A Mixed Reality Shopping System Providing Hand Scale Preview and Gesture Interaction

MA RUICHEN

44181601-1

Master of Engineering

Supervisor: Prof. Jiro Tanaka

Graduate School of Information, Production and Systems

Waseda University

September 2020

Abstract

Online shopping is becoming an indispensable part of people's life because of its effectiveness and convenience. These years, the development of mixed reality techniques provides us more possibilities for designing mixed reality shopping systems. In mixed reality environment, users are allowed to use their hands and voice to interact with virtual contents while seeing the real environment, which provide users with an immersive experience. One of the most important features of mixed reality shopping systems is the product preview, which helps users see the product's real effect before the purchase. However, current mixed reality shopping systems are still limited on shopping and preview experience, that lack of natural interactions and flexibility on preview interface.

In this paper, we proposed a mixed reality shopping system with a hand scale preview interface and a new hand gesture set for interactions. We intended to design a more intuitive preview interface to display the virtual product directly over user's palm in a suitable size, while the user can use a series of related hand gestures to manipulate the virtual object. This system also allow users to switch two kinds of preview modes: hand scale preview and real size preview, to observe the virtual products freely in their rooms. Besides, we introduced instinctive gestures which include multiple hand shapes in one gesture so that users would feel easier to interact with virtual components without remember too many specific gestures. We connect a mixed reality head-mounted display to a hand tracking sensor to enable our system display and real time hand detection.

In the last stage, we have carried out an user study to let participants use our system and give some feedback. The preliminary evaluation result shows that our system improved user's online shopping experience.

Keywords: mixed reality, gesture interaction, immersive shopping

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Prof. Jiro Tanaka, who provided me much guidance and valuable advice on my research. It's a great honor to have him led me into the field of HCI. It is his patience and kindness that helped me keep making progress on my study. Without him, my master research would not have gone that smoothly during these days. His challenging academic attitude infected me to become more creative and persist in exploring.

Secondly, I would like to thank all IPLAB members. I am glad to have meet you and spent a precious time together during these 2 years. You have provided me so much help not only on my study, but also on my life away from home country. Last but not least, I want to thank my parents. Thank you for giving me financial support and encouragements throughout my school life. I am appreciated to have your warm accompany all the time.

Contents

List of figures vi

List of tables viii

1 Introduction 1

1.1 Introduction 1

1.2 Organization of the thesis 2

2 Background 3

2.1 Mixed Reality 3

2.2 Mixed Reality Shopping System 5

2.3 Hand Gesture Interaction 5

3 Related Work 7

3.1 Related work of Mixed Reality Shopping System 7

3.2 Related work of Hand Gesture Interaction 9

4 Research Goal and Approach 10

4.1 Research Goal 10

4.2 Research Approach 10

4.2.1 Approach 10

4.2.2 Use Case 12

4.2.3 Novelty 13

5 System Design 15

5.1 Interface Design 16

5.2 Hand Gesture Set 20

5.2.1 Gesture Set Design 20

5.2.2 Gesture Classification 21

5.2.3 Summary of Gesture Set Design . . . 24
5.3 Interaction Method . . . 25
5.3.1 Interactions for Interface . . . 26
5.3.2 Interactions for Hand Scale Preview . . . 28
5.3.3 Interactions for Real Size Preview . . . 29

6 System Implementation 30
6.1 System Hardware . . . 30
6.2 Software Environment . . . 31
6.3 UDP Connection . . . 32
6.4 Hand Gesture Setup . . . 34
6.4.1 Hand Gesture Recognition . . . 34
6.4.2 Hand Gesture Interaction . . . 38
6.5 Voice Search . . . 41
6.6 Interface Configuration . . . 42
6.7 Preliminary Evaluation . . . 44
6.7.1 Participants . . . 44
6.7.2 Method . . . 45
6.7.3 Result & Discussion . . . 47

7 Conclusion and Future Work 49
7.1 Conclusion . . . 49
7.2 Future Work . . . 50

References 51

List of figures

2.1 Representation of the “Virtuality Continuum” 4
2.2 Microsoft HoloLens 4

3.1 AR shopping systems 8
3.2 Lu’s shopping systems 9

4.1 Searching Interface 12
4.2 Previewing view 13

5.1 System Design Overview 15
5.2 Interface design concept 16
5.3 Start searching panel 17
5.4 Result list panel 18
5.5 Product detail information panel 19
5.6 Preview Interface 19
5.7 Level 1 Gestures: Pointing, Palm up 22
5.8 Level 2 Gesture: Scroll 23
5.9 Level 3 Gestures: Press, Grab 23
5.10 Interaction flow of start searching panel 27
5.11 Hand manipulation in hand scale preview 28
5.12 Hand manipulation in real size preview 29

6.1 HoloLens and Leap Motion Setup 30
6.2 Hand data collection process 32
6.3 Hand JSON data collected by Leap Motion 33
6.4 Hand joint tracked by the depth sensor for gesture recognition 35
6.5 Hand gesture interaction operation flow 38
6.6 The concept of three kinds of pointers 39
6.7 Pointer event action flow 41

6.8 Voice search process 41
6.9 Structure of result list 43
6.10 Scaling algorithm for hand scale preview models 44
6.11 Questionnaire 46
6.12 Average Score of Question 1 to 8 47

List of tables

5.1 The hand gesture set for MR shopping system 21
5.2 Hand gesture set classification 22
5.3 Gesture set comparison of our system and Lu's system 24
5.4 Comparison of similar gestures in our system and Lu's system 25

6.1 The PC setup information 31
6.2 MRTK v2 articulated hand shapes 36
6.3 Articulated hand poses included in gestures 37
6.4 Gesture interaction happening time 40

Chapter 1

Introduction

1.1 Introduction

As augmented reality (AR) technology provides more and more possibilities to different kinds of applications, mixed reality (MR) shopping system, as a new trend of e-commerce, also becomes popular in research field. In a mixed reality scenario, the user is allowed to see both virtual objects and the real world, which enables an immersive experience for interacting with virtual interfaces based on the user's surroundings.

Previewing experience is one of the most important part for developing e-commerce applications. 2D images could not satisfy most users because of the limited information it shows. However, in an augmented environment, seeing 3D virtual products in preview can make up for this kind of insufficient to some extent. With a hand detection system, users can even use their own hands to move the virtual product to view how it looks like in different sides or place it in the room.

Currently, researches on the preview experience of MR shopping system are mainly focused on displaying the real size product model in the user's room, besides the hand interaction method on virtual objects still has less flexibility. But in the case of previewing some large items like furniture, real size model would take up lots of space, and for some very small objects, it is hard to see it clear. So it is a better way to add another preview mode by rescaling the virtual objects in different sizes to a proper size.

In this research, we designed a hand scale preview interface for HMD (head-mounted display) based mixed reality shopping system. The user is able to view the rescaled product's virtual model over the hand palm while using the other hand to manipulate it. We implemented a new set of hand gesture set based on hand tracking techniques for interacting with the virtual interface and virtual product model. We used instinctive hand gestures that users do not need to remember specific gestures but interacting with virtual objects freely like as in the real life. This system provides real size preview as well, so that users can switch from two preview modes during purchase and move or rotate the virtual products to have a better shopping experience at home.

1.2 Organization of the thesis

The rest of the thesis is organized as follows: In chapter 2, we will introduce the background of the thesis. In chapter 3, we will introduce some related works. In chapter 4, we will describe the goal and approach of our research and show the use case of designing our system as well. In chapter 5, we will introduce the concept of the system design. In chapter 6, we will show the implementation details for the system development and a preliminary evaluation process. In chapter 7, we will make a conclusion and talk about future possibilities.

Chapter 2

Background

2.1 Mixed Reality

Mixed reality (MR) is the result of merging physical world with the digital (virtual) world to produce interactive new environments in which physical and virtual objects co-exist and interact in real time. Mixed reality systems are becoming more prevalent because of the intuitive interface. Unlike in virtual reality, MR interface allow users to see the real world and surrounding objects by wearing a hand-held or head-mounted display (HMD). Many MR systems also focused on providing similar interactions to manipulating real objects, like providing haptic feedbacks and physical constraints to virtual contents.

The term of mixed reality was firstly introduced by Paul Milgram and Fumio Kishino[1] in 1994. Their paper introduced a description of virtuality continuum (Fig. 2.1), which explained the definition of mixed reality environment. They also introduced a definition that MR is “a superset in terms of a ‘mix of real and virtual objects within a single display.’”

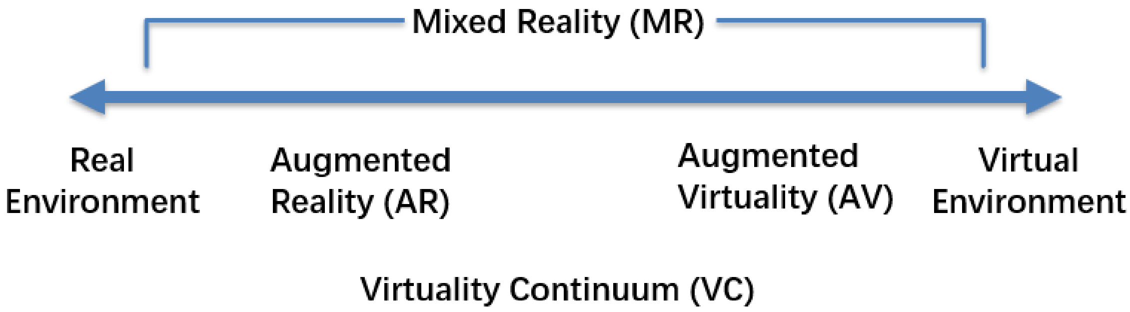


Fig. 2.1 Representation of the "Virtuality Continuum"

Microsoft are pushing MR as a new technology as they launched a MR HMD HoloLens[2] (Fig. 2.2), and then expand it to a range of MR devices based on Windows platform, along with the Mixed Reality Toolkit for developers to build new MR applications. They deem mixed reality experiences as a combination of computer processing, human input and environmental input[3], which highlighted a direction of the development of mixed reality interactions.

Fig. 2.2 Microsoft HoloLens

2.2 Mixed Reality Shopping System

Current online shopping systems are mostly limited to showing 2D images and text information of products to customers, usually users cannot see a digital preview of items in physical world. Recently, with the development of computer vision, natural language understanding and other mixed reality related techniques getting evolved, E-commerce sites on the Internet are starting to seek for approaches through AR methods to make their shopping systems become more convenient and enhance the shopping experience to customers. Both IKEA[4] and Amazon[5] have launched AR applications on smartphone that allow users to place their product at home before purchasing. These systems not only made it easier for users to better observe the product before buying, but also increased their tendency of buying the product.

