# ハンドジェスチャインタラクションを用いた複数ユーザによるお買い物インタフェース

彭　昶[†] 李　雲　爽[†] 岩　淵　志　学[†]

益　子　宗[††] 田　中　二　郎[†]

オンラインショッピングを通じて、より安価に商品を購入できたり、商品のさまざまな情報を得ることができる。そのため、オンラインショッピングの需要は近年急速に増加している。しかし従来の WEB 上の提示方法（テキスト、画像等）では、ユーザが商品のデザイン等の情報を把握出来ない、またリアルタイムに他人の推薦を受け取れないという問題がある。この問題を解決するため、我々は屋内環境において複数人利用が可能な、仮想ファッション商品（衣類、帽子、バッグなど）試着型ショッピングインタフェースを提案、実装した。本システムでは、ユーザは定義されているジェスチャーを用いて、大型ディスプレイ上の仮想オブジェクトを移動、選択することができる。そして、ユーザは他のユーザに選択した仮想オブジェクトを手渡すことができる。身体に仮想商品をフィッティングさせることによって、ユーザは実店舗における試着と似たフィードバックを得ることができる。

# Multi-User Shopping Interface Using Hand Gesture Interaction

CHANG PENG,[†] UNSEOK LEE,[†] SHIGAKU IWABUCHI,[†]

SOH MASUKO,[††] and JIRO TANAKA[†]

Because of the numerous benefits, online shopping has been rapidly increased in recent years. Users can get better prices or product variety through the online shopping. However, an interactive problem is that users cannot get real-time recommendation or try on products through the traditional operation method (text, image, etc.). To resolve this problem, we present an interactive shopping interface which supports multiple users to operate virtual fashion products (piece of clothing, hat, bag, etc.) in an indoor environment. In this system, user can select, move a virtual object through the defined gestures on a large display. Then, user can hand out the selected virtual object to another user. Finally, user can use gesture to fit the virtual object so that obtain a vivid feedback, as if wearing the real fashion product.

## 1. INTRODUCTION

In recent years, people prefer online shopping rather than go to a traditional shop. The online shopping interface is mainly based on images and text. The user looks at the commodity information through the input of the mouse and keyboard or performs interactions, including the purchase of products. However, the existing shopping interface has several limitations. First, the controller is static. In the existing shopping methods, the mouse and keyboard represent the controller. Only simple input, like clicking or typing, is possible. Second, the traditional operation methods can only process a single user's task. This means that users cannot cooperate with other people, and obtain only limited feedback. To solve this problem, we propose Coordinated Shopping which supports multi-users to perform real-time interactions together. In our approach, selecting, fitting, and delivering virtual objects are possible by using hand gestures. Two users can operate on the same selected object.

Our paper is organized as follows. In section 2, the overview of this system is introduced. In section 3, we illustrate 2 types of hand gestures that are used in the system and the method for recognizing and interaction. In section 4, we illustrate how hand gesture works in the designed operation flow. In section 5, we perform an evaluation of the system. Finally, we offer conclusions and consider future work.

† 筑波大学 大学院 システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba

†† 楽天株式会社楽天技術研究所
Rakuten Institute of Technology, Rakuten, Inc.

## 2. SYSTEM OVERVIEW

In this paper, we present a hand gesture interaction system for multi-user shopping. Through this system, the users can see themselves with the system interface on a big display (50 Inch). They can use 2 types of hand gestures to select, control, and fit virtual objects which are also displayed on the screen. The crucial part of this system is allowing two users to operate on the same selected object.

This system is proposed to be used in a local fashion shops. First, the users can obtain recommendations from friends or salespeople. Then, they can quickly try on the desired product (piece of clothing, hat, bag, etc.) while standing in the front of a large display without going to a real dressing room. Finally, the users can survey the compatibility with the fashion product while watching the combination effect.

### 2.1. USER INTERFACE

The user interface of Coordinated Shopping contains two operational zones: Activity zone and Panel zone.
Activity zone is a screen area that is designed to let users see themselves in the physical environment, and perform interactions with virtual objects through hand gestures.
Panel zone is a screen area that is designed to allow users to launch different functions quickly ("Buy it", "Dress Room", etc.)


Figure1. User Interface operational zones: Activity and Panel zone

### 2.2. SYSTEM DESCRIPTION

The system hardware is constructed with one desktop PC, one big display and one depth camera (Microsoft Kinect). This confirms that our system is easy to install and adopt in a room-based environment. As for the software, we use the Kinect for Windows SDK. We obtain the depth data, color video streaming, and skeleton joint positions from Kinect SDK (the hand gestures are based on our own algorithms).

The depth camera is set in the center of the big display. The distance between the ground and the depth camera is approximately 180 cm. By setting up the system working environment, we ensure a good and steady operating performance, so that users can be detected from an effective distance (160cm ~ 220cm).

