Augmented Reality Shopping System Through Image Search and Virtual Shop Generation

Zhinan Li[1](✉), Ruichen Ma[1], Kohei Obuchi[1], Boyang Liu[1], Kelvin Cheng[2], Soh Masuko[2], and Jiro Tanaka[1]

[1] Waseda University, Kitakyushu, Japan
lizhinan@fuji.waseda.jp,
{mizutsu-ma,kohei76752109}@akane.waseda.jp,
waseda-liuboyang@moegi.waseda.jp, jiro@aoni.waseda.jp
[2] Rakuten Institute of Technology, Rakuten, Inc., Tokyo, Japan
{kelvin.cheng,so.masuko}@rakuten.com

Abstract. In this paper, we introduce an intelligent augmented reality shopping system that supports image search and 3D virtual shop generation. By capturing an image of the product that the user is looking at, and using image search, we can recommend similar products to the user. Based on the spatial understanding techniques, we developed a 3D virtual shop interface that would be automatically generated in the user's physical room, where products are arranged around the room at appropriate locations. With the virtual shop display, users would be able to interact with the products' life-size virtual models by full-hand manipulation. Our system provides more natural interactions such as grabbing and touching. We also carried out an evaluation study for using our system, the participants' feedback shows that our system provides a better shopping experience with its intuitive search process, immersive shop display, and natural interactions.

Keywords: Immersive shopping · Intelligent system · Image search · Spatial understanding · Full-hand manipulation

1 Introduction

After entering the 21st century, the advancement of Internet technology has promoted the vigorous development of e-commerce. Because of the flexibility and the ability to connect to a large information network, users are able to purchase products whenever and wherever they want [1, 17]. However, there exists limitations in current online shopping systems [2]. Users are usually limited to viewing product information in 2D, such as through photos on the PC or smartphone, instead of directly in the physical world [3]. Additionally, the method of searching products depends heavily on textural keywords. At the same time, many consumers still prefer offline shopping, which provide a more authentic shopping experience.

In our previous work, we have already developed an intelligent shopping assistant system in an augmented environment, which supports quick scene understanding of products' recommendation, using voice control and two-finger manipulations [4]. Based on

S. Yamamoto and H. Mori (Eds.): HCII 2020, LNCS 12184, pp. 363–376, 2020.
https://doi.org/10.1007/978-3-030-50020-7_26

this work, we observed that the 2D interface was inadequate in an immersive augmented shopping system, and two-finger manipulation [5, 6] is still very limited so that more natural interactions with the products' 3D models are needed in order to provide a better previewing experience.

Moreover, the search method was also limited to keywords extraction from voice input, and the use of the user's environment scene [3]. Users may sometime wish to purchase a similar product to the ones that they saw, as they go about their daily life, or that they have the desire to replace an old product. In these cases, simply using keywords to search for products is inadequate and imprecise. Understanding the shape and color of the product, and search by image search may be more effective.

Current online shopping systems are mostly limited to 2D information and usually cannot let users get an instant preview in the physical world [11]. Recently, emerging technologies that make use of Augmented-Reality (AR) and computer vision techniques brings new opportunities for intelligent shopping environments.

However, some consumers may still prefer offline shopping [12]. Offline shopping can provide a more authentic shopping experience. They can compare the size and shape of different products more clearly in offline shopping. Also, some offline furniture shop provides a sample room, where customers can have an immersive user experience. For example, in the real shop, the customer can see the true size of the products. To combine the benefit of online and offline shopping, we propose to bring the real shop to their home, virtually. Our system allows users to compare the color, shape, and size of each product as if they were in the real shop. Furthermore, customers do not need to leave their home. Users can place virtual items in their rooms directly to see if it fits their room style.

The system in our current work is aimed at building an augmented reality shopping system that supports image search, full-hand manipulation, and virtual shop generation. It can recommend similar products to users and bring the real shop to users' homes. Using our system, users will have an immersive shopping experience at their home and find it easier to interact with virtual product previews.

2 Goal and Approach

In this work, we envision an augmented reality shopping system that provides users with a better shopping experience at their own home. We consider the use of image search as well as 3D virtual shop display to enable a more effective way of searching a similar product and immersive previewing experience. At the same time, this system also involves hand-tracking based full-hand manipulation [7, 8], which provides a more natural interaction with the virtual shop interface and 3D product models.

