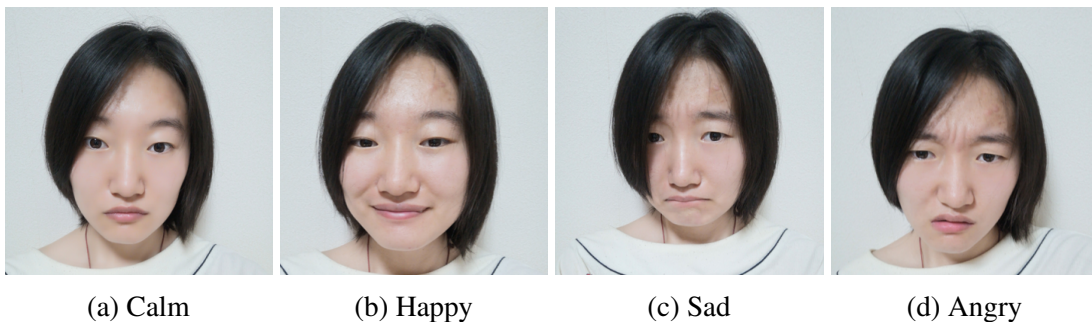


Fig. 4.15 Four kinds of facial expression

To detect users' emotions through voice is actually to detect the physical properties of audio. For example, being short of breath and speaking too fast means being angry. Relative, flat tone and monotone voice could point to calm. Therefore, we use Empath API to identify 4 emotions from voice in real-time regardless of language. Our system will collect a 5-second audio clip every 5 seconds and upload it to the server for detection. The returned results will be stored in the database.

After getting two sets of emotion-related values, we need to combine these two sets of values. When avatar does not detect user voice input, the final result is consistent with the

result of facial expression detection. When the user's voice and facial expression are used as input, we assign a coefficient of 0.5 to the respective results, and then add the two parts to obtain the final result. Fig.4.16 shows a scenario that a specified emotional trend trigger avatar's feedback: The user looked so sad and his voice also sounded sad. And the sad value gradually increases, at this time avatar will give verbal and non-verbal feedback to help users improve their mood.



Fig. 4.16 The specified emotional trend was detected

4.3.3 Speech Recognition

Detecting whether the user's voice contains keywords is another way to trigger avatar feedback. When the user uploads their speech, we will use the voice recognition API to convert the speech to text and then detect whether the text contains some keywords. For example, Fig.4.17 shows our user said he finished his work to avatar, here, work and finished are detected as keywords so that it will trigger avatar's behaviour as the feedback.

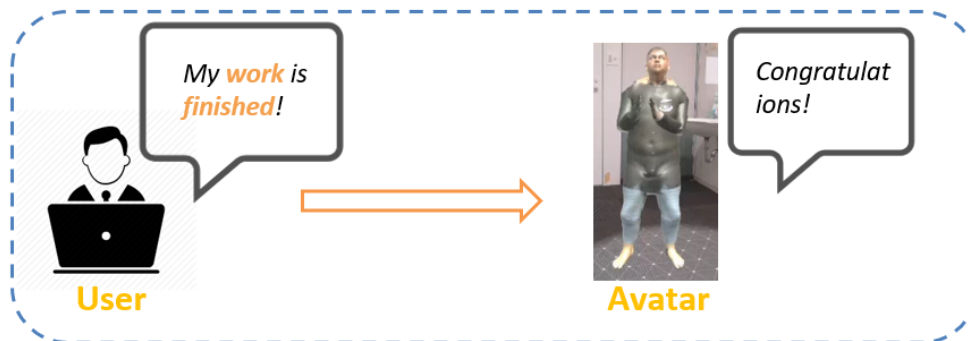


Fig. 4.17 The key words of user's audio clip were detected

4.3.4 Place and Manipulate Avatar in AR Environment

Plane detection and model placement is the first step of all interaction in AR environment. With the help of AR technology, the system can automatically detect the plane in the picture through the phone's rear camera. The detected plane is marked with a white grid. All the user has to do is select their preferred location and click on the screen to place the model in the real world (shown as Fig.4. 18).



Fig. 4.18 Place the avatar into real world

After placement, the user can change the size and direction of the avatar at any time. As Fig.4. 19 shows, users can use two fingers to make the model as large as a real person, or as small as a bottle. Similarly, users can rotate the model by touching the screen with one finger.



(a) Modify avatar's size

(b) Rotate avatar

Fig. 4.19 Avatar Manipulation

4.4 Avatar-Based Work Incentive System

After we generate our personalized human avatar and define the interaction between user and avatar, we give this interactive avatar an application scenario, a work incentive system using avatar as a motivator to boost user's work motivation. We introduced the basic workflow of work incentive system in section 4.1:

1. User upload their video to build their personalized avatar, then they can place avatar in the AR environment and perform basic operations such as zooming in, zooming out and rotating.
2. When the incentive system starts to run, the facial expression and voice of the user will be recorded and recognized in real time. At the same time, system will do emotion analysis.
3. After that, the results of emotion analysis will drive avatar to give feedback to users.

We have already introduced the parts related to avatar. Therefore, in this chapter, we will focus on the emotion analysis and user scene part.

4.4.1 Emotion Analysis

In order to better analyze the emotional trend of users, the collected data are presented intuitively in the form of line chart. As Fig.4.20 shows, four different colored lines represent four different emotions. The horizontal axis represents time, and the two adjacent points differ in value by five seconds. The vertical axis represents the degree of emotions, which ranges from 0 to 50. The smaller the value is, the weaker the degree is, while a larger number indicates a stronger degree.

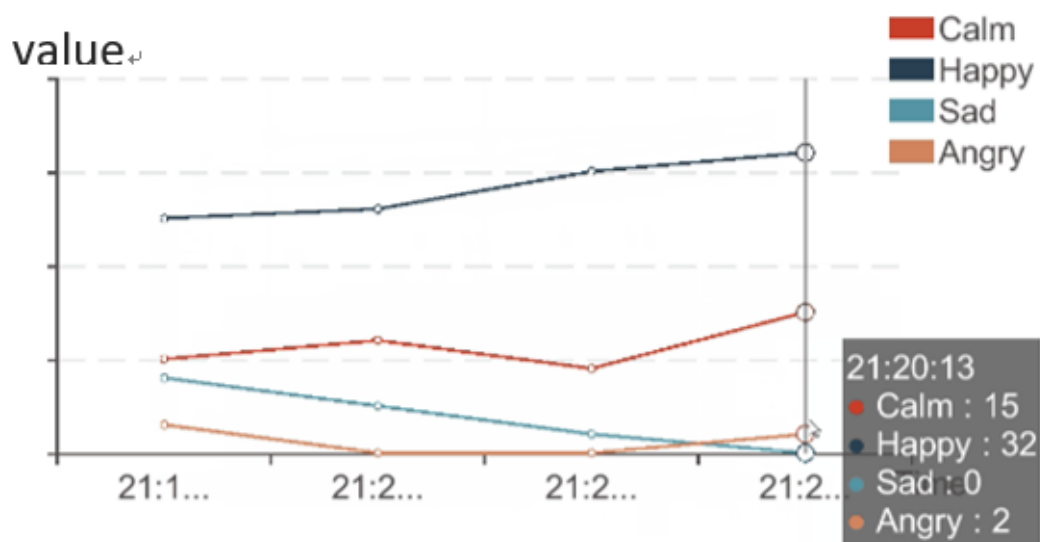


Fig. 4.20 A line chart used to indicate emotions

The system will analyze the emotion recorded every 20 seconds. To do analysis, we first divide the four emotions into two categories: positive emotion including calm and happy, negative emotion including sad and angry. Then we add the trend line of each emotion to analyze user sentiment trends.