Along with the applications developed for E-commerce sites, using mixed reality techniques to enhance retail shopping experience is another orientation. Some people still prefer shopping in physical stores rather than shopping online[6]. Dou and Tanaka[7] introduced an MR shopping system to decorate the user's room into a real shop which shows different styles based on the brand of the shop. While some researchers focused on providing a support for "disadvantaged shoppers" to help them experience the real shop view and service. Ohta, et al.[8, 9] proposed a mixed reality shopping system with panoramic view inside the physical store which is displayed on an HMD. It allows users to tap on a smartwatch to select products during shopping.

As we can see, mixed reality indeed provides more possibilities to online shopping experience. In our study, we designed a preview experience for MR shopping system, which could help users acquire a more comprehensive understanding before their purchase.

2.3 Hand Gesture Interaction

Hand gesture has been considered as a powerful and promising interaction method between human and computers. It overcomes the limitations of touch input interfaces in AR and VR scenarios while offering an efficient interaction and richer experience. The applications of hand gesture are diverse, ranging from medical surger[10], exhibition[11], sign language[12] and so on.

Gesture data collection is an important procedure in hand gesture recognition. According to the works have been done in this field, the methods can be mainly divided into two categories: glove-based and vision-based. The gloves consisted of multiple sensors can measure flexion of hand joints thus collect hand gesture data. Using data gloves makes it possible to provide various tactile feedbacks[13], while vision-based recognition aimed at providing visual feedbacks[14]. Unlike data glove device, vision-based gesture recognition usually uses cameras or depth sensors to collect real time hand data image sequence. In this case, users are not required to wear any other hardware device apart from the display device, so that they are allowed to use free hands to realize control. Popular sensors and technology such as Kinect, Depth-Sense and Leap Motion has been widely used in related researches.

Gestures can be static and dynamic. As to vison-based dynamic gesture recognition, approaches can be classified into two classes, feature-based and model-based approach[15]. The main idea of the former approach is to find a correspondence between 2D image features and 3D world coordinate system. Boulahia et al.[16] proposed to extract the higher level HIF3D features on the hands trajectories to model 3D gestures. Model-based approaches use a model of features of tracked hands for predicting gestures. Lee et al[17], introduced a method to reconstruct a hand pose model through a real-time algorithm that recognizes fingertips with six-degree-of-freedom information.

Recently, hand skeleton data are used in many studies for dynamic gesture recognition[18, 19], the articulated hand model can also be displayed in AR scenario for teaching user gestures or indicate hand position and poses. In our work, we also applied this method for our hand gesture recognition to enable more detail dynamic gesture interaction and giving visual and audio feedbacks in real time thus provide users with a flexible and effective shopping experience.

Chapter 3

Related Work

3.1 Related work of Mixed Reality Shopping System

With the advent of the latest hardware display devices and corresponding toolkits with different interaction methods, many researchers and companies started to use augmented reality (AR) techniques to develop various applications. Online shopping, which becomes a hot area because of its convenience, attracts a huge number of customers so that E-commerce companies find their chances on applying AR technology in their applications to enhance online shopping experience by providing 3D product preview and multiple interactions. "IKEA place" (Fig.3.1 (a))is an AR application on smartphone[4], which allows users to place the furniture they like in their room before buying it and adjust it by rotation. This helps reduced their customer's dissatisfaction of buying an unsuitable furniture. However, smartphone based AR applications are still limited on small screen display and touch interactions.

(a) IKEA Place

(b) “ShowSomething” shopping system

Fig. 3.1 AR shopping systems

Wearable mixed reality HMDs, like Microsoft HoloLens[3], gives possibilities to perform spatial interaction with the controller or hand gestures[20]. Gaze, voice and hand gesture inputs are considered as some intuitive ways to interact with AR systems, many works did efforts to use these interactions to improve AR shopping experience. Mirae Kim et al[21], proposed a gaze-based interaction shopping assistance system, which would provide users with selected product information, other buyers’ comments and product related animation. Dou and Li’s system[22], introduced a MR shopping system named “ShowSomething”(Fig. 3.1(b)), in which they applied speech recognition techniques to extract user’s voice command and keywords, thus provide real-time response for searching and filtering products. Their system also provides users with an immersive preview by displaying life-size three-dimensional virtual products in users’ room while using spatial understanding techniques to realize automatic placement. The user can use two hand gestural manipulation to operate virtual products. This work introduced some use cases and solutions for MR shopping system, but their interface is still limited to 2D panels that cannot merge well with the physical surroundings. Besides, the life-size preview is also not suitable for previewing very large or small items. We think that hand gestures can be more flexible and multiple for users to interact with 3D UIs and operate virtual products in the preview interface.

Fig. 3.2 Lu's shopping systems

3.2 Related work of Hand Gesture Interaction

To figure out what kind of hand gestures are needed in a mixed reality shopping system, we referred to Lu's work[23]. This work introduces a gesture set for a virtual reality shopping system, as well as a gesture classification method to make the hand gesture set more structural. As we are aiming at designing a gesture set for mixed reality system, we did some adjustment in our gesture set to suit the mixed reality shopping operation. Furthermore, the gestures used in this system are fixed to specific hand postures, too many gestures could be difficult for users to remember, so that it would be hard for them to remind one of the gestures when they want to manipulate a virtual object. In our system, we used instinctive gestures to overcome this problem and make it easier for users to interact with virtual contents naturally.

Chapter 4

Research Goal and Approach

4.1 Research Goal

This research is aimed at designing a hand scale preview interface and a set of new hand gesture interactions for MR shopping system.

According to the previous researches on MR shopping systems, both the user's preview experience and the virtual contents interaction methods are still limited on their flexibility. To provide a better preview experience for users shopping by a mixed reality system, we plan to design our system with following features.

We intend to rescale different size virtual products to a suitable size for closely preview, in order to make the online shopping preview more intuitive. Besides, we intend to implement a set of new hand gestures and related interactions for users to interact with virtual contents directly and instinctively.

4.2 Research Approach

4.2.1 Approach

Previewing the product at home before purchasing is a big requirement for customers. As the development of AR and VR techniques, it becomes possible for people to view 3D virtual objects in an immersive world by wearing an HMD. As for online shopping systems, users want to buy some real items instead of virtual things, so it is better for us to display the products' preview in the real world. Thus, in our research, we built the shopping system in mixed reality environment to provide users with multiple interactions and preview experience.

People are used to picking things up with their hands when they want to observe them closely, so we considered that scaling virtual objects to a size which is comfortable for users to manipulate by hands would be more intuitive for preview. In our system, we rescale the product virtual model into a hand scale size (about 10cm in the bottom, maintaining the original proportion), and display it over a detected hand palm. We connected a hand tracking sensor to an AR HMD to ensure precise hand gesture input in real-time. We also implemented a real size preview by displaying the original size of the item, in this mode, user can place the virtual product in the room to see its real effect. We configured depth rendering to make the virtual item looks in different size when they are at different distance from the user.

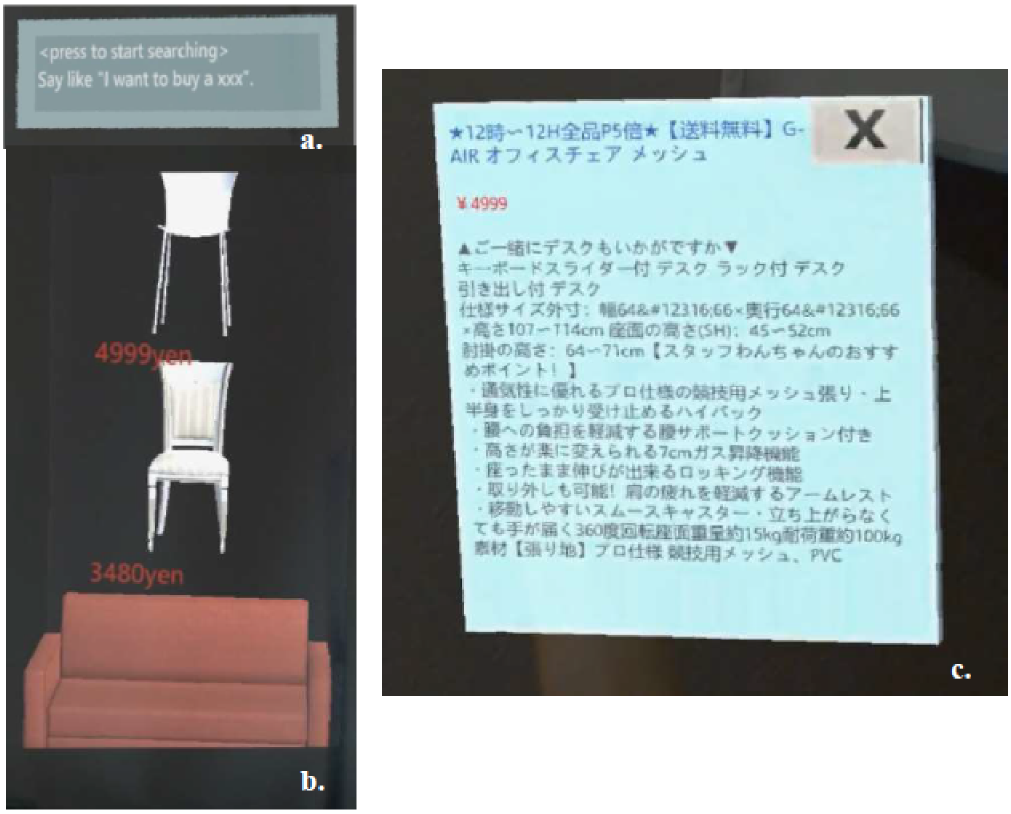
To help users use their free hands to interact with the interface and virtual product models, we designed a new hand gesture set for the MR shopping system. We used some instinctive gestures that users do not need to remember the specific hand shape but interact with the virtual object directly like in real world. All the gesture interactions need real time hand detection through the depth sensor, we defined them based on Mixed Reality Toolkit (MRTK) hand poses. The hand tracking sensor captures the hand joint data including 21 hand joints three-dimensional coordinates and rotation information, then transfers them to the HMD worn by user for recognizing the gesture and triggering the related output event. The system would generate an articulated hand model at the same time to simulate the physical hand touch or show a cursor looks like a ray comes from the index fingertip.