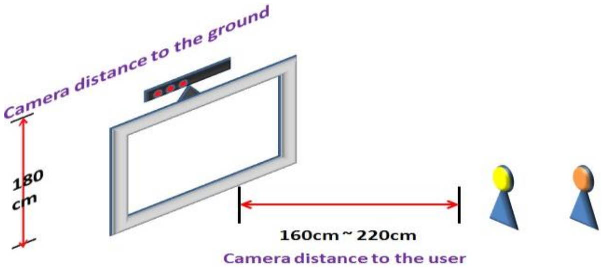



Figure2. System installation with big screen and depth camera, distance measured by the camera distance to the ground and camera distance to the user

## 3. HAND GESTURE

The aim of the system is to let users simulate a veritable shopping experience by interacting with virtual objects only by hand gestures, and with no wearable device requirement or complex construction.

### 3.1. HAND GESTURE TYPE

The recent studies show that the hand gestures can be of 3 types: Natural, Sign Language and Symbolic. This system has adopted the use of natural language which can express syntactic and semantic information directly and it is easy to remember for users.

### 3.2. HAND GESTURE RECOGNITION

We use depth camera to obtain the user's skeleton joint positions. We fetch some of the skeleton parameters (hands, shoulder, etc.), then calculate the hand's movement in different directions (x, y, z) to define hand gesture's toggle condition, and use skeletal tracking to fit the user's movement.

### 3.3. HAND GESTURE INTERACTION

#### 3.3.1. SWING HAND GESTURE

In this system, we use *swing hand gesture* to browse through virtual objects.

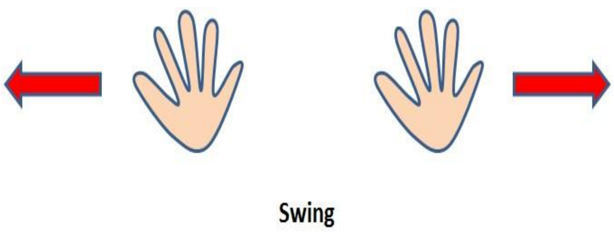
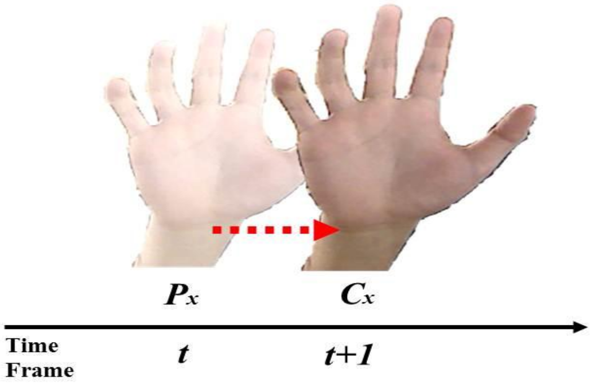



Figure3. Swing hand gesture: shift hand's position to the left or right to flip the virtual objects.

### 3.3.2. SWING HAND GESTURE PROCEDURE

To recognize Swing Hand Gesture, we can create a time frame so that we trace the hand movement continuously.




*Px* indicates previous hand's *X* position value, while *Cx* indicates current hand's *X* position value.

$$P_x - C_x < -Threshold \quad (1)$$

The threshold value is defined by the system and it is a standard value needed in recognizing the gesture.

In case (1), as above, the system recognizes the shift to the right gesture. *Px* is assumed to be 0; *Cx* is a positive number because of moving left to right.
Therefore, the value of (*Px* - Cx) is a negative number.

$$P_x - C_x > Threshold \quad (2)$$

In the same way, in case (2), the value of (*Px* - *Cx*) is a positive number. Consequently, the system recognizes a *Swing Hand Gesture* if the whole value of (1), (2) is in the Threshold value range.

### 3.3.3. PUSH HAND GESTURE

In this system, we use *push hand gesture* to select and fit the virtual object.

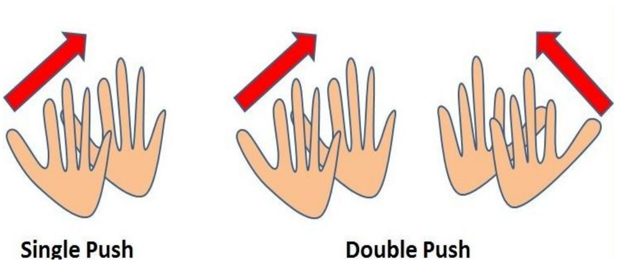
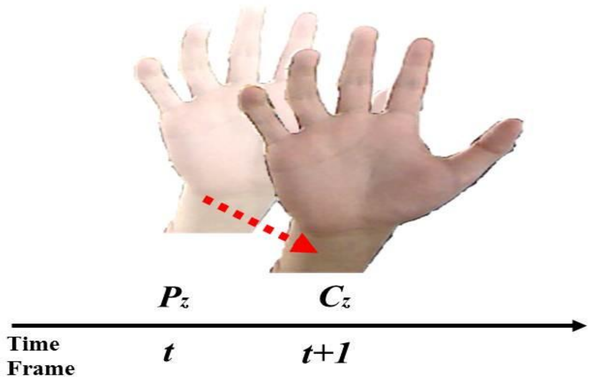



Figure4. Single Push hand gesture: shift one hand's position forward to select the virtual object. Double Push hand gesture: shift both hands' position forward to fit the virtual object.