The following three questions briefly describe the process of emotion analysis.

1. What is the current trend for each emotion?

2. What is the overall emotional trend?
3. Is the trend good or bad for work?

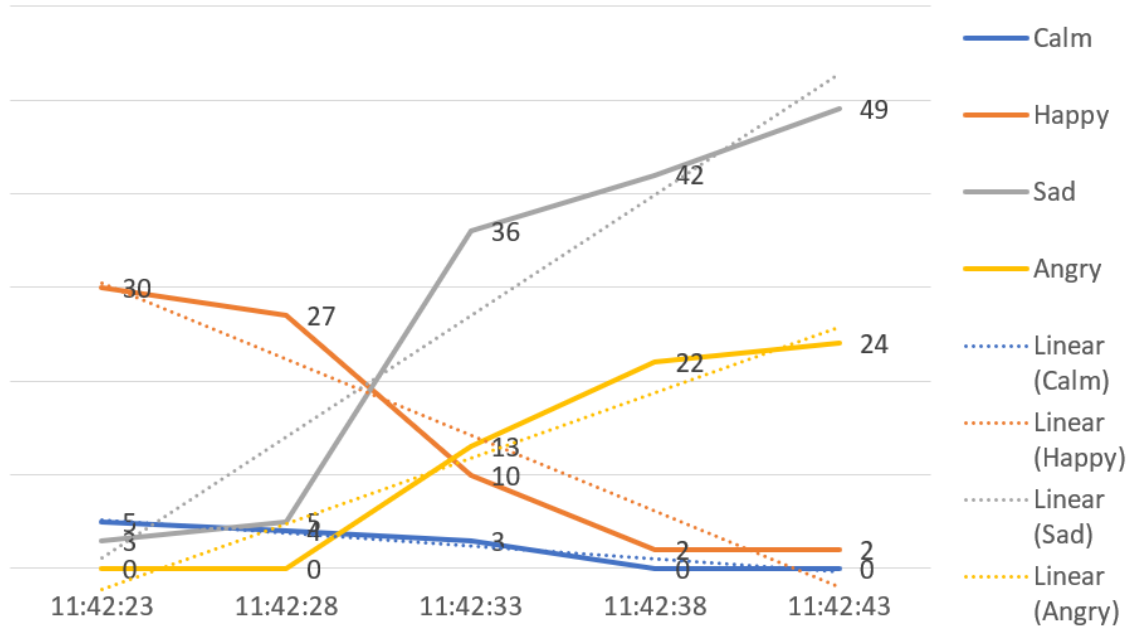


Fig. 4.21 Trend line

Fig.4.21 shows that in addition to the four lines that record emotions, there are four trend lines drawn with dashed lines to predict emotional trends. We further analyze the chart, the trend line of calm tends to be flat and not volatile, the trend line of happy shows a downward trend, and its slope is negative. By combining these two trend lines, we can conclude that positive emotions are gradually decreasing. Corresponding to this, the slopes of the trend lines of sad and angry are positive numbers, indicating an upward trend. Combining these two trend lines, we conclude that negative emotions are gradually increasing. From the above information we get final conclusion that users' emotions are gradually changing from positive to negative, and avatar needs to help users improve their emotions.

4.4.2 Use Scene

The Fig.4.22 shows the main interface after user placing their avatar. There are two kinds of augmented reality information in the interface:

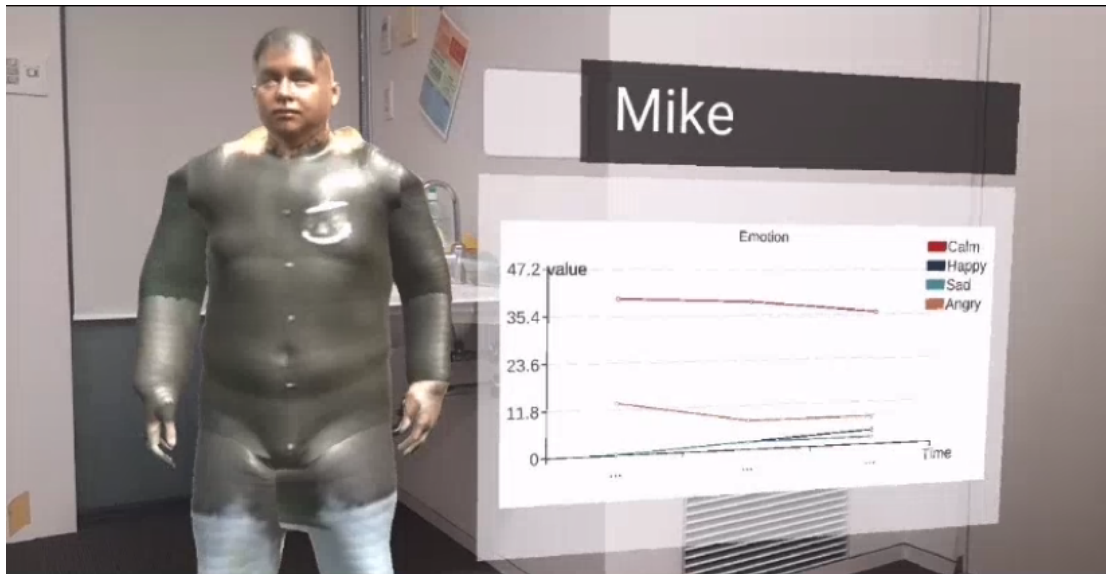


Fig. 4.22 Main interface

The model on the left is the personalized avatar system generated according to video user uploaded. The user can change its size and direction at will. Body language and sound feedback is given by avatar.

The panel on the right is used to visualize emotion data. The system updates data every five seconds. Users can easily check their current work emotional status and adjust themselves timely.