To make sure the system provides an immersive shopping experience, we designed an 3D result list for showing the searched items' model and applied some 3D UIs in the interface. As for the searching method, since virtual keyboard takes more time to use and needs more space for display comparing with speech input, we implemented voice search that extract keywords from user's speech and then search the target items from the shopping website.

4.2.2 Use Case

A user wants to buy a chair for his new house on the internet, he needs to check whether the size and the style of the chair fits his room or not, so he starts a mixed reality shopping system at home with his AR headset.

He presses the guiding button to start searching by saying “I want to buy a chair”. The words he said shows in the panel and then a result list containing searched products’ models with their prices appears in front of him. He uses a scroll gesture to look through the chairs and find out one he interests. He pressed the product, another panel showing its detail information displays in the left. He checks it but he is still not sure how it looks like in real world.



(a) Start searching panel (b)Result list (c)Detail information panel

Fig. 4.1 Searching Interface

So he puts up his left palm to start preview, a hand scale model of the chair shows over his hand, just like he is holding it. He uses his right hand to grab the chair and rotate it to check it on different sides. It looks good, he wants to put it into his room to see if the size is suitable. He presses on the virtual chair and it transforms to its real size. He grabs the chair to place it on the floor, and then he grabs it with his hands to adjust it by move and rotate it.

He thinks the chair looks good, but he would like to view if other products are better. Therefore, he selects another chair in the list and places it in the room just like the steps he did before. He found the second one is better, so he presses on the first chair to remove it.

(a) Hand scale preview

(b) Real size preview

Fig. 4.2 Previewing view

4.2.3 Novelty

The novelty of this research mainly reflects in these aspects:

1. We introduced a framework for HMD based mixed reality shopping system, which supports two kinds of preview modes and intuitive hand gesture interactions.

2. We designed a hand scale preview interface for mixed reality shopping system by rescaling the virtual product model then display it over the detected hand palm, while allowing users to use the other hand to manipulate the virtual object;

3. We introduced a new hand gesture set containing instinctive gestures with a series of interaction methods for interacting with the virtual interface and 3D product models in our system by connecting a hand-tracking sensor with a see-through type of HMD;

Chapter 5

System Design

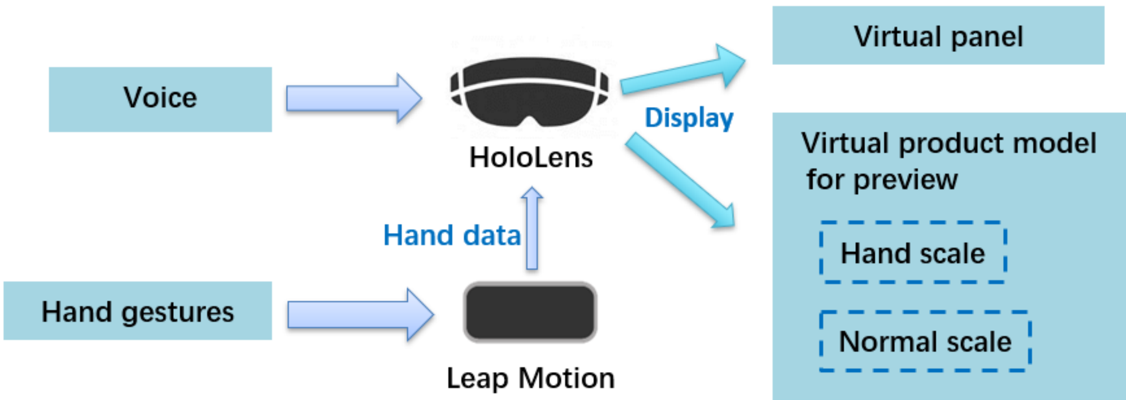


Fig. 5.1 System Design Overview

In this chapter, we will introduce the system design of our mixed reality shopping system from three aspects: interface design, hand gesture set, and the interaction method. Figure 5.1 shows the overall structure of our system.

There are two kinds of inputs in our system, voice and hand gestures. We apply voice input for searching products, and hand gestures for interacting with interface and manipulating virtual items. This system is displayed on a mixed reality HMD, HoloLens, in a mixed reality world, the user is able to see his hands and his surroundings. We use a hand tracking sensor to collect hand tracking data and send it to the HMD in real time. The HMD would collect voice input and analyze the hand data as well. The searching voice dictation and searching results would be shown on virtual panel, while the product preview would be displayed as product's 3D virtual model apart from the panel in its hand scale or normal scale.

5.1 Interface Design

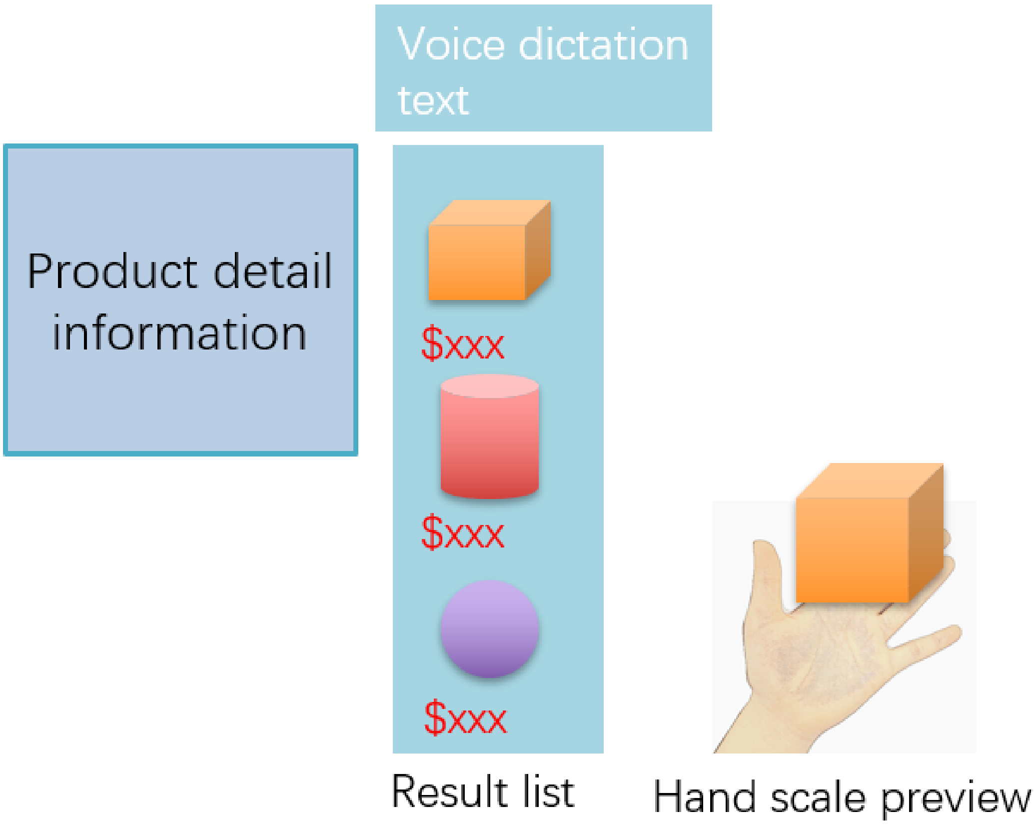


Fig. 5.2 Interface design concept

Our system interface mainly contains two parts, searching panels and preview virtual models. The interface design concept is shown in Fig. 5.2, which shows the layout of the main interface of this shopping system.

In our system, we assume that e-commerce sellers provide 3D model information for products. Even though current online shopping systems does not provide much product models in their database, as we mentioned in chapter 2, we can see a trend that many e-commerce companies are working on that. Besides, in order to better integrate the panel with the real scenery and provide an immersive experience, we applied 3D UIs in our system and added lots of visual and audio feedbacks to them, so that users can better recognize their actions while using this system.

Searching panels includes a start searching panel on the top, a result list showing searched items in mini-sized 3D models, and a product detail information panel.

Start Searching Panel

(a) Start searching guide (b) Real size preview (c) Real size preview

Fig. 5.3 Start searching panel

Fig. 5.3 shows the start searching panel. When the user firstly starts this system, it would simply show the start searching panel which has some guiding texts to help the user get start with the voice searching process. We design it to be a button so that user can press on this panel to start speaking. In this panel, it would show text feedbacks during the system is listening to the user's speech and the dictation result when it detects that the user finished speaking.

Result List

After the user completing the first step, the system would extract the keywords from the detected text from speech, and search it in the shopping website database. Then it would generate a result list panel for showing the searched items, like it shows in Fig. 5.4 In this panel, the items are shown with their 3D models and their prices. We design this list to turn pages by being scrolled instead of using buttons, so that users can operate faster when browsing without precisely touches. The user can also press on each product to see its detail information and preview.

Fig. 5.4 Result list panel

Detail Information Panel

After the user selected a product, the detail information panel would be displayed on the left of the list, as shown in Fig. 5.4. This panel contains the title, price and text description for the product. It is available for users to move it or close it by hands.

All these panels are designed to face to the user in a radical field and follow user's head, so that they can be easily found around the user and would not be too much disturbing during the preview.

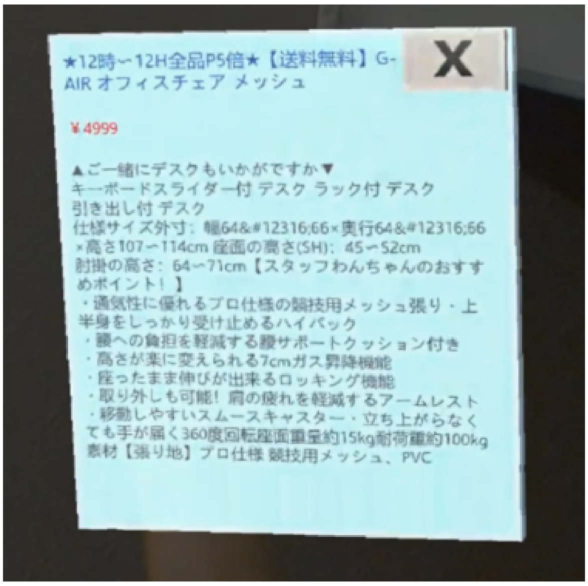
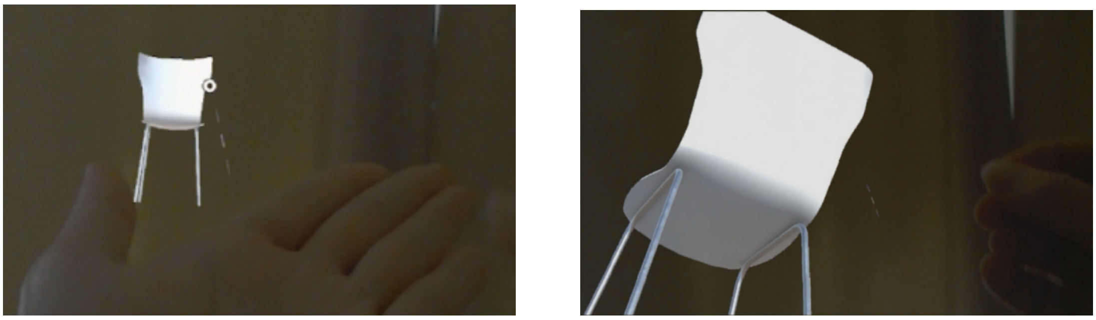


Fig. 5.5 Product detail information panel

Preview Interface

The other part of the interface is for preview. We designed two preview modes, hand scale preview and real size preview. In preview interface, the system would display the 3D model of the selected product in the real world.