Our image search component is aimed at recommending similar products to the product which the user is looking at. By capturing an image of the current product, the system can analyze the color and type of the item. We realized this by using the image search method. It first analyzes the image taken by the camera from the see-through type head-mounted display (HMD) [9, 14], recognize the objects in the image, then it would search the specific kind of products on the online shopping website.

Virtual shop generation is aimed at integrating an immersive shopping experience of the users. It allows users to decorate their own home into a virtual shop. The virtual object would then be automatically placed in the room according to the user's surroundings. For example, it would detect the largest flat surface in the room and show the first several products in the result in 3D form. We realized this with a spatial understanding method. We use the depth camera of the HMD to capture the surroundings and place the virtual products in the real room. Also, users can easily adjust the position of these 3D virtual objects and manipulate them manually.

Full-hand manipulation provides users with a more authentic experience of manipulating virtual products, users are able to interact with the interface and 3D objects through pushing, grabbing and touching, which is a more natural way comparing with two-finger manipulation in our previous work. This is realized by installing a full-hand motion detection device [10] on an ordinary HMD. These hand movements are also supported by the Mixed Reality Toolkit (MRTK) which we are using. Our system recognizes these hand movements by the hand tracking system and the operation result would be shown through the HMD.

This system connects all these features to make it easy for users to shop at their home as if in a real shop, or even better when they want to find a similar item and view the product in their rooms.

3 Intelligent Shopping Assistant

3.1 Image Search

In order to enable users to obtain information about the products around them at any time, we provide the image search feature. Compared with the previous manual text input method, this feature can help users get the information of the products around them faster. For example, if users want to buy a new television to replace their old one, one way is to type the letter "television", another way is to use image search. Obviously, taking a photo is faster than typing letters when using HMD.

When the user gives the voice command "Search this". The camera will take a photo of the current product. The image will be uploaded to the image search platform. Then

Fig. 1. A water bottle

the result will be returned to the user. As Fig. 1 shows, when the input product is a water bottle, the recommended products are also water bottles (Fig. 2).

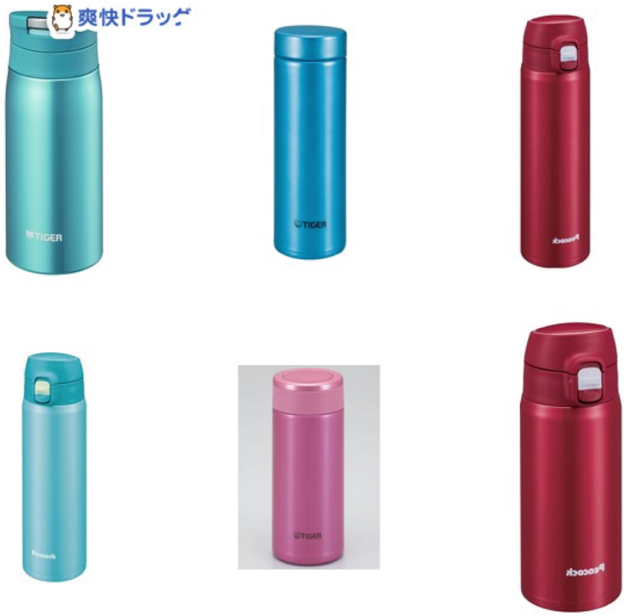

Fig. 2. Recommended products

3.2 Virtual Shop Generating

In order to let users visually see whether the products that they are considering are suitable in their current environment, we designed an approach to generate a virtual shop in the real world.

Before getting the recommended products, we analyze the current scene and extract the dominant objects in the real environment. By recording the location of these key objects, we can place the recommended items in the physical location nearby. Different products have their own placement rules. For example, decorative paintings and racks are usually hung on the wall. Vases are usually placed on a table or other flat surfaces, and chandeliers are usually hung from the ceiling. In order for the virtual object to automatically follow the placement rules for placement, we needed a higher level of understanding about the user's environment.

We solved this problem by introducing the spatial mapping function. Using this, we analyzed the basic structure of the user's current environment and identify surfaces such as walls, ceilings, and floors. After scanning the whole room, the system will notify users to stop scanning, and the spatial understanding component in Microsoft Mixed Reality Toolkit will analyze the scanned data. Through this analysis, we can get the orientation (vertical or horizontal), position and size of the corresponding surface in 3D space.