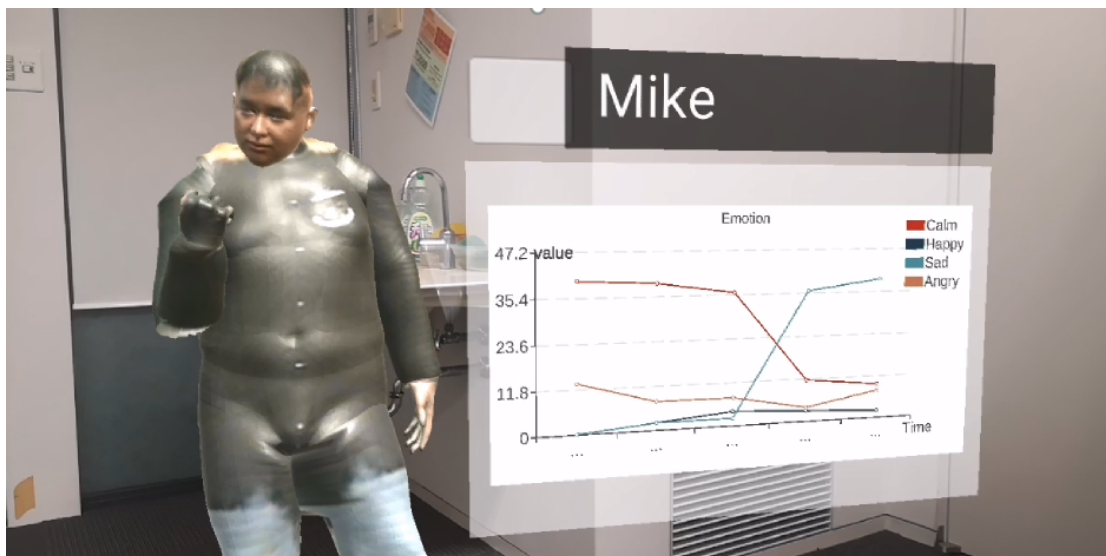


Fig. 4.23 Encourage user

Fig 4. 23 shows the use scene that there is a rapid increase in the value of sad over a short period of time. As can be seen from the line chart in the figure, the anger value increases rapidly, while the calm value decreases significantly. Through sentiment analysis we can conclude that the influence of negative emotions on users gradually increases. The system determines that the user is emotionally unstable and unfit to continue the work. So avatar will encourage user using their movement and speech.

Chapter 5

System Implementation

5.1 Hardware

Some hardware are needed to create our system.

- A smartphone with a front and rear camera.
- A computer to program our system.

To realize our system, the smartphone need to support Android 7.0 at least. Our system needs to recognize facial emotions in real time, so mobile phone cameras are necessary. Here we choose HUAWEI P20 as the device.

And computers need to have some computing power and support android development. The configuration of the computer we use is shown in the Table 5.1.

Category	Information
Operation System	Microsoft Windows 10
CPU	Intel(R) Core(TM) i7-6500U CPU @2.5GHz 2.59GHz
RAM	8 GB

Table 5.1 The information of PC

5.2 Development Environment

We also use some programming tools to complete code writing. The main development tool is Unity 3D, Visual Studio and ARCore. ARCore is a software development kit which allows building augmented reality application. It has been integrated into our smartphone.

The other technical supports are:

- OpenPose, it detects 18 human body joints from video frames and return JSON files.
- Baidu human body analysis API, it extracts the human body outline from the video frame and binarizes it.
- Baidu emotion recognition API, it detects key points in faces from images and uses them to identify emotions.
- Baidu speech API, it converts text to the avatar's voice and convert user's speech to text.
- Epath, it calculates the properties of speech and recognizes the emotions in it.

In the process of generating the avatar, we also used Blender and Mixamo to rig and animate the model.

5.3 Framework

The Fig.5.1 shows how the system framework is.

First, user need to take a video with a person moving around and upload it. Then the system starts to detect the body joints and human outline in the video frames. After that, naked body model and human silhouette model will be generated in turn. Also, system will generate texture for the model.

The information about use's emotion including facial expression and speech will be collected in real time. These information will be sent to API in order to recognize emotion. Then emotion and its value will be stored in database. At the same time, system will start

analysing emotion and then triggers the animation and speech of avatar as the feedback to user.

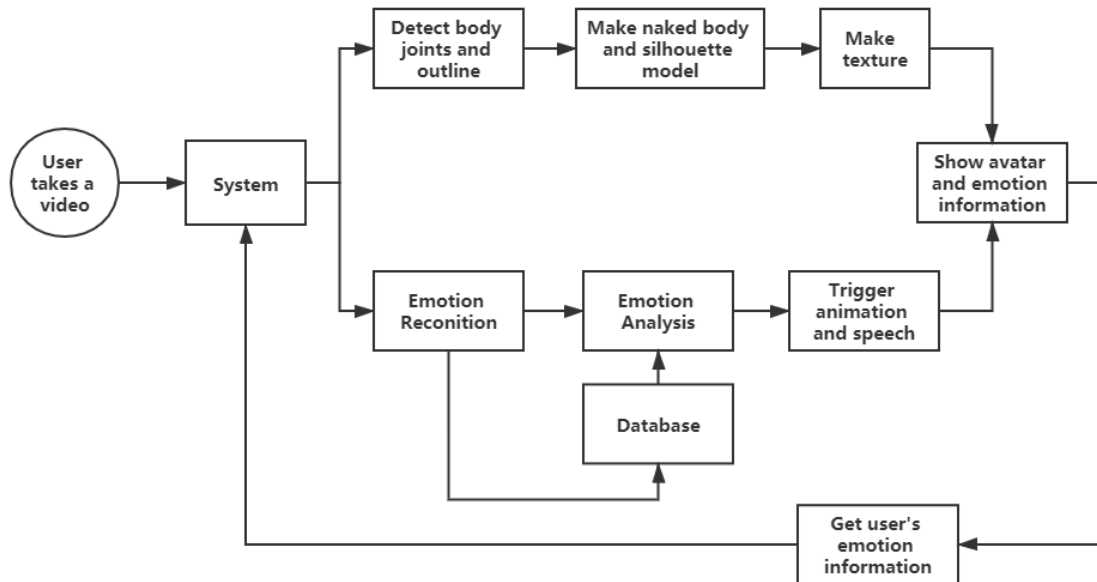


Fig. 5.1 System framework

5.4 Avatar Generation

The avatar generation part is an important part of the whole system. In order to make the generated avatar more realistic and close to the real person, the method we use can be divided into two parts, one is to generate the silhouette model according to the outline of the video frame, the other is to extract the pattern in the frame to generate the final texture.

5.4.1 Model Generation

In setion 4.2, we briefly introduce the preprocessing and steps required to generate the model. The entire process diagram for generating the model is shown in Fig.5.2.