(a) Hand scale preview

(b) Real size preview

Fig. 5.6 Preview Interface

Hand Scale Preview

The hand scale preview interface, as shown in Fig. 5.6 (a), starts automatically after the user selecting an item. The user can put a palm up to see the virtual product over their hands, and put it down to suspend.

To make the virtual products in different size looks suitable on hands as a primary stage preview, we rescale all the product models into a size of 10 cm2 in the bottom, keeping the initial proportion. The virtual product model would follow the user's hand based on the hand tracking sensor and a solver algorithm in MRTK. Besides, the user can use the other hand to move or rotate the hand scale model to have a better preview in different sides.

Real Size Preview

Although we designed a hand scale preview interface in this system, we noticed that it is still needed to provide real size preview to let user place the product in their room to see how it looks like in real world. So we also add this to our system and implement it with multiple interactions for manipulation.

Fig. 5.6 (b) shows the real size preview interface, the user can switch to this mode by pressing on the hand scale model and remove it by tap on it again. The real size model would be rendered with depth effect that looks smaller at a farther distance.

5.2 Hand Gesture Set

5.2.1 Gesture Set Design

Many kinds of hand gestures have been designed to be used for interacting with AR and VR systems, as well as digital devices like smartphone, PC and HMD. In an augmented environment, use free hand gestures instead of controllers to interact with virtual contents would be a better choice for users. Basically, Microsoft HoloLens (1st gen)[3] can only support two kinds of gestures: air-tap and bloom. While in our system, we intend to provide multiple interactions based on different hand gestures, so we used a hand tracking sensor to detect hand joint data in real time to support a new gesture set that has more flexibility.

According to the use case for using a mixed reality shopping system we imagined in chapter 4, we designed 5 gestures in our mixed reality shopping system: (1) pointing, (2) press, (3) scroll, (4) palm up, (5) grab. The functions we designed for these gestures are shown in Table 5.1.

Number	Gesture	Function
1	Pointing	Enable a cursor for indicating the pointing point of index finger.
2	Press	Press at a button; Switch to the real size preview; Remove the preview model.
3	Scroll	Scroll the result list.
4	Palm up	Show the selected virtual model over the hand palm.
5	Grab	Move the panel; Manipulate 3D virtual models in preview.

Table 5.1 The hand gesture set for MR shopping system

In addition, we applied some instinctive gestures in this hand gestures set, like scroll and grab. Instinctive gesture means that users can interact with the virtual object instinctively like in real life. In this case, users do not need to remember the specific gestures when using the system. These gestures are realized by displaying hand feedback and define several different hand poses that people are tend to use for achieving the same function in one gesture.

5.2.2 Gesture Classification

We used the gesture classification method in Lu's work [24] to classify hand gestures in our system to make this gesture set has a clearer structure. According to Lu's work, the gestures can be classified in three levels based on their characteristics:

- Level 1: Core static hand postures. In level 1, gestures are just hand shape without hand motion. The classic example is pointing gesture.
- Level 2: Dynamic palm motions are divided into the level 2. In level 2, we only care about the palm movement, while not care about the shapes of fingers.

- Level 3: Combination hand gestures are divided into the level 3. Combination hand gestures combine the features of level 1 and level 2 gestures. In level 3, we care about the motion and shapes of fingers and the motion of palm.

Based on these levels, we classified the gestures in our system in Table 5.2 In Fig. 5.7, Fig. 5.8, Fig. 5.9, we show the hand shape and motion for each gesture.

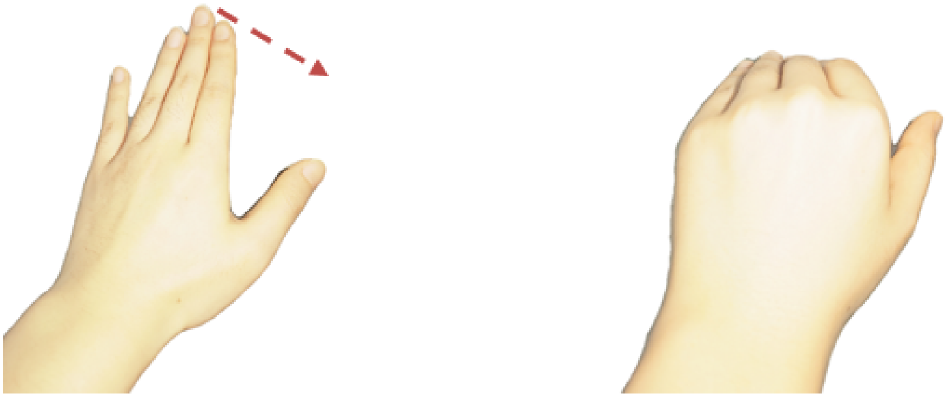
Level	Gesture	Description
1	Pointing	Stretch out index finger from a fist pose.
1	Palm up	Palm side face to the top, all fingers are extended.
2	Scroll	Fingers stretched out, palm face to the front, then fingers bend to the palm.
3	Press	Start from a pointing pose, then index finger presses down.
3	Grab	Start from a stretched hand, grip by thumb and index finger, or use three or five fingers to grip.

Table 5.2 Hand gesture set classification

(a) Pointing

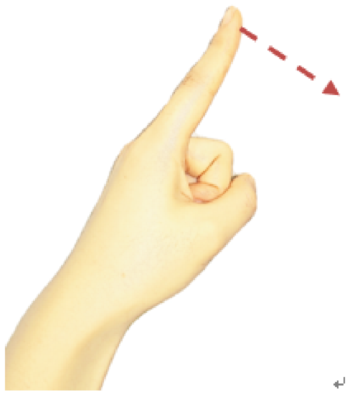
(b) Palm up

Fig. 5.7 Level 1 Gestures: Pointing, Palm up

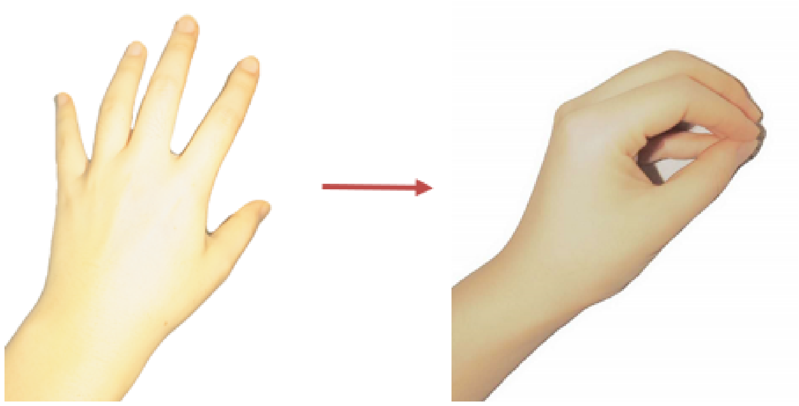


(a) Start scroll pose (b) End scroll pose

Fig. 5.8 Level 2 Gesture: Scroll



(a) Press



(b) Grab

Fig. 5.9 Level 3 Gestures: Press, Grab

5.2.3 Summary of Gesture Set Design

We tried to design a hand gesture set for the users to interact with the mixed reality shopping system by using free hands. In Table. 5.3, we compared our gesture set with Lu's gesture set, as we both designed hand gesture set for HMD based shopping systems.

	Our System	Lu's System
System Environment	AR	VR
Number of gestures	5	14
Core gesture	Pointing	
Gestures for interacting with interface	Press Scroll Grab	Click Scroll/Swipe Opening/Closing
Gestures for interacting with virtual object	Palm up Press Grab	Grab Holding Drag Rotation Zoom in/out
Gestures for interacting with space	/	Changing area

Table 5.3 Gesture set comparison of our system and Lu's system

Since the two systems are developed in different environments (Lu's system is in VR, our system is in MR), our gesture set does not include any gestures for manipulating the virtual space. We did not implement zooming gestures as well because we have designed two kinds of preview interface in our system that can scale the virtual object to a suitable size automatically. Besides, our system focuses on interacting with 3D interface and virtual product preview models, we developed more flexible scroll and grab gestures for providing users with a natural experience. It can be noticed that we all implement scroll and grab gestures in our gesture set, but we defined different gestures to achieve that in our systems. In Table 5.4, we analyzed the difference of these two gestures in our system and Lu's system based on their description, function and gesture classification level.

Gesture		Our system	Lu's system
Scroll	Description	Bend fingers to the palm.	Stretch out index finger and middle finger to swipe.
	Function	Scroll the whole list.	Move the menu slider to scroll.
	Level	2	3
Grab	Description	Grip by thumb and index finger; use three or five fingers to grab.	Bend other four fingers close to the thumb.
	Function	Hold/move/rotate the virtual object	Make object move close to hand.
	Level	3	3

Table 5.4 Comparison of similar gestures in our system and Lu's system

5.3 Interaction Method

We have introduced our system interface and hand gesture set in previous sections. Based on these components, we designed interaction methods for different parts of the virtual contents to make our system intuitive and interactive. Our system interaction methods are designed for three parts: interface, hand scale preview, and real size preview, we will introduce each part in the following sections.

5.3.1 Interactions for Interface

The main interface of our system is three virtual panels, start searching panel, result list, and product detail information panel. To interact with these virtual panels, users can use both voice and hand gestures. As shown in Table. 5.4, the gestures include pointing, press, scroll and grab. In the following paragraphs, we will introduce the interaction method of the three panels.