Currently, not all the products have 3D models (Fig. 3). In these cases, 3D cubes with 2D images are being used instead (Fig. 4).

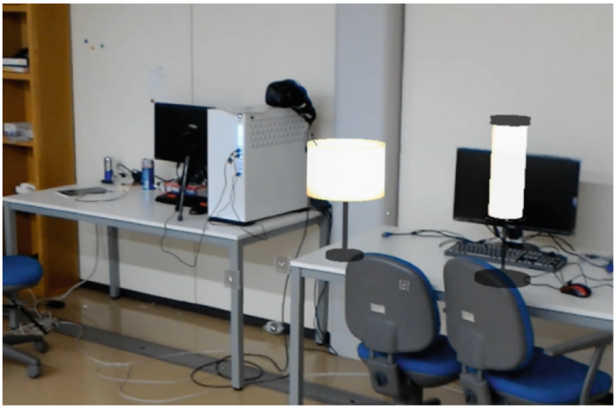

Fig. 3. 3D models displayed in the system interface

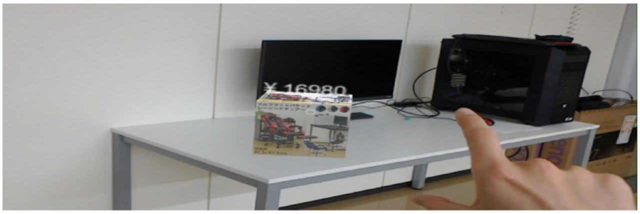

Fig. 4. 3D cube with 2D picture

3.3 Natural Interactions with Full-Hand Manipulation

In order to interact with an augmented reality system interface, hand gestures are often used for specific manipulation. In this work, it is necessary to use hands to move or rotate 3D objects for the preview experience. Some of the augmented reality (AR) systems use two-finger manipulation [5], like air-tap gesture. However, in daily life, we often use the entire hand for grabbing and placing objects. Two-finger manipulation does not conform to human natural behaviors. Therefore, monitoring and simulating full-hand manipulation is important to make it closer to human daily movements and let users act naturally when using our system.

In our previous work, we applied two-hand manipulation in the shopping system for dragging and rotating 3D virtual objects, but it was based on two-finger gestures: air-tap and hold [9]. This is very limited since it does not have much flexibility, we need to use both hands if we want to do a rotating operation.

To improve the manipulation method for interacting with our virtual shop interface and 3D virtual objects, we adopted an approach to enable hand motion detection by connecting an optical hand tracking module (Leap Motion) to an augmented reality HMD (Microsoft HoloLens). Apart from this, we applied Microsoft Mixed Reality Toolkit Version 2 (MRTK v2) [13] to enable user experience components (Table 1).

Table 1. Interactions of 3D objects

Interactable 3D objects	Related UI components	Input interactions
Buttons	Pressable button	Press
		Touch
Virtual products	/	Grab
	/	Air-tap
	Bounding box	Scale
		Rotate
	Tooltip	Gaze

As for the buttons in our system interface, we choose to use 3D pressable buttons (Fig. 5) to replace the previous 2D tap buttons. The pressable buttons provides feedback for touching and pressing. With pressing and touching enabled, the user can use one finger to interact with buttons, which is much easier than using an air-tap gesture.

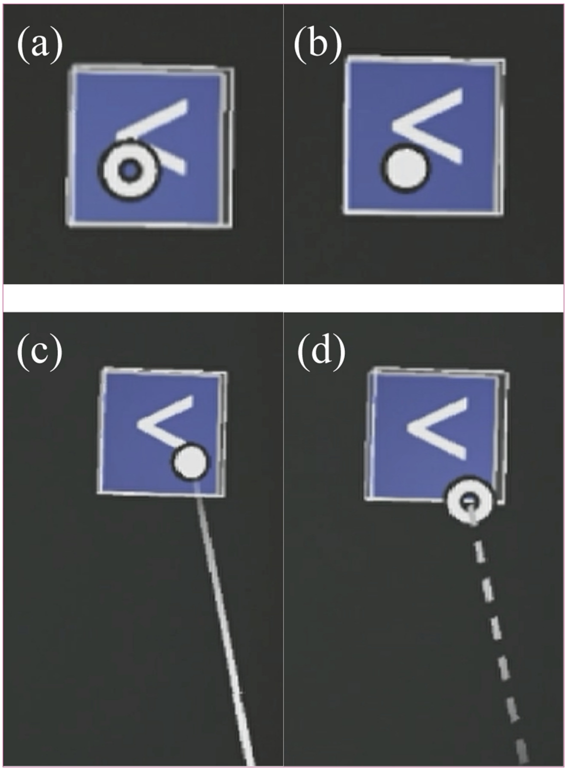

Fig. 5. Pressable button: (a) Near interaction, being released. (b) Near interaction, being pressed. (c) Far interaction, being released. (d) Far interaction, being pressed.