The main steps are:

1. Break the video into frames.

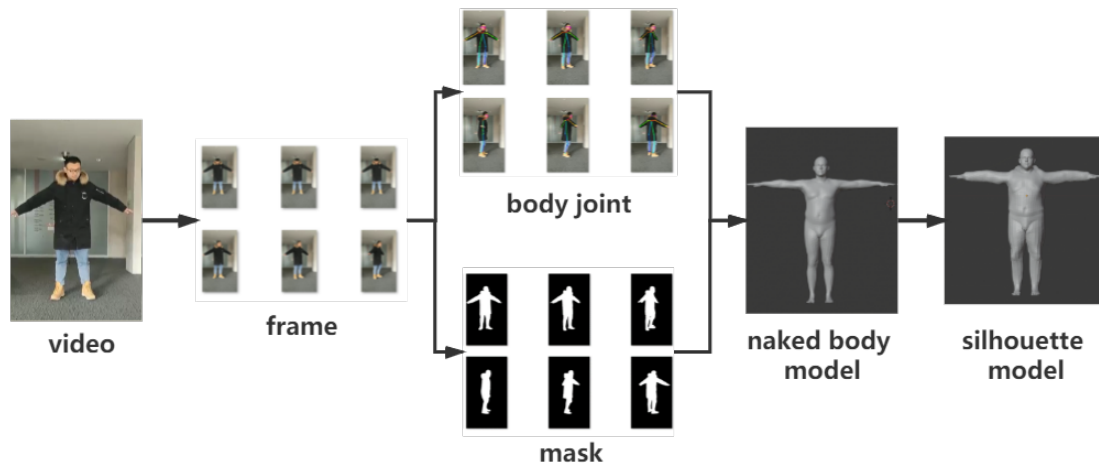


Fig. 5.2 Model generation

2. Detect body joints frame by frame.

In our system, we use OpenPose which provide a windows portable program to detect keypoints. The body joints model comes in many formats, and OpenPose provides three, body25, COCO, and MPI. COCO model is used to present 18 keypoints. Fig.5.3 shows the executable code to get body joints and save the result as the JSON file.

```
OpenPoseDemo.exe --image_dir C:\YUAN_Jingyi\model\sun\frame --model_pose
COCO --write_images C:\YUAN_Jingyi\model\sun\pic --write_json
C:\YUAN_Jingyi\model\sun\joint_coco
```

Fig. 5.3 OpenPose example

3. Extract human body silhouette and binarize the frame.

We use Baidu Body Analysis API to complete this part. Before we use the API, we need to follow the instructions to apply for API key. After this, the platform automatically generates and returns the API key and the corresponding API secret key. Fig.5.4 shows one example of extract human body from background.

4. Generate naked body model and human silhouette model.

```

client = AipBodyAnalysis(APP_ID, API_KEY, SECRET_KEY)

imgfile = 'mask/00020.png'
ori_img = cv2.imread(imgfile)
height, width, _ = ori_img.shape

with open(imgfile, 'rb') as fp:
    img_info = fp.read()

seg_res = client.bodySeg(img_info)
labelmap = base64.b64decode(seg_res['labelmap'])
nparr = np.fromstring(labelmap, np.uint8)
labelimg = cv2.imdecode(nparr, 1)
labelimg = cv2.resize(labelimg, (width, height), interpolation=cv2.INTER_NEAREST)
new_img = np.where(labelimg == 1, 255, labelimg)
cv2.imwrite(imgfile, new_img)

```

Fig. 5.4 Extract body outline and binarize it

To generate a body model, we need to process two sets of data. One is the JSON file that holds the body joints information, and the other is the generated mask file. These two sets of data will be compressed to HDF5 file for later use.

We use the method of All to generate naked model and silhouette model. The core idea is to preprocess the model by selecting 5 frames at medium distance from all video frames. On this basis, the model is optimized frame by frame. After generating the naked model, the system fits the frame array to build the silhouette model.

5.4.2 Texture Generation

Part texture will be got from each frame and will be projected to the silhouette model. In the code, the UV mapping between the model and the texture is already defined, so the generated texture can be used directly on the final avatar. Fig.5.5 shows an example of texture we have generated.

At the same time, the quality of the texture is closely related to the light in the video, the clutter of the background pattern, and the resolution of the image. For better texturing, we can use a white wall or a green screen as the background.



Fig. 5.5 Texture example

5.4.3 Surface Subdivision

The surface of the directly generated model is rougher due to the small number of vertices that generate the model. To make the model more realistic, we need to make the surface smoother. Loop subdivision surface algorithm [11] is used for skinned mesh in our system. It adds a vertex on each edge, and the vertices in the same triangle are connected with new vertices to form a new triangle. As the number of vertices increases, the surface of the model becomes smoother. Fig.5.6 shows the process of surface subdivision.

5.5 Interact with Avatar

We provide nonverbal and verbal interaction for user. For nonverbal feedback given by the avatar, we make a animation controller to switch avatar's movements. For verbal feedback, we preset some text and try to convert these texts to speech.

```

var nmodel = new Model();
for (int i = 0, n = model.triangles.Count; i < n; i++)
{
    var f = model.triangles[i];

    var ne0 = GetEdgePoint(f.e0);
    var ne1 = GetEdgePoint(f.e1);
    var ne2 = GetEdgePoint(f.e2);

    var nv0 = GetVertexPoint(f.v0);
    var nv1 = GetVertexPoint(f.v1);
    var nv2 = GetVertexPoint(f.v2);

    nmodel.AddTriangle(nv0, ne0, ne2);
    nmodel.AddTriangle(ne0, nv1, ne1);
    nmodel.AddTriangle(ne0, ne1, ne2);
    nmodel.AddTriangle(ne2, ne1, nv2);
}

```

Fig. 5.6 Surface subdivision

5.5.1 Animation Controller

We preset few movements for our avatar, such as stand on the plane, clap hands, fist pump and nod head. Therefore, an animation controller is needed to control when and what movement of avatar to trigger.

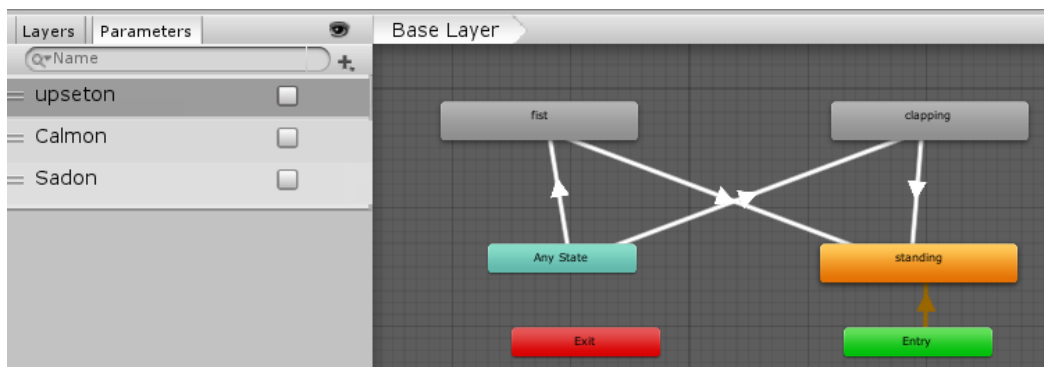


Fig. 5.7 Animation controller

As we can see from Fig.5.7, the avatar has an initial default animation. When the user places the model on a plane for the first time, or when it is currently in a non-interactive state, the avatar will loop the standing movement. Our system can detect four kinds of emotion

and we have already introduced emotion trend line in our system in section 4.4.1. The design of animation controller is related to emotion analysis.

For example, the current user is in a sad mood, which is negative. Moreover, negative emotions are growing rapidly and it is not good for the user, so we need to switch to fist pump movement to encourage user.