Start searching panel: To start searching some items, the user needs to firstly put up his hands to pointing at the panel, when the cursor is on this button, the user can press on the panel to start the searching process. When the system shows the "Listening" text on this panel, it means that the user can say something he want to purchase. The system would generate the dictation text and display it in the panel. Then the system would extract the keywords from the dictation text through a text analysis API and search them in the shopping website. If the system cannot detect the keyword or cannot find any result related to the keyword, it would show a warning in the panel. The interaction process of this panel is shown in Fig. 5.10.

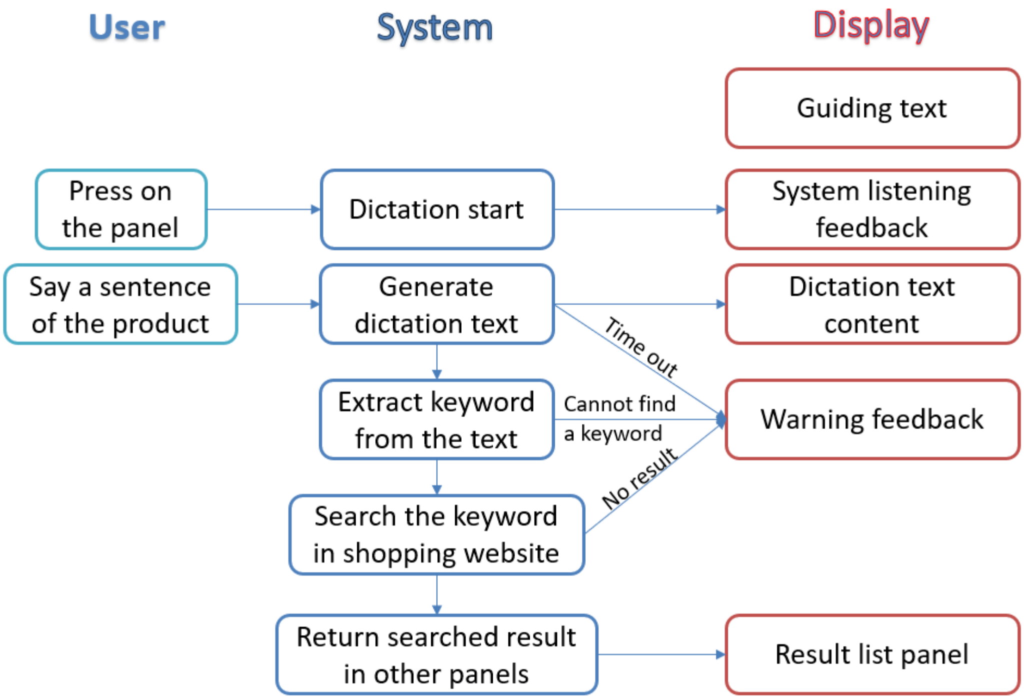


Fig. 5.10 Interaction flow of start searching panel

Result list: If the system received some searching results from the former process, it would start to display this panel. The system would download the product information one by one from the website and show the product model and price in each button in the list. The user can use scroll gesture to make the list sliding with a fall off velocity, so that he can view other items. When the user's hand is pointing at one of the buttons, a guiding text would show on the right of the button to help the user understand how to see other information of this product in the next step. The user can press on a button to display the detail information panel and preview it. In addition, the user would get sound feedback when he scrolls the list or press on buttons.

Detail information panel: If the user pressed on a product button, this panel would display. When the user's hand is pointing on this pad, it would render the fingertip point with a proximity light as a visual feedback. This pad can be grabbed to drag it to other places. The user can also press on the button with closing icon to turn off this panel.

5.3.2 Interactions for Hand Scale Preview

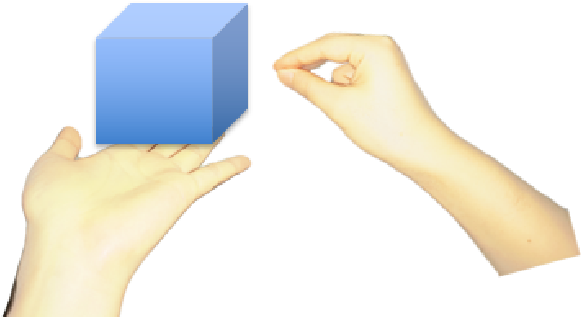
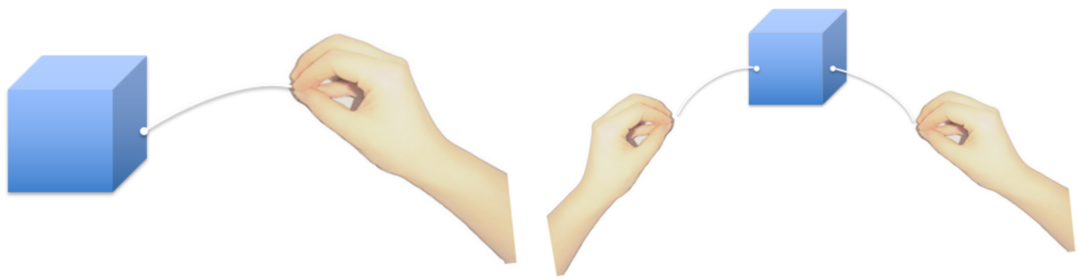


Fig. 5.11 Hand manipulation in hand scale preview

If the user has pressed the button of an item, he can use palm up gesture to start hand scale preview. As shown in Fig. 5.11, in hand scale preview, the system would display the virtual model over the user's palm based on the detected hand palm position, and make it follow this hand. The user needs to keep this gesture like holding the object. If the system cannot detect the palm up gesture anymore, it would automatically stop displaying the hand scale preview model. During the hand scale preview, the user can put up the other hand with a grab gesture to rotate or move the virtual model at a near distance. In this mode, the model would be rotated about its bottom center while keep standing up. It would be switched to a real size preview when the system detects a press on the hand scale model. The user can also press on other objects in the list to change the current previewing product.

5.3.3 Interactions for Real Size Preview



(a) One hand manipulation (b) Two hands manipulation

Fig. 5.12 Hand manipulation in real size preview

Press on the product's hand scale model can start the real size preview interface. In this mode, the user can use one hand or two hands manipulation to interact with the product virtual model at a distance, which is shown in Fig. 5.12. There would be a hand ray cursor comes from the user's hand to indicate the grab point when the user is trying to grab the model. As for one hand manipulation, the user is able to move the model or rotate it about the grab point by holding the grab gesture and move the hand. In the case of using two hands to manipulate the object, the user can hold the grab with to hands while the cursor is on the object to move and rotate it about the object center. The user could press on the object to remove it or keep it to start to preview other items and compare them together in the room.

Chapter 6

System Implementation

6.1 System Hardware

In our mixed reality shopping system, we use Microsoft HoloLens (1st gen) as the mixed reality HMD device. HoloLens is equipped with a microphone and an immersive earphone to enable the voice input and audio feedback. The depth camera on it can detect user's hand and gaze point position, to collect data for generating the system interface.

Fig. 6.1 HoloLens and Leap Motion Setup

We use Leap Motion controller (LMC) as the hand-tracking sensor to detect real time hand data for generating articulated hand models and hand gesture recognition. Leap Motion can observe a hemispherical area within 1 meter distance, so that users can have a wide space for hand interactions. In order to maintain the relative position of LMC and HoloLens and ensure our system to be hand free, we implemented a 3D print head mounted frame to mount the LMC on the top of the HoloLens, as shown in Fig. 6.1.

As the Leap Motion itself cannot provide electricity and it needs a USB cable for data transfer, while the HoloLens' USB-port does not support OTG protocol. We connect the Leap Motion to a laptop PC as a server in order to send data to HoloLens wirelessly (UDP protocol)[23]. Table 6.1 shows the information of the PC we use for building this system. This laptop and HMD Unit are in the same network environment to ensure a good data transfer speed.

Operation System	Windows10.0.18362
CPU	Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz
Graphics Card	Intel(R) HD Graphics 520
RAM	8.00GB

Table 6.1 The PC setup information

6.2 Software Environment

The Software Environment support is:

1. Unity 2018.4.23f1, it provides the development platform for applications based on Universal Windows Platform (UWP) which can be run on HoloLens. Unity is also implemented with an XR SDK which helps our system to handle input and output events.

2. Mixed Reality Toolkit 2.3.0 (MRTK v2), it supports basic hand poses, interface UXs and some basic configuration modules for our mixed reality system.

3. Leap Motion Orion SDK 3.2.1, it helps collect and analyze the raw data from Leap Motion sensor.

4. Microsoft Azure text analytics service. We use this API for extracting the keywords from user's speech to get the searching target.

5. Rakuten Ichiba (E-commerce website) web shopping API. We use this API to get product information for displaying in our system searching interface.

Besides, we used Visual Studio 2019 with C# for scripting, debugging and simulation.

6.3 UDP Connection

Since the LMC needs a physical USB cable to ensure the data transfer speed and Microsoft HoloLens Micro-USB port cannot be used with peripherals. We set up a UDP connection between these two devices to stream the sensor data from LMC to the HoloLens.

The Leap Motion SDK installed on the computer with a Leap Motion controller provides raw tracking data through WebSocket. We made a WebSocket Server to broadcast JSON data of the Leap Motion, and we also developed a client on UWP platform for running on the HoloLens to get receive these data. In addition, because the hand data for hand gesture recognition are handled by the MRTK v2, there is a transformation process to convert the JSON raw data into the related hand data in MRTK after the HoloLens client received the data. The whole hand data collection process is shown in Fig. 6.2.