Since the virtual shop display is an important part of our system for previewing 3D virtual products in the user's room, we added plenty of interactions for manipulating these 3D objects (Table 1). In previous work, we need to use an air-tap gesture to move a virtual object, but now we can use the grabbing gesture as well. People usually grab things to move objects in daily life, so grabbing is a more natural way to interact with virtual objects. The grabbing gesture is not only be used for moving objects' position but also available for other direct manipulations like rotating the object by using hand movements. All these direct manipulations can be achieved by using the grabbing gesture (Fig. 6).

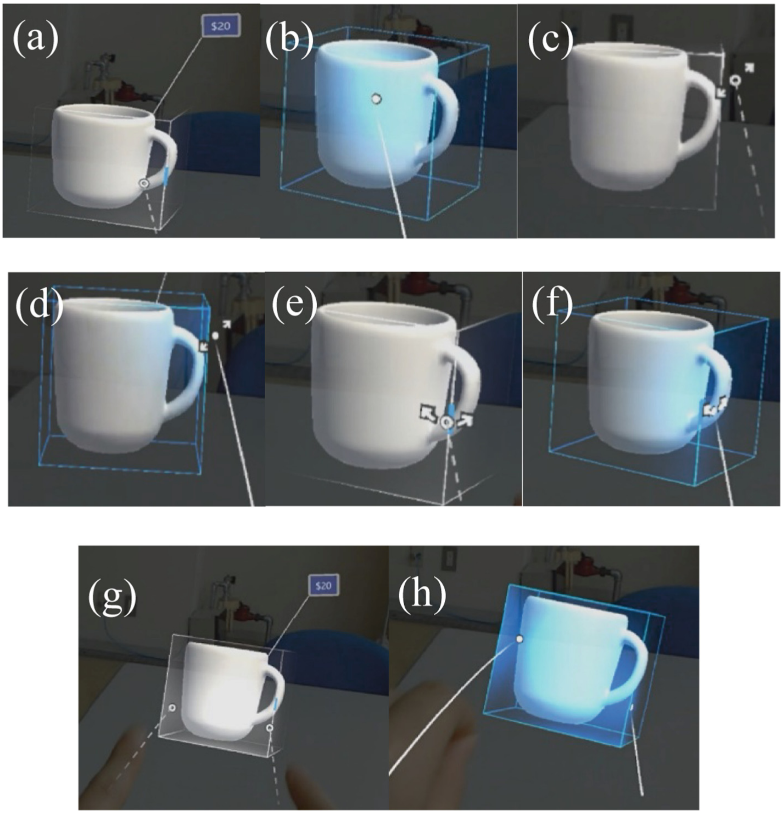

Fig. 6. Interactions of 3D virtual products: (a) Bounding box and tooltip. (b) Object being grabbed. (c), (d) Scaling the object by the bounding box. (e), (f) Rotating the object by the bounding box. (g), (h) Manipulating the virtual object with two hands. (Blue bounding box means object is being manipulated.) (Color figure online)

To make the operations understandable, we attached bounding boxes to virtual object models, which is invisible when the object is not being gazed at. The bounding box size is related to the size of the virtual objects. It enables handlers for scaling from eight apexes of the box and rotating from each edge's midpoint. We set the scaling range to be

1 to 5 times, according to the original scale of the object, so that users can see details by zooming in, and see the initial size by zooming out to the minimum size. With the new style of the bounding box, it is possible for users to do scaling and rotating operations with one or two hands, which is much more flexibility than our old system where the rotation operation must use completed with two hands. The accuracy of these operations is also improved by using the bounding box.

We also used tooltips for showing the price of the products. It is visible only when the object is being gazed at, so that users will not be overloaded with too much text or lines in the initial interface which may disturb the user's previewing experience.