5.5.2 Add Voice to Avatar

Verbal feedback is another interaction we provide for user. We implement this part using Baidu text-to-speech API. At first, we should get our API key and secret key from their official website. Then, the speed, volume and the tone of the speech should be set. We completed this part by referring to the official API documentation. Also, text that needs to be converted should be saved in a txt file. Fig. 5.8 shows the process we convert text to speech.

```
def tts(str, id):
    client = AipSpeech(APP_ID, API_KEY, SECRET_KEY)
    result = client.synthesis(str, 'zh', 1, {'spd': 7, 'vol': 9, 'per': 0, 'aue': 103})
    filename = 'vocal/audio'
    if not isinstance(result, dict):
        with open(filename + '{}'.format(id) + '.wav', 'wb') as f:
            f.write(result)
```

Fig. 5.8 Convert text to speech

As we have already introduced in section 4.3, real time facial expression recognition and voice recognition are needed in our system. We use Baidu emotion recognition API and Empath API to do that.

5.5.3 Facial Expression Recognition

The facial expression API recognizes users' emotions by detecting 72 key points in their faces. The value and possibility of emotion will be returned as a part of face attributes which will be stored as the JSON format. To detect the user's facial expressions in real time, we take an image from the smartphone's camera every five seconds and converted it to Base 64 format for processing. After getting the results, we need to extract the emotional data we

need from the JSON data and store it in the database. Fig.5.9 shows the example of facial expression recognition.

```
string image = System.Convert.ToBase64String(image64);
var image_type = "BASE64";
var options = new Dictionary<string, object>{
    {"face_field", "emotion"}
};
try
{
    var result = client.Detect(image, image_type, options);
    Debug.Log(result);
    string[] msgArr = result.ToString().Split(',');
    //extract classname
    for (int i = 0; i < msgArr.Length; i++)
    {
        if (msgArr[i].Contains("emotion"))
        {
            string[] strArr = msgArr[i].Split(':');
            detectedMotionMsg.text = strArr[2];
            //store database
            InsertFacialEmotion(strArr[2]);
            break;
        }
    }
}
```

Fig. 5.9 Facial expression recognition

5.5.4 Voice Recognition

Similar to facial expressions, voice recognition calculates the physical properties of audio to determine a user's emotion. We use Empath API to do that. To detect the user's speech in real time, the system automatically records a five-second audio recording every five seconds and uploads it via an HTTP request. When API returns the result, we convert the JSON data into empath data format and extract what we need from it. After that, the emotional data will be stored into database. Fig.5.10 shows an example of voice recognition.

5.5.5 Multiple Aspects Emotion Recognition

As Fig5.11 shows, after getting initial results from facial expression recognition API and vocal recognition API, we need make a combination for these two aspects to get the final

```
IEnumerator Upload(byte[] wavbyte)
{
    WWWForm form = new WWWForm();
    form.AddField("apikey", empath_apikey);
    form.AddBinaryData("wav", wavbyte);
    string receivedJson = null;

    using (UnityWebRequest www = UnityWebRequest.Post
        ("https://api.webempath.net/v2/analyzeWav", form))
    {
        yield return www.SendWebRequest();
    }

    EmpathData empathData = ConvertEmpathToJson(receivedJson);
    empathResult.text = ConvertEmpathDataToString(empathData);
}
```

Fig. 5.10 Voice recognition

result. We assign 0.5 as coefficient to two different initial recognition results, then calculate the final results.

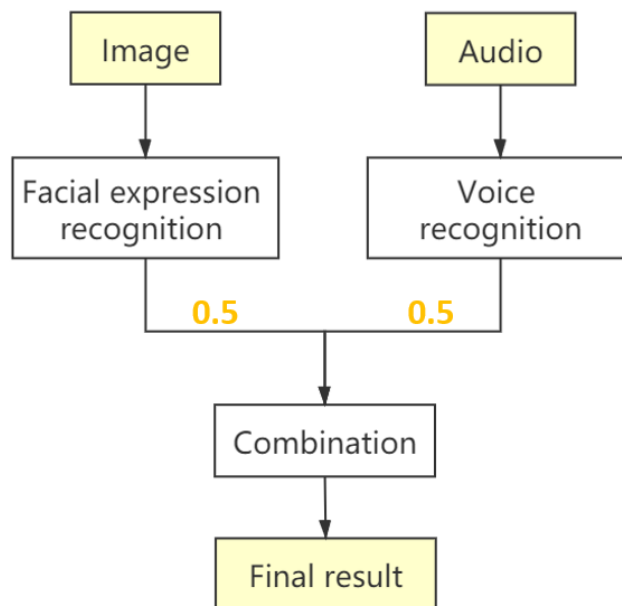


Fig. 5.11 Combine two initial result

5.5.6 Speech Recognition

We use speech recognition to detect if there are key word in user speech. First thing we need to do is processing recording data to the format of PCM16 encoding. Then we convert recording speech to text. The API will return the result of recognition and we need to detect if the text contains key words we need. Fig.5.12 shows the example of speech recognition.

```
UnityWebRequest unityWebRequest = UnityWebRequest.Post(url, wwwForm);

unityWebRequest.SetRequestHeader("Content-Type", "audio/pcm;rate=" + recordFrequency);

yield return unityWebRequest.SendWebRequest();

if (string.IsNullOrEmpty(unityWebRequest.error))
{
    asrResult = unityWebRequest.downloadHandler.text;
    Debug.Log(asrResult);

    if (Regex.IsMatch(asrResult, @"err_msg.:. success"))
    {
        Match match = Regex.Match(asrResult, "result:... (.*)...");
        if (match.Success)
        {
            asrResult = match.Groups[1].ToString();
        }
    }
}
```

Fig. 5.12 Speech recognition

5.6 Store Database

To store the emotional data for analysis, we designed a simple database with two separate entities. One is used to store data related to facial expressions, and the other is used to store data related to sounds.

5.6.1 Entities

- **Facial expression** Facial expression entity refers to the data we captured using facial expression recognition API. Table 5.2 shows this entity stores information including time and four emotion value.

Name	Type	Description
ID	INTEGER	The ID of the facial expression.
Time	TIME	The system time when the data is recorded.
Calm	INTEGER	A value used to indicate the degree of calm.
Happy	INTEGER	A value used to indicate the degree of happy.
Sad	INTEGER	A value used to indicate the degree of sad.
Angry	INTEGER	A value used to indicate the degree of angry.

Table 5.2 Facial expression table

Name	Type	Description
ID	INTEGER	The ID of the voice.
Time	TIME	The system time when the data is recorded.
Calm	INTEGER	A value used to indicate the degree of calm.
Happy	INTEGER	A value used to indicate the degree of happy.
Sad	INTEGER	A value used to indicate the degree of sad.
Angry	INTEGER	A value used to indicate the degree of angry.

Table 5.3 Voice table

- **Voice** Voice entity refers to the data we captured using Empath API. Table 5.3 shows this entity stores information including time and four emotion value.