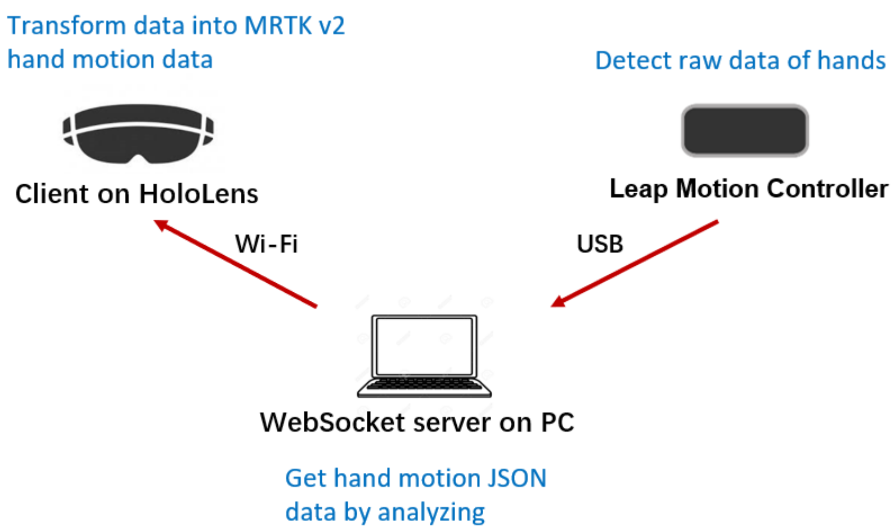


Fig. 6.2 Hand data collection process

```
public class LeapHand
{
    public int Id;
    public bool IsLeft;
    public bool IsRight { get { return !IsLeft; } }
    public float Confidence;
    public float TimeVisible;
    public float GrabStrength;
    public float GrabAngle;
    public float PinchStrength;
    public float PinchDistance;
    public float PalmWidth;
    public Vector PalmPosition;
    public Vector PalmVelocity;
    public Vector PalmNormal;
    public Vector Direction;
    public Vector WristPosition;

public class LeapBone
{
    public float Length;
    public float Width;
    public Vector Center;
    public Vector Direction;
    public LeapQuaternion Rotation;
```

```
public class LeapFinger
{
    public int Id;
    public bool IsExtended;
    public float TimeVisible;
    public float Width;
    public float Length;
    public Vector TipPosition;
    public Vector Direction;
    public Vector TipVelocity;

public class LeapArm
{
    public float Length;
    public float Width;
    public Vector Elbow;
    public Vector Wrist;
    public Vector Center;
    public Vector Direction;
    public LeapQuaternion Rotation;
```

Fig. 6.3 Hand JSON data collected by Leap Motion

The hand data collected by leap motion sensor (shown in Fig. 6.3) comes from snapshots of the motion occurring across frames, we need to use motion factors to help figure out the relative motion based on the previous frame. There are three motion factors for handling rotation, scale and translation data. We compute these factors by following expressions(r: a 3x3 rotation matrix, s: a scale factor, t: a 3-element translation vector):

$$rotation = r_{currentframe} * r^{-1}_{previousframe}$$

$$scale = e^{s_{currentframe} - s_{previousframe}}$$

$$\overrightarrow{translation} = \vec{t}_{currentframe} - \vec{t}_{previousframe}$$

To correctly generate the articulated hand model in MRTK based on the collected hand data, we also configured some properties in MRTK including the LMC orientation, enter and exit pinch distance to make our hand interactions more accurate and smooth.

6.4 Hand Gesture Setup

In most studies, dynamic hand gesture recognition methods exploit the temporal character of hand motion, by considering the gesture as a sequence of hand shape [24]. In this case, dynamic hand gesture recognition usually has these following steps: data collection, hand feature extraction, hand gesture classification and gesture prediction.

In our system, we use the Leap Motion to track and collect hand joint data and transfer these data to the HoloLens, then MRTK v2 configured in the system would recognize each gesture and generate an articulated hand skeleton model in real time to respond to user's hand interactions. We will introduce both hand gesture definition and hand interaction details in the following.

6.4.1 Hand Gesture Recognition

The hand gesture set in our system is recognized based on an articulated hand model generated by MRTK v2 with 21 hand joints overlapped our real hands, and all the hand joint position data are detected by the Leap Motion controller in real time. The hand joints tracked by the LMC for gesture recognition is shown in Fig. 6.4.

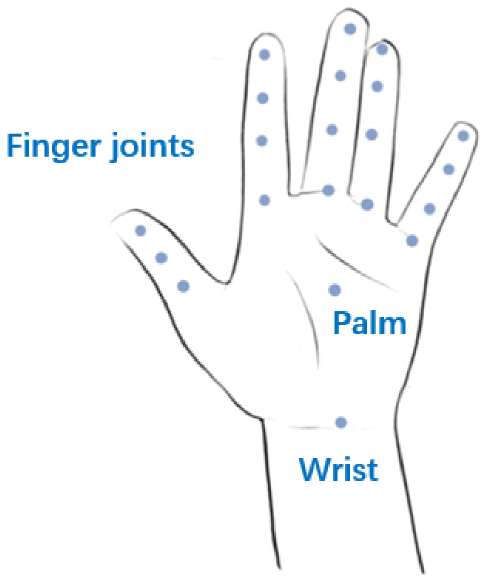


Fig. 6.4 Hand joint tracked by the depth sensor for gesture recognition

(Joints are shown as blue points on the hand.)

As we mentioned before, LMC can detect static hand poses from single frames, so the hand gestures can be defined by defining the hand poses contained in each gesture. In MRTK v2, it has already defined some basic hand shapes based on static joint poses which can be quickly recognized and used for articulated hand interaction events, which are listed in Table. 6.2[25], we used some of these hand shapes to define the hand gestures in our system.

Name	Description
Flat	Flat hand with fingers spreads out.
Grab	Grab with whole hand, fist shape.
None	Unspecified hand shape.
Open	Relaxed hand pose.
OpenSteadyGrabPoint	Relaxed hand pose, grab point does not move.
Pinch	Index finger and thumb touching, grab point does not move.
PinchSteadyWrist	Index finger and thumb touching, wrist does not move.
Poke	Index finger stretched out.
ThumbsUp	OK sign.

Table 6.2 MRTK v2 articulated hand shapes

The complexity of gestures in each level is variable, so that different descriptions are needed for defining gestures in different levels. For level 1 gesture, we only need to set one hand pose to it in the definition script. For level 2 gesture, it needs to define a start hand pose and a description of hand motion. For level 3 gesture, we should define all the included hand poses from start to end of the gesture along with the moving method. Three gestures in our gesture set are defined based on this method, we listed the included hand poses for each gesture in Table. 6.3.

Gesture Level	Gesture Name	Included hand poses
1	Pointing	Poke
2	Press	Poke
3	Grab	Grab; OpenSteadyGrabPoint; Pinch; PinchSteadyWrist

Table 6.3 Articulated hand poses included in gestures

With the predefined hand shapes and hand joint data in 3D coordinate, we can also define gestures by setting the following key points in scripts:

- Handedness (left hand/ right hand/ both hands);
- Tracked hand joints position and hand palm orientation;
- Joints' relative position;
- Threshold comparing with a defined hand pose (flat hand pose).

We use the three dimensional coordinate vectors to describe the joint position and their movements, when we calculate the relative positions, we scale the data to [-1, 1] as the normalization, thus helps eliminate noise and improve accuracy for recognition.

The scroll and palm up gestures are defined based on this method. Both of them can be operated by using any one of the hands, we set constraints and thresholds of hand joints comparing with a flat hand pose.

As for the scroll gesture, we calculate the angle of index fingertip and palm center in three dimension vector data and compare it with the flat hand angle value to determine whether the hand fingers are bending to the palm. Based on this definition, the user can trigger the scroll interaction by bending one or several fingers which provides more flexibility to this gesture.

The palm up gesture can be simply described as a flat hand pose with hand palm facing up, while we also add two constraints to it to specify it and suit for the use case. The first one is by checking if the triangle's normal formed from the palm, to index finger tip, to ring finger tip roughly matches a flat hand based on the set threshold. The other one is to set a facing threshold for the angle of gesture rotation and camera forward direction, in order to check if the gesture is facing to the user.

6.4.2 Hand Gesture Interaction

As we have defined all the needed gestures for our gesture set in this system, the next step is to let the system respond to each gesture with a specific interaction after it has been recognized by MRTK. The gesture recognizer in Unity would generates gesture events when a gesture is detected as started or finished. We use these gesture event APIs in related interaction methods and then add these components to target virtual objects to enable output events. An overview gesture interaction flow is shown in Fig. 6.5. For some dynamic gestures, the interaction output events could happen in different stage of the gesture.

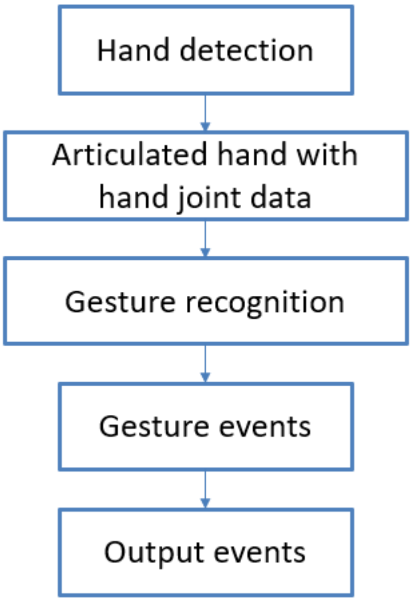
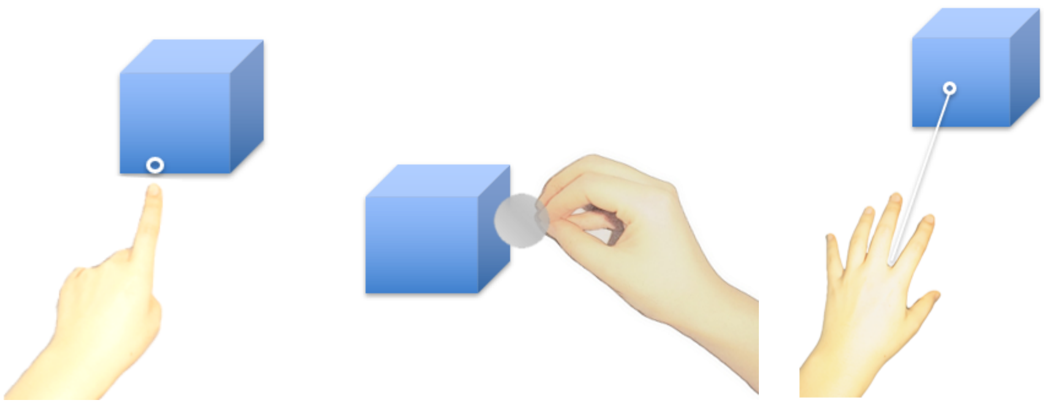


Fig. 6.5 Hand gesture interaction operation flow

The hand gestures used for manipulating virtual objects and interacting with the panels need a pointer to indicate the position that the hand is interacting with, so that we implement different kinds of pointers to the articulated hands when using different gestures. Pointers are used to generate cursor feedback and pointer events. Pointer events includes touch events and grab events, as we treat most interactions come in the form of touches and grabs.

We used three kinds of pointers in our system: poke pointer, sphere pointer and hand ray pointer (Fig. 6.6).