Apart from close-up interactions such as touching and pressing, we also enabled distant interactions for both buttons and 3D objects, so that users can operate 3D objects from afar. Far interactions are supported by the hand tracking module, together with the feedback visualization from MRTK v2, which is shown as an imaginary line starting from the finger and ending with a circular pointer. Far interactions include grabbing, pointing and air-tap gestures. Each of them can be triggered by one or two hands. In our case, users can interact with buttons, bounding box or the object itself with the far interaction method, the user can feel more natural when interacting with our immersive shopping system.

3.4 Previous Work (Scene Understanding)

In previous work, we introduced a way to search for products based on the user's scene. By understanding the scene that the user is currently looking at, and extracting the detected scene information, the system can recommend related products that could potentially interest the user (Fig. 7).

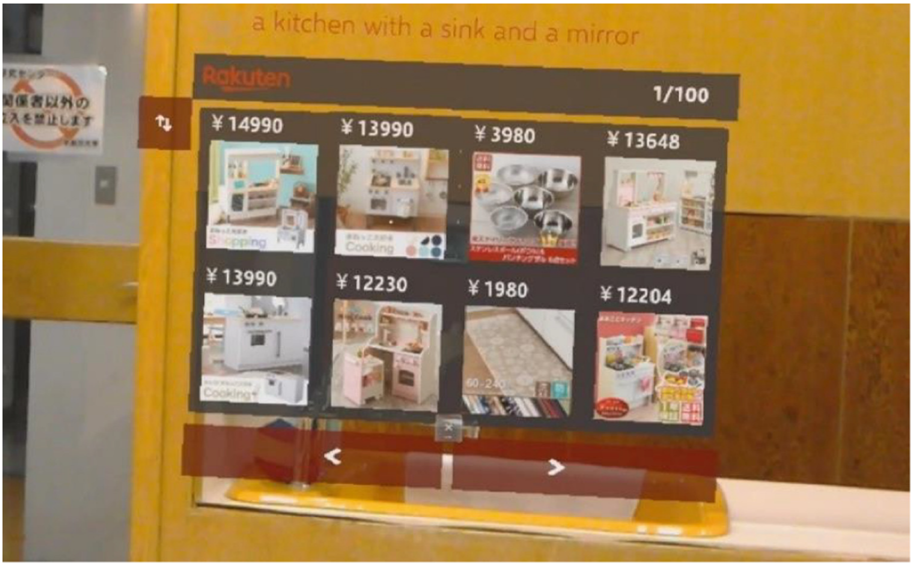

Fig. 7. The system recommends relevant products to the user according to the current scene in the user's environment. Scene recognition results: a kitchen with a sink and a mirror.

A scene is a view of real-world environment where users are physically located, in which contains multiple surfaces and objects being organized in a meaningful way. By applying computer vision techniques, the scene can be understood quickly. After generating an overall description of the scene (e.g. kitchen, or restroom) in this way, the system would be able to recognize dominated objects (e.g. desktop computer, or desk) from the ambient environment. The keywords of the current scene (e.g. kitchen) can then be used to search products on an e-commerce platform which contains the related categories as the dominated objects in the scene (instead of the dominated objects themselves). In this way, this system recommends users with products that are strongly associated with the current scene.

In our system, the recommendation process is triggered by a speech command, "Show Something". After the detecting and searching process, the results would then be displayed in augmented reality in a virtual panel that floats in the real world by an HMD. Users can visually see the links between recommended products and the real world, and view the details of the recommended items, or directly filter the recommended items by voice command and quickly find the products they are interested in.

4 Implementation

Figure 8 shows the overview of our system. Our system includes the HMD client and the server.