5.6.2 Connect Database and Store Data

In our system, we choose SQLite as our database engine. In order to use SQL databases in Unity, we need to define our own utility classes for adding, deleting, and modifying data. Fig.5.13 shows the class used to insert data to specified table.

```

public SqliteDatabase InsertValues(string tableName, string[] values)
{
    //Gets the number of fields in the table
    int fieldCount = ReadFullTable(tableName).FieldCount;
    //An exception is thrown when the inserted data length is not equal to the number of fields
    if (values.Length != fieldCount)
    {
        throw new SqliteException("values.Length!=fieldCount");
    }
    string queryString = "INSERT INTO " + tableName + " VALUES (" + values[0];
    for (int i = 1; i < values.Length; i++)
    {
        queryString += ", " + values[i];
    }
    queryString += " )";
    return ExecuteQuery(queryString);
}

```

Fig. 5.13 Insert data to specified table

Fig.5.14 shows an example of SQLite database storing emotional data.


	 ID	Time	Calm	Happy	Sad	Angry
1	1	2020-05-22 21:19:58	10	25	9	3
2	2	2020-05-22 21:20:03	12	26	5	0
3	3	2020-05-22 21:20:08	9	30	2	0
4	4	2020-05-22 21:21:13	15	32	0	2
5	5	2020-05-22 21:21:18	10	40	0	4

Fig. 5.14 SQLite database

5.6.3 Update Line Chart

when the database store new data, we also need to update our line chart. To update it, we need to define query sentence so that emotion data will be read from form in database.

Fig.5.15 shows the process we define query sentence and read data from form.

```

SqliteDataReader reader;

reader = sql.ReadTable("face_emotion", new string[] { "Calm", "Happy", "Sad", "Angry" },
    new string[] { "ID" }, new string[] { "=", "=" }, new string[] { i.ToString() });

while (reader.Read())
{
    emotion_value[0] = reader.GetInt32(reader.GetOrdinal("Calm"));
    emotion_value[1] = reader.GetInt32(reader.GetOrdinal("Happy"));
    emotion_value[2] = reader.GetInt32(reader.GetOrdinal("Sad"));
    emotion_value[3] = reader.GetInt32(reader.GetOrdinal("Angry"));
    Debug.Log("Update emotion line chart.");
}

```

Fig. 5.15 Read data

5.7 Augmented Reality

We need to detect the plane in the real world to put our avatar on it. We use ARCore to recognize the environment and detect the platform we needed. To make use of ARCore, we should download ARcore SDK for Unity and configure build setting.

At first, a plane prefab need to be set in Unity. Then we should set a texture for this plane, which is the grid shown to user when they detect the plane in real world. Also, we need to add the plane render script which ARcore has already provided for us. At the same time we need to specify the object that we want to place, which is use's customized avatar. Fig.5.16 shows configure the plane in Unity.

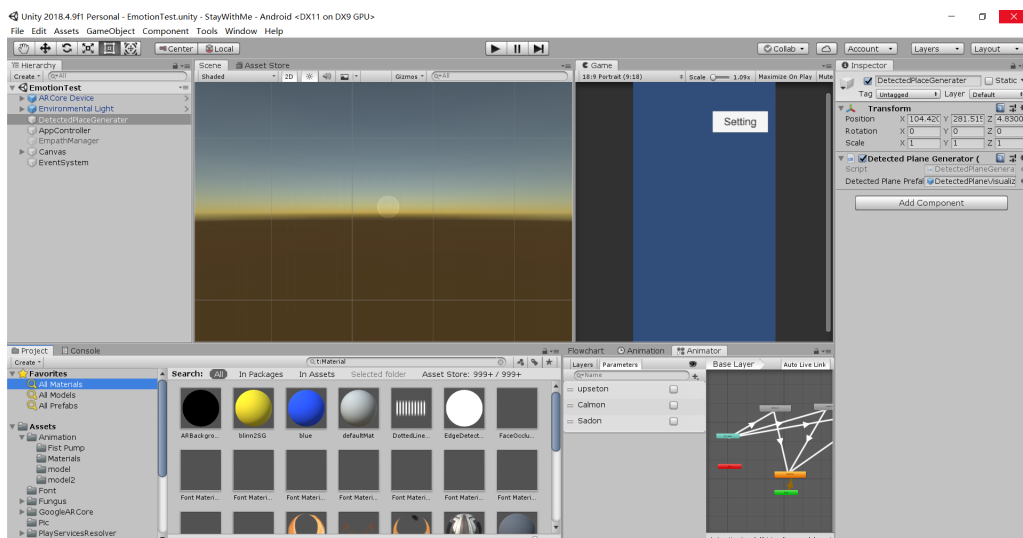


Fig. 5.16 Configure plane in Unity

Chapter 6

Related Work

In this chapter, we will introduce related work about our system. Our research focuses on the role of avatar in motivating users' work motivation. The main method is to detect the user's emotion in real time and give feedback. Therefore, there will be three kinds of related work. The first is the work about personalized human avatar. The second is about emotion recognition and analysis. The third is about work incentive system.

6.1 Related Work on Personalized Human Avatar

With the development of HCI field, the avatars have become increasingly used in virtual communities to provide users with a more natural and favorable online experience, so promoting a close human-avatar relationship has been a necessity to the sustainable development of virtual communities [12]. Such avatars that can communicate well with user can replace human beings to complete some work, such as helping users control obesity by detecting changes in human body shape [13] or use them to augment language exposure for 6-12 month old infants [14]. These studies show that well-designed human-avatar interaction can create a more efficient and engaging communication environment than a normal user interface. Realistic avatars are the basis of interaction which can give user a sense of reality.

So how to generate a more human-like avatar will be the first problem. As we have introduced in section 2.1, there are three common ways building avatar. The first method uses

a 3D scanning device similar to Kinect. The work of Jing Tong et al. [15] provides a method which captures 3D full human body models by using multiple Kinects. The disadvantages are obvious. That is, users need to have multiple hardware devices to capture different parts of the body. Therefore, it is expensive and inconvenient for ordinary users. The second method builds human avatar by adjust the parameter of user's body and head. This approach is often used in large games where players can customize their own models to quickly blend into the game environment and feel a strong sense of autonomy and control [16]. However, this does not apply to our system, as we need to not only customize the mannequin but also add corresponding texture to the model.

The third method uses machine learning technology to reconstruct human body and appearance based on an image or a video. Zhong Li et al. proposed a pipeline [17] that reconstructs 3D human shape avatar at a glance. The avatars generated by this method often include not only the body, but also other human attributes, such as appearance, hair color and clothing. But the drawback of this approach is obvious. It does not contain deep information about the human. One solution to this is to use video instead of pictures which can get more comprehensive information than pictures. Also, high-quality avatars should feature a natural face, hairstyle, clothes with garment wrinkles, and high-resolution texture [18]. The method we use meets the above two conditions and can generate a realistic avatar based on video.

In addition to entertainment such as video game enjoyment, the avatars are also used in other research fields. In terms of virtual shopping, JKapur et al. [19] proposed a method to help users visualize how a wearable article will look on the their bodies. WonSook Lee et al. [20] allow user to user their photo and body shape parameter as the input to see how themselves dressed.

Avatars are also often used in research related to psychology and health. Kiani and Massi Joe E [21] use avatar to encourage improvement of physical health and academic performance. Wenbing Zhao et al. [22] devotes to build the next generation virtual avatar-based life coaching system for children with Autism Spectrum Disorder (ASD).

Personalized virtual avatar can also be used for social communication. Zichun Guo et al. [23] develop a virtual communication system for disabled people to improved their face-to-face communication ability.