(a) Poke pointer (b) Sphere pointer (c) Hand ray pointer

Fig. 6.6 The concept of three kinds of pointers

Poke pointer: Poke pointer is used to enable touch interactions in near distance. It uses a small sphere cast on user's index fingertip to determine the closest touchable element. The main property is the touchable distance.

Sphere pointer: Sphere pointer is used to enable grab interactions to nearby objects. It uses a bigger sphere cast on user's grab point. The main property is the sphere cast radius.

Hand ray pointer: The hand ray pointer comes out from the center of the palm. The ray is treated as an extension of the hand. The object that the cursor lands on can then receive gestural commands from the hand, like press and grab, so that user can interact with virtual objects from a distance.

By applying different pointers, the same gesture can be used for both near and far interactions. Near interaction in our system means the interacting object is within arm's length (about 50cm), in this case, the hand ray pointers would be turned off automatically. Far interaction means the target object is farther than 50cm, the hand rays would be turned on.

In near interaction, we use the poke pointer in touch, press and interaction which triggered by pointing, press and scroll gesture. Sphere pointer is used for grab gesture interactions. In far interaction, we use hand ray pointer for all these four gesture interactions.

The system firstly generates gesture event after recognizing the gesture and then analyze the pointer events happened to respond with interaction outputs. The interaction output events can be happened in different timing during the gestures, we listed them in Table. 6.4.

Gesture name	Output event happening time
Pointing	Touch begin; Touch end;
Press	Button pressed; Button released;
Grab	Grab started; Holding after a time period; Grab ended;
Scroll	Touch begin; Scroll started; Scroll ended;
Palm up	Hand is detected; No hand is detected; First hand detected; Last hand lost

Table 6.4 Gesture interaction happening time

There are three key steps to listen for input events on a game object in Unity: (1) Register the relevant pointer in the main MRTK configuration profile; (2) Add the appropriate grab or touch script component and Unity collider to the game object; (3) Implement an input handler interface on an attached script to the related game object to listen for the interaction events. In Fig. 6.6, we show the pointer input event action flow in our system.

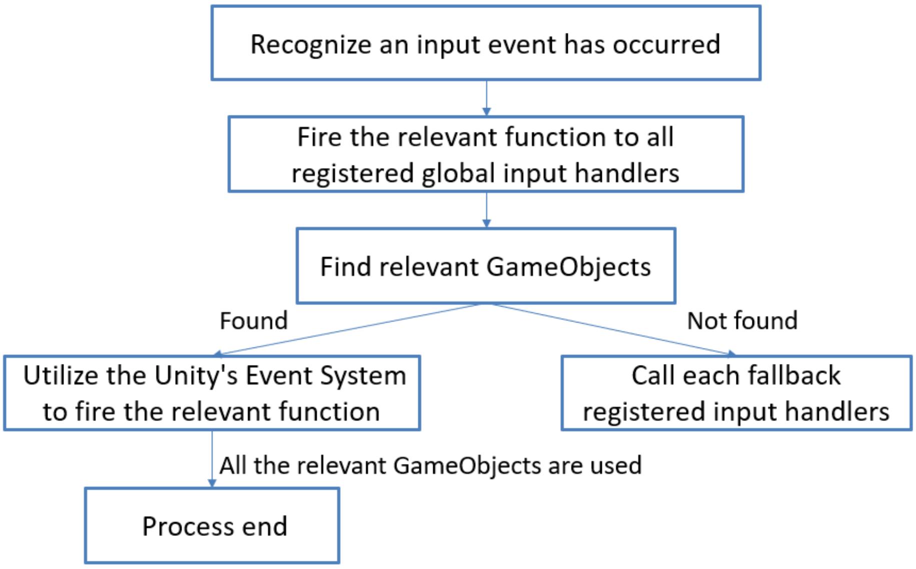


Fig. 6.7 Pointer event action flow

6.5 Voice Search

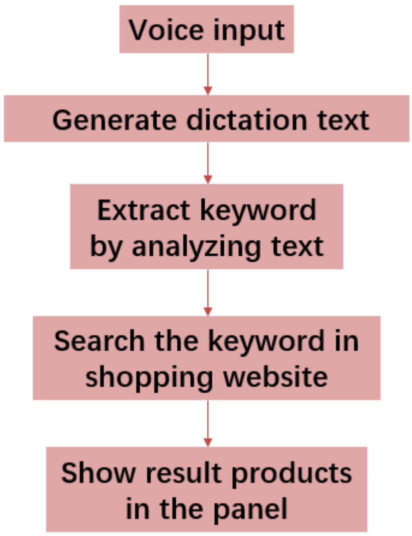


Fig. 6.8 Voice search process

The voice searching process shows in Fig. 6.8. We subscribe the Unity dictation service to transform user's input voice into a text string. Then the system would send the text to

Azure text analytic service of Microsoft through API to extract keywords from the speech text. Then the system would use the keywords of the item to search target product in Rakuten Ichiba shopping website through searching related APIs thus get products' prices and detail information. Then the result information would be displayed in the related part of our system interface panels.

6.6 Interface Configuration

To make our system interface fuse well with the real environment and present better visual effect during users' interactions, we configured a number of functional components to the system interface which we will introduce in this section.

Solvers are MRTK components that facilitate the means of calculating an object's position and orientation according to a predefined algorithm. Furthermore, the solver system can help the system to define an order of operations for the transform calculations as Unity could not offer a way to specify the update order for components. In our system, we used solvers to help with tracking head and hands.

As for head tracking, we attached a solver with a radical view component to the system interface panels. This makes the panels always face to the user, thus it could be easier for users to check information on panels. Besides, we used the hand tracking solver to make the hand scale model follow user's hand palm.

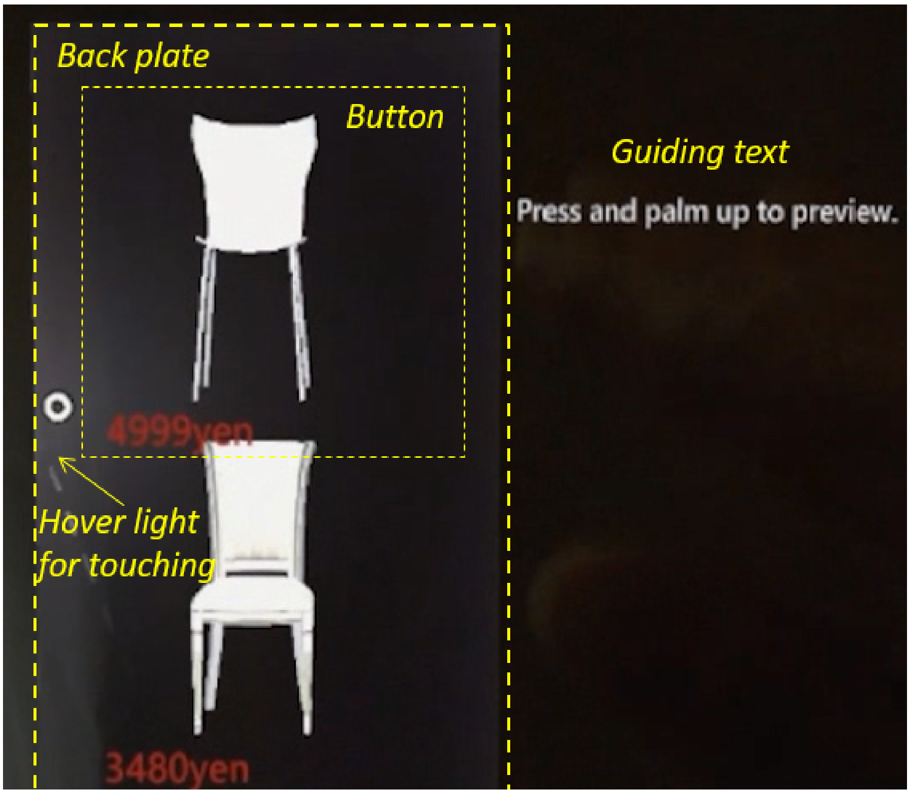


Fig. 6.9 Structure of result list

To build the result list, we combined four kinds of objects as contents: back plate, button, product model and clipping box. The back plate show the frame of the list with a close transparency background, which renders the buttons inside the plate. It receives pointer events related to scroll gesture, and it can also receive hand touch and pointing. We integrated product minimized models with price texts into buttons in the list to make is looks more intuitive. We added some guiding text beside the buttons which would be displayed when the button is being touched. There is an invisible front mask of buttons to show hover lights as hand touch feedback. Since our system receive more than one page result of searched products, to make the scrollable list show product information fluently, we use a 3D clipping box with the same length and width of the back plate to cut out extra buttons outside the back plate.

We also configured some properties of the result list related to the effect for the list being scrolled. The scrolling direction of the list is vertically, and the velocity properties for scrolling includes velocity dampen, velocity multiplier and type which is set to be fall off per frame.

As for the scaling operator we used to generate hand scale preview models, we set the object bottom plane size to be a 0.2 centimeters square, which is close to human's palm size. The algorithm we scale the virtual model is shown in Fig. 6.10.

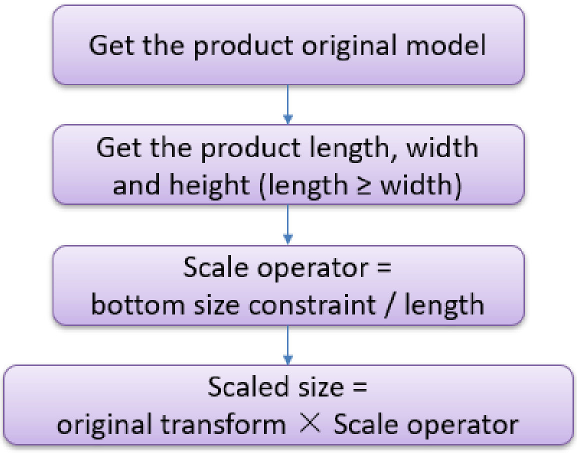


Fig. 6.10 Scaling algorithm for hand scale preview models

6.7 Preliminary Evaluation

In this section, we will introduce a preliminary user study of our system and the result in analysis. We asked our participants to use our system to shopping online in order to verify whether our system can provide better shopping experiences both on its preview experience and hand interactions when compared with the smartphone or PC based online shopping. Our discussion is based on the received results and received feedback from a questionnaire.

6.7.1 Participants

We invited in a total of 12 participants (6 females and 6 males), ranging from 20 to 27 years of age. All participants have the experience of shopping online and using AR applications.

6.7.2 Method

We introduced our system to participants firstly and let them experience head mount display basic operations. After a simple tutorial and question answering session, we let each of our participants use our shopping system for searching and previewing products.