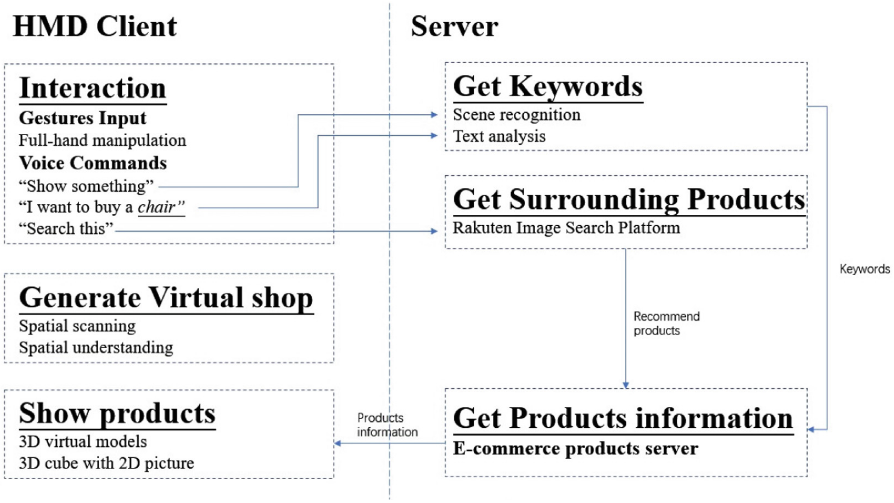

Fig. 8. System overview

4.1 Image Analysis

When the voice command "Search this" is issued, the camera of the HMD will capture the current product. The image will first be stored in the memory, and it will then be sent to the Rakuten Image Search Platform API. This API will analyze the image and return several similar products. We will visualize these products to the user.

4.2 Spatial Mapping

The system prompts users to walk around and start scanning the surrounding physical environment. Within the HMD, the camera group in front of the device are used to perceive the surface information in the surrounding environment. MRTK's spatial understanding component is used to analyze the metadata captured by the camera group to identify specific ceilings, floors, walls, and other flat information. Combined with the preset information for the virtual item placement feature, we calculate the location that is suitable for placing the virtual products and automatically place them.

4.3 MRTK v2

We built our system based on the Unity engine and Microsoft Mixed Reality Toolkit version 2 (MRTK v2). MRTK v2 for Unity is an open-source cross-platform development kit for accelerating the development of mixed reality applications. We used MRTK v2 for supporting camera system, input system, spatial mapping, UX and other functions. Compared with MRTK v1 we used in our previous system, MRTK v2 is modularized by different kinds of tools, which makes it possible for us to streamline our system. Since it configures many commonly used profiles in a foundation profile, our building procedure is simplified as well. Furthermore, our UX for full-hand manipulation and pressable buttons are supported by MRTK v2 which cannot be done in the previous version.

4.4 Hand Motion Detection

We used an optic hand tracking device, Leap Motion, to detect hand movements, and a see-through type HMD, Microsoft HoloLens 1, to detect user's head position information [15]. The Leap Motion controller is connected to a laptop PC through USB, and the PC and HoloLens are linked by Wi-Fi. We recreated a supporting frame [16] by 3D printing

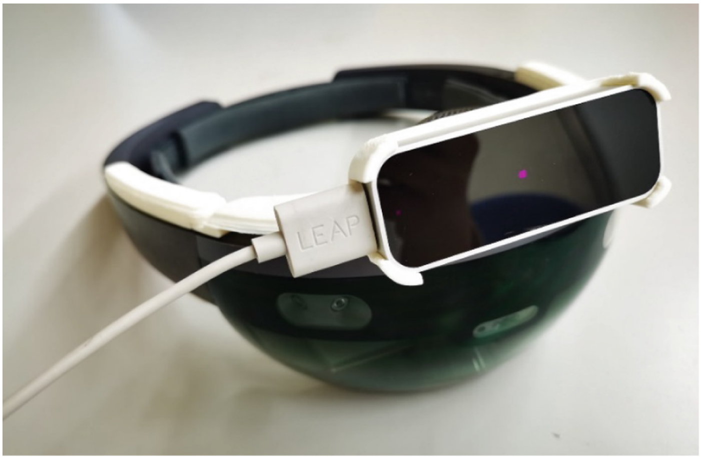

Fig. 9. Device assembly for attaching Leap Motion to HoloLens 1.

to fix the Leap Motion camera onto the HoloLens (Fig. 9) so that the Leap Motion camera would obtain a stable view as it can be moved with the HMD. Leap Motion camera captures the user's hand movements and sends it to the computer to analyze them, then exchanges the results with HoloLens to realize real-time feedbacks for hand interactions.

5 Preliminary Evaluation

In this section, we introduce our user study and result analysis. In our previous system, we asked participants to accomplish their shopping tasks. It shows the product in the form of a 2D panel. Users use two fingers to manipulate virtual products. The main purpose of this study was to test whether our new system can provide better display experience and better manipulate experience compared to the previous system. This study can help us figure out whether our system interests them. We will also discuss the received feedback from a questionnaire.