6.2 Related Work on Emotion Recognition and Analysis

Interacting with others by reading their emotional expressions is an essential social skill in humans [24] which can also apply to avatars. So we add emotion recognition and analysis mechanism into our system in order to make avatar better perceive the changes of users' working state. Emotions can be sensed through facial expression, speech and physiological sensory processes. Facial expression and perception have long been the primary emphasis in emotion recognition field. However, there is a growing interest in other channels like the voice and touch [25].

Facial expressions and speech are expressive way humans display emotions [26]. We recognize the facial expression and voice through detecting the keypoints and physical attributes. Here are some related work about emotion recognition. Caridakis, George et al. [27] proposed and trained a model which can recognize 8 kinds of emotion from facial expression, gestures and speech. Yu, Chuang et al. [28] propose a emotion recognition framework for robot using multimodel data from facial expressions and human gait data.

Chapter 7

Preliminary Evaluation

In this chapter, we introduce our preliminary evaluation method and the analysis results to evaluate the avatar. The main purpose of this evaluation is to test whether our human avatar is realistic or satisfying enough and whether the avatar can boost user's motivation to work by interacting with it. 8 participants are asked to fill in a questionnaire. After that, we will discuss the results from the questionnaire.

7.1 Participants

8 participants aged from 20 to 26 are invited to conduct the experiments. All of them have the experience with of AR application and a few of them have experience with virtual avatar.

7.2 Method

We give a brief introduction of the avatar and our system to all participants. Then the participants need to try to use our system at work. Then our system will use avatar to detect their emotion and help them adjust their bad moods.

After using the system, participants will be asked to fill out a five-question questionnaire. The 5 questions are shown as below.

1. The avatar looks realistic from the following aspects.
 - Body shape
 - Clothes
 - Face
2. It is useful or interesting to interact with avatar.
3. The avatar can understand or recognize emotion easily.
4. The avatar can give appropriate feedback to boost work motivation.
5. The avatar can work as a motivator effectively.

Fig.7.1 shows our questionnaire using 5-point Likert scale.

7.3 Result

Our participants were asked to rate each question and the results we gathered shown as the following table 7.1 and Fig.7.2.

Question	1	2	3	4	5
Q1-1:The avatar looks realistic from the body shape.				5	3
Q1-2:The avatar looks realistic from the clothes.			2	5	1
Q1-3:The avatar looks realistic from the face.				6	2
Q2:It is useful or interesting to interact with avatar.					8
Q3:The avatar can understand or recognize emotion easily.				5	3
Q4:The avatar can give appropriate feedback to boost work motivation.				4	4
Q5:The avatar can work as a motivator effectively.				6	2

Table 7.1 Investigative questions after using the system

Question 1 is used to test how realistic our virtual avatar looks like. We test the model from three aspects: body shape, clothes and face. The average score are 4.375, 3.875 and 4.25 respectively. The result proves that most participants think our avatar looks very similar

QUESTIONNAIRE

Name: Age: Gender: Date:

QUESTIONS

The questions are based on 5-point scale.

Answer the following questions by circling the most appropriate answer.

1. The avatar looks realistic from the following aspects.

➤ **Body shape**

Strongly Disagree Disagree Neutral Agree Strongly Agree

➤ **Clothes**

Strongly Disagree Disagree Neutral Agree Strongly Agree

➤ **Face**

Strongly Disagree Disagree Neutral Agree Strongly Agree

2. It is useful or interesting to interact with avatar.

Strongly Disagree Disagree Neutral Agree Strongly Agree

3. The avatar can understand or recognize emotion easily.

Strongly Disagree Disagree Neutral Agree Strongly Agree

4. The avatar can give appropriate feedback to boost work motivation.

Strongly Disagree Disagree Neutral Agree Strongly Agree

5. The avatar can work as a motivator effectively.

Strongly Disagree Disagree Neutral Agree Strongly Agree

How could avatar or system be improved?

Fig. 7.1 Questionnaire

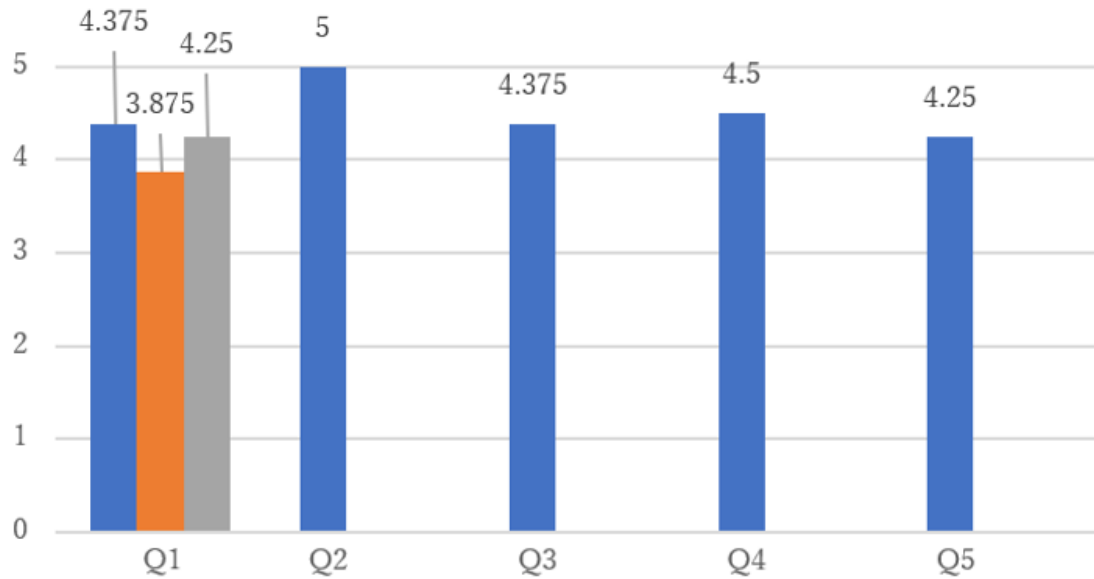


Fig. 7.2 Questionnaire results

to the real people. But two participants think it would be better if the clothes on avatar looks more natural.

Question 2 is used to determine if the interaction between user and avatar can give a good experience to the user. The average score of question 2 is 5. It proves that all the participants agree that it is useful or useful to interact with avatar.

Question 3, question 4 and question 5 are used to judge whether the personalized avatar can boost user's motivation in work incentive system. Question 3 is used to test if our avatar can understand user's emotion easily and the average score is 4.375. The average scores of question 4 and question 5 are 4.5 and 4.25 which shows our avatar works well in work incentive system.

Overall, we got relatively positive feedback from our participants. We also got some feedback and suggestions from our participants:

- I think if the avatar could give some facial expression during the interaction will be good.

-
- The voice can be improved. You can use voice from the user as its avatar's voice, which may make the interactions more intuitive.
 - Maybe you can add a variety of different types of feedback. Also, add more interaction with avatar, according to user's emotion, avatar can give different response. For example, when user feel sad, avatar can sing a song for user.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

In this thesis, our research tries to build an interactive personalized human avatar. On this basis, an avatar-based work incentive system is constructed.