After that, the participants will be asked to fill in a questionnaire as shown in Fig. 6.11 The questionnaire has following questions. Question 1 to 8 use the 5-point Likert scale. 5 options from strongly disagree to strongly agree represents the points from 1 to 5. In question 9, we let our participants to compare our shopping system with online shopping system on other platforms: smartphone shopping application and PC shopping website. Finally, we asked some comments from participants for using our system.

Questionnaire

Name: Gender:

Date: Age:

Questions

Question 1-8 are based on 5-point scale. Question 9 is a preference selection.
Answer the following questions by marking the most appropriate answer in your own perspective.

1. The system is easy to use.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

2. The system is easy to learn how to use.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

3. Gestures in our system are easy to use.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

4. The hand scale preview is useful.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

5. The real size preview is useful.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

6. It is easy to interact with the shopping system interface panels.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

7. It is easy to manipulate the virtual product model in hand scale preview.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

8. It is easy to manipulate the virtual product model in real size preview.

   Strongly Disagree ○---○---○---○---○ Strongly Agree

9. Which one you prefer for online shopping:

   - Computer based shopping websites.
   - Smartphone based shopping apps.
   - Our mixed reality shopping system.

Fig. 6.11 Questionnaire

6.7.3 Result & Discussion

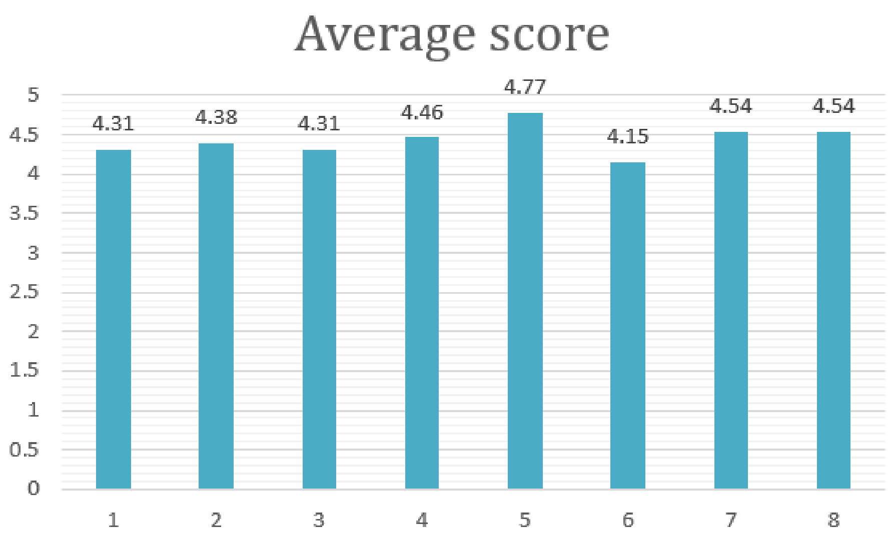


Fig. 6.12 Average Score of Question 1 to 8

In Fig. 6.12, we analyzed our evaluation result of question 1 to 8 through the average score of each question. All the average score is more than 4.0, which means that our participants thinks our system is easy to use and provide them with better shopping experience in preview and interactions.

Question 1 to 3 are used to evaluate our system overall operation difficulty for users. The average score of them are 4.31, 4.38 and 4.31, which suggests that users would not feel much difficulty of using our system or learning the basic operations.

Question 4 and 5 are used to evaluate user's demand for the preview interface in our system. The scores of Q4 is 4.46 and the score of Q5 is 4.77, which shows that users are requiring preview experience when shopping online. It also reflects that mixed reality shopping system is really needed for online shopping as it gives better preview experience.

In question 6 to 8 we evaluate the interaction methods for the system interface panels, hand scale preview and real size preview. The average scores are 4.15, 4.54, 4.54, which shows the interactions in our system is understandable and easy for users to interact with virtual contents.

According to the result of question 9, 91.7% participants prefer to use our system for shopping comparing with smartphone and PC platforms. The comments also shows that users like to view 3D virtual products at home, this helps them gain a better sense of how the product looks like. Besides, they think the hand scale preview function is useful for seeing product details and the manipulation methods are easy for them to use.

Overall, we got positive feedback from the preliminary user study result, we can consider our system preview interface and gesture interactions are easy and useful, and it provides users with a better shopping experience.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this research, we proposed a mixed reality shopping system with a hand scale preview interface and a new hand gesture set for interacting with MR shopping system. Based on the hand tracking technology, we display the virtual model over user's detected hand palm during the hand scale preview, to allow users observe the product in a suitable size at close distance. We also designed a hand gesture set for interacting with our system interface and virtual product models. According to the gesture classification method, we divided our gestures into three levels to make the gesture set more structural. Besides, we compared our gesture set with a virtual reality shopping system's gesture set in the related work to show different requirements when users shopping in different environments.

As for the interactions in mixed reality system, we combined hand gestures with pointer events to fire output events. We configured our system to generate output events during different period of the gestures. Besides, we used near and far interactions for enabling two kinds of interaction methods with related feedback, in order to help users easily control the virtual object they intend to interact with. We also applied voice search to extract keyword of products from user's speech. In addition, we designed a 3D shopping interface containing 3D UIs and product mini-sized models for showing product information.

In summary, according to our work on the system and the preliminary evaluation result, we think our system provides user a better shopping experience especially with its intuitive preview interface and flexible hand interactions.

7.2 Future Work

In the future, we plan to improve our mixed reality shopping system. We will add more gestures to our gesture set to fit more needs of user in online shopping. We will also add more functional features to make it more convenient to use. Furthermore, we will keep doing evaluations to compare the shopping experience of our system with other mixed reality shopping systems.

References

[1] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. vol.E77-D(12. 1321-1329), 1994.

[2] Microsoft Documentation. What is mixed reality? https://docs.microsoft.com/en-us/windows/mixed-reality/mixed-reality., 2018.

[3] Microsoft. Hololens. https://www.microsoft.com/ja-jp/hololens/developers.

[4] IKEA. IKEA place. https://www.ikea.com/au/en/customer-service/mobile-apps/say-hej-to-ikea-place-pub1f8af050.

[5] Amazon. View in your room. https://www.amazon.com/adlp/arview.

[6] Yuhui You and Kelvin Cheng. Rare shop: Real augmented retail bookshop experience using projection mapping. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, page 415–420, New York, NY, USA, 2019. Association for Computing Machinery.

[7] Jiro Tanaka Hao Dou. A mixed-reality shop system using spatial recognition to provide responsive store layout. In *22nd International Conference on Human-Computer Interaction (HCI International 2020)*, 2020.

[8] Masaya Ohta and Nagano. Mixed-reality web shopping system using panoramic view inside real store. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*, SA '15. Association for Computing Machinery, New York, NY, USA, 2015.

[9] Masaya Ohta and Nagano. Mixed-reality shopping system using hmd and smartwatch. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, page 125–128, New York, NY, USA, 2015. Association for Computing Machinery.

[10] Luis E. Bautista R. and Maradei G. Augmented reality user interaction to computer assisted orthopedic surgery system. In *Proceedings of the 7th Mexican Conference on Human-Computer Interaction*, MexIHC '18, New York, NY, USA, 2018. Association for Computing Machinery.

[11] Gun Lee Huidong Bai and Mark Billinghurst. Free-hand gesture interfaces for an augmented exhibition podium. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, page 182–186, New York, NY, USA, 2015. Association for Computing Machinery.

[12] Duc-Hoang Vo and Huynh. Automatic hand gesture segmentation for recognition of vietnamese sign language. In *Proceedings of the Seventh Symposium on Information and Communication Technology*, SoICT '16, page 368–373, New York, NY, USA, 2016. Association for Computing Machinery.

[13] Hsieh and Yi-Ta. Tactile interaction gestures: A wearable hand tactile system for interacting with ubiquitous information. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, page 539–544, New York, NY, USA, 2015. Association for Computing Machinery.

[14] Kim Heesun and Lee Dongeon. Omni-gesture: A hand gesture interface for omnidirectional devices. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, IUI '18 Companion, New York, NY, USA, 2018. Association for Computing Machinery.

[15] Muriel Pressigout and Eric. Marchand. Hybrid tracking algorithms for planar and non-planar structures subject to illumination changes. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 52–55, 2006.

[16] Said Yacine Boulahia and Eric Anquetil. Dynamic hand gesture recognition based on 3d pattern assembled trajectories. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2017.

[17] Tobias Hollerer Taehee Lee. Handy ar: Markerless inspection of augmented reality objects using fingertip tracking. In *2007 11th IEEE International Symposium on Wearable Computers*, pages 83–90, 2007.

[18] Ioannis Kapsouras and Nikos Nikolaidis. Action recognition by fusing depth video and skeletal data information. In *Multimedia Tools and Applications 78, 2 (2019)*, page 1971–1998, 2019.

[19] Kim Chwee Lim, Swee Heng Sin. Nguyen, and Matthew Chua. Video-based skeletal feature extraction for hand gesture recognition. In *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, ICMLSC 2020, page 108–112, New York, NY, USA, 2020. Association for Computing Machinery.

[20] Ahmad Byagowi Barrett Ens. Combining ring input with hand tracking for precise, natural interaction with spatial analytic interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, page 99–102, New York, NY, USA, 2016. Association for Computing Machinery.

[21] Min Kyung Lee Mirae Kim and Laura Dabbish. Shop-i: Gaze based interaction in the physical world for in-store social shopping experience. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, page 1253–1258, New York, NY, USA, 2015. Association for Computing Machinery.

[22] Minghao Cai Kelvin Cheng Soh Masuko Jiro Tanaka Hao Dou, Zhinan Li. Show something: Intelligent shopping assistant supporting quick scene understanding and immersive preview. In *21st International Conference on Human-Computer Interaction (HCI International 2019),Lecture Notes in Computer Science*, volume 11570, pages 205–218. Springer, 2019.

[23] Juin-Ling Tseng. Development of a low-cost 3d interactive vr system using sbs 3d display, vr headset and finger posture motion tracking. vol.5:6–12, 08 2016.

[24] Chunmeng Lu and Jiro Tanaka. A virtual shopping system based on room-scale virtual reality. In *The Twelfth International Conference on Advances in Computer-Human Interactions (ACHI 2019)*, pages 191–198, 2019.

[25] Hazem Wannous Quentin De Smedt and Jean-Philippe Vandeborre. Skeleton-based dynamic hand gesture recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1206–1214, 2016.