5.1 Participants

We asked 12 participants (5 females and 7 males), ranging from 20 to 26 years of age to participate in our experiment. All of them have basic computer skills. We divided them into two groups evenly.

5.2 Task and Procedure

Before the study, we introduced the basic operation of Microsoft HoloLens to each participant. We gave every participant 15 min to get familiar with the HoloLens. Then we asked them to search for three products (chair, water bottle, and monitor) in the room. Group 1 was asked to first use the old system for 15 min, and then use our new system for 15 min. Group 2 was asked the opposite, where they used the new system first. We asked all 12 participants to fill out a questionnaire with 5 questions to obtain qualitative feedback. Participants rated each question from 1 to 5 (1 = very negative, 5 = very positive).

5.3 Result and Discussion

After the experiment, we collected 36 sets of data from 12 participants using two systems. Then we analyzed the data we collected and performed preliminary evaluation.

As shown in Table 2, we divided the results into two parts, one is using our new system, and the other is using our old system. We calculated the average score of each question.

Question 1 judges whether the image search is faster. In our new system, the user tries to take a photo of the product and search. In the old system, users use voice command to search the product. From the result, we can see that when searching for a product that around us, the new system is faster than the old one. By using the image search, users can search the surrounding products quickly.

Table 2. Questionnaire

	Question	New system	Old system
1	I could search the products quickly	4.5	3.7
2	Search results are very accurate	4.5	4.5
3	I like the display interface very much	4.6	3.5
4	I felt the system operation was very easy	4.0	3.7
5	Gestures are very close to the real world	4.0	3.4

Question 2 is used to judge the accuracy of our new system. As we can see from the results, the participants gave the new system the same rating as the old one. It shows that although we use image search, the search results are not worse than the old system. Combining question 1 and question 2 we can know that we improved search speed while still maintaining high search accuracy.

Question 3 is used to compare the display interface of these two systems. The old system uses a 2D panel to show the products and the new system uses a 3D cube. As the result shows, most of the participants thought that the interface of our new system is better.

Question 4 and Question 5 is used to judge whether our new manipulate method MRTK v2 performs well. The new system's rating is higher than the old one. This is likely because when using the old system, the user has to use two fingers to manipulate, which is not similar to real life. Our new system uses full-hand manipulation, which is more realistic.

In general, all participants rated our new system higher than the old one. This signifies that our new system design is more reasonable and practical than the old one. It demonstrates that our new system can provide a faster, better and more realistic shopping experience than the previous version.

6 Related Work

6.1 Intelligent Shopping Assistant System

One of our related work is the intelligent shopping assistant system, which is also our previous work [3]. This work introduces an immersive shopping system supported by quick scene understanding and augmented reality 3D preview. This system can recommend related products that the user might interest in by understanding the scene the user is looking at. After returning the results of potential target products, the system provides users with an augmented reality preview experience. It automatically puts products to a suitable place in front of the user by using life-size 3D virtual products and spatial understanding. Users can use two-hand gestural manipulation to operate virtual products. It also allows searching and filtering for specific products by voice commands and keywords. We updated this system to a new environment to support MRTK v2 new UXs and more natural hand interactions to replace old type interface and two-hand manipulation methods. Our current work also extends the automatic placement method to generate our virtual shop interface at a proper position.

6.2 Spatial Recognition Based MR Shopping System Providing Responsive Store Layout

Another related work is a spatial recognition based mixed-reality (MR) shopping system providing responsive store layout [18]. This work is about building an MR shop system which could recognize the space and generate a virtual shop in the real environment. This system provides immersive virtual preview and interaction in the real environment. It also includes some in-store characteristics such as store layout, decoration, music, and a virtual store employee. It describes a new kind of online shop system by mixing virtual in-store characteristics and real environments, which may be the possible direction of the future online shop system. This research also introduced a new spatial understanding algorithm and layout mechanism to support a responsive spatial layout. This work inspired us to use spatial understanding techniques to generate our 3D virtual shop at a proper place in order to provide users with a better preview experience. In our work, we used spatial understanding to detect different kinds of surfaces in the user's room by the HMD and generate the augmented 3D virtual shop at a large platform in the user's environment, so that the user can have an immersive shopping experience.