The avatar will be generated with the realistic body shape, lifelike appearance and clothes and it is based on the video user uploaded. In order to make the model interact with people more naturally, we also need to add some movements and voice to the avatar. After that, we build an application which uses personalized avatar as the motivator to boost their work motivation by adjust their mood.

Technically, we build a video-based human model with the technical support of Alldieck's work. Then we do some optimization after getting the basis model, including optimizing the model's surface using loop subdivision algorithm and generating a texture for the model to get the final avatar. Then the avatar will be rigged to make the movements and also given the voice. We also design two kinds of user behavior which can trigger avatar's feedback. One is to detect if there are key words in user's speech. Another is to judge whether there is a specified emotion trends when user is at work.

User can consider the avatar as the second incarnation of anyone they want. On this basis, we build a work incentive system using the avatar as a motivator. The application uses

user's facial expression and voice as the input. These data will be collected and analyzed in real time. The results of the analysis will drive the feedback of avatar.

Last, preliminary evaluation was performed to test how realistic our virtual avatar looks like and if the interaction between user and avatar can give a good experience to user.

8.2 Future Work

The directions we are going to work on is shown as below:

- Although we have generated a realistic avatar based on a real person, the face of the avatar is still need to be improved in terms of the feedback we got from our participants.
- We want the avatar to look as real as possible. Therefore, it is better to add various facial expression such as smile and frown to avatar.
- We can generate avatar's voice based on the real person's voice, which may make the interactions more intuitive.
- Besides, the interaction between user and avatar can be more diversified.

We focus on build an human like avatar, so our next work will focus on how to build a more realistic model on the existing basis.

References

- [1] Guo Freeman, Samaneh Zamanifard, Divine Maloney, and Alexandra Adkins. My body, my avatar: How people perceive their avatars in social virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8, 2020.
- [2] John M Gibbons. Employee engagement: A review of current research and its implications. Conference Board, 2006.
- [3] Kori M Inkpen and Mara Sedlins. Me and my avatar: exploring users’ comfort with avatars for workplace communication. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 383–386, 2011.
- [4] Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, and Gerard Pons-Moll. Video based reconstruction of 3d people models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8387–8397, 2018.
- [5] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6):1–16, 2015.
- [6] Hu Chen, Ye Gu, Fei Wang, and Weihua Sheng. Facial expression recognition and positive emotion incentive system for human-robot interaction. In *2018 13th World Congress on Intelligent Control and Automation (WCICA)*, pages 407–412. IEEE, 2018.
- [7] Rashid Ahmed Khamis Al Naqbi PROF, Rosman Bin Md Yusoff Dr Fadillah, and Binti Ismail. The effect of incentive system on job performance motivation as mediator for public sector organization in uae. *International Journal of Engineering & Technology*, 7(4.7):380–388, 2018.
- [8] Empath. Vocal emotion recognition test by empath. <https://webempath.net/lp-eng/>. Accessed April 4, 2020.
- [9] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1812.08008*, 2018.
- [10] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In *European Conference on Computer Vision*, pages 561–578. Springer, 2016.

- [11] Charles Loop. Smooth subdivision surfaces based on triangles. *Master's thesis, University of Utah, Department of Mathematics*, 1987.
- [12] Yi Zhao and Weiquan Wang. Attributions of human-avatar relationship closeness in a virtual community. In Miltiadis D. Lytras, John M. Carroll, Ernesto Damiani, and Robert D. Tennyson, editors, *Emerging Technologies and Information Systems for the Knowledge Society*, pages 61–69, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
- [13] Angelos Barmpoutis. Automated human avatar synthesis for obesity control using low-cost depth cameras. *Studies in health technology and informatics*, 184:36–42, 2013.
- [14] Brian Scassellati, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkuhn, Barbara Manini, Adam Stone, Geo Kartheiser, Arcangelo Merla, Ari Shapiro, David Traum, and Laura-Ann Petitto. Teaching language to deaf infants with a robot and a virtual human. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, page 1–13, New York, NY, USA, 2018. Association for Computing Machinery.
- [15] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3d full human bodies using kinects. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):643–650, 2012.
- [16] Keunyeong Kim, Michael G Schmierbach, Mun-Young Chung, Julia Daisy Fraustino, Frank Dardis, Lee Ahern, et al. Is it a sense of autonomy, control, or attachment? exploring the effects of in-game customization on game enjoyment. *Computers in Human Behavior*, 48:695–705, 2015.
- [17] Zhong Li, Lele Chen, Celong Liu, Yu Gao, Yuanzhou Ha, Chenliang Xu, Shuxue Quan, and Yi Xu. 3d human avatar digitization from a single image. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [18] Chulhan Lee, Hohyun Lee, and Kyoungsu Oh. Real-time image-based 3d avatar for immersive game. In *Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '08, New York, NY, USA, 2008. Association for Computing Machinery.
- [19] Jay Kapur, Sheridan Jones, and Kudo Tsunoda. Avatar-based virtual dressing room, May 9 2017. US Patent 9,646,340.
- [20] Frédéric Cordier, Wonsook Lee, Hyewon Seo, and Nadia Magnenat-Thalmann. Virtual-try-on on the web. *Laval Virtual*, 2001.
- [21] Massi Joe E Kiani. Avatar-incentive healthcare therapy, May 7 2019. US Patent 10,279,247.
- [22] Wenbing Zhao, Xiongyi Liu, Tie Qiu, and Xiong Luo. Virtual avatar-based life coaching for children with autism spectrum disorder. *Computer*, 53(2):26–34, 2020.

- [23] Zichun Guo, Xueguang Jin, and Rui Hao. Avatar social system improve perceptions of disabled people's social ability. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pages 483–488, 2019.
- [24] Tobias Grossmann. The development of emotion perception in face and voice during infancy. *Restorative neurology and neuroscience*, 28(2):219–236, 2010.
- [25] Annett Schirmer and Ralph Adolphs. Emotion perception from face, voice, and touch: Comparisons and convergence. *Trends in Cognitive Sciences*, 21(3):216 – 228, 2017.
- [26] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S Chen, and Thomas S Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and image understanding*, 91(1-2):160–187, 2003.
- [27] George Caridakis, Ginevra Castellano, Loic Kessous, Amaryllis Raouzaïou, Lori Malatesta, Stelios Asteriadis, and Kostas Karpouzis. Multimodal emotion recognition from expressive faces, body gestures and speech. In Christos Boukis, Aristodemos Pnevmatikakis, and Lazaros Polymenakos, editors, *Artificial Intelligence and Innovations 2007: from Theory to Applications*, pages 375–388, Boston, MA, 2007. Springer US.
- [28] Chuang Yu and Adriana Tapus. Interactive robot learning for multimodal emotion recognition. In Miguel A. Salichs, Shuzhi Sam Ge, Emilia Ivanova Barakova, John-John Cabibihan, Alan R. Wagner, Álvaro Castro-González, and Hongsheng He, editors, *Social Robotics*, pages 633–642, Cham, 2019. Springer International Publishing.