7 Conclusion and Future Work

An augmented-reality online shopping system which allows users to experience in-store shopping at home is proposed. The system can recognize a specific object by camera and recommend similar products through an HMD. This system is also configured with a method for generating a virtual shop at the user's home. Users can use their full hand to manipulate the virtual object when using this system. This helps users to perform manipulation such as dragging and rotating more conveniently and naturally, and therefore the system enable the user to have an immersive preview for checking the details through the real model of the product that they want to buy.

Currently, the system is a single-user system, so it is not possible for others in the same room to view the virtual shop display during one is searching and manipulating the products. This would cause inconvenience when several users want to exchange opinions during shopping. For future work, we will try to make it into a multi-user system, in order to support two or more users sharing their preview information in the same room.

References

1. Luo, P., Yan, S., Liu, Z., Shen, Z., Yang, S., He, Q.: From online behaviors to offline retailing. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 175–184. ACM, August 2016
2. Close, A.G., Kukar-Kinney, M.: Beyond buying: motivations behind consumers' online shopping cart use. J. Bus. Res. **63**(9–10), 986–992 (2010). Lu, Y., Smith, S. (2007)

3. Zhao, Y., Guo, L., Wang, X., Pan, Z.: A 3D virtual shopping mall that has the intelligent virtual purchasing guider and cooperative purchasing functionalities. In 8th International Conference on Computer Supported Cooperative Work in Design, vol. 2, pp. 381–385. IEEE, May 2004
4. Dou, H., Li, Z., Cai, M., Cheng, K., Masuko, S., Tanaka, J.: Show something: intelligent shopping assistant supporting quick scene understanding and immersive preview. In: Yamamoto, S., Mori, H. (eds.) HCII 2019. LNCS, vol. 11570, pp. 205–218. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22649-7_17
5. Tanikawa, T., Arai, T., Masuda, T.: Development of micro manipulation system with two-finger micro hand. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 1996, vol. 2, pp. 850–855. IEEE, November 1996
6. Liu, J., Au, O.K.C., Fu, H., Tai, C.L.: Two-finger gestures for 6DOF manipulation of 3D objects. Comput. Graph. Forum **31**(7), 2047–2055 (2012)
7. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**(5), 6380–6393 (2013)
8. Jailungka, P., Charoenseang, S.: Intuitive 3D model prototyping with leap motion and microsoft HoloLens. In: Kurosu, M. (ed.) HCI 2018. LNCS, vol. 10903, pp. 269–284. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91250-9_21
9. Microsoft HoloLens | Mixed Reality Technology for Business (2019). https://www.microsoft.com/en-us/hololens. Accessed 14 Dec 2019
10. Leap Motion (2019). https://www.leapmotion.com/. Accessed 14 Dec 2019
11. Lu, Y., Smith, S.: Augmented reality e-commerce assistant system: trying while shopping. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 643–652. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73107-8_72
12. Levin, A.M., Levin, I.P., Weller, J.A.: A multi-attribute analysis of preferences for online and offline shopping: differences across products, consumers, and shopping stages. J. Electron. Commer. Res. **6**(4), 281–290 (2005)
13. Getting started with MRTK version 2 - Mixed Reality (2020). https://github.com/Microsoft/MixedRealityToolkit-Unity/releases. Accessed 26 Jan 2020
14. Garon, M., Boulet, P.O., Doironz, J.P., Beaulieu, L., Lalonde, J.F.: Real-time high resolution 3D data on the HoloLens. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 189–191. IEEE (2016)
15. Wozniak, P., Vauderwange, O., Mandal, A., Javahiraly, N., Curticapean, D.: Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality. In: Optics Education and Outreach IV, vol. 9946, p. 99460P. International Society for Optics and Photonics, September 2016
16. Ababsa, F., He, J., Chardonnet, J.R.: Free hand-based 3D interaction in optical see-through augmented reality using leap motion, October 2018
17. Zhu, W., Owen, C.B., Li, H., Lee, J.H.: Personalized in-store e-commerce with the promopad: an augmented reality shopping assistant. Electron. J. e-Commer. Tools Appl. **1**(3), 1–19 (2004)
18. Dou, H.: Spatial recognition based mixed-reality shop system providing responsive layout. Unpublished master's thesis, Graduate School of Information, Production and System, Waseda University, Japan, September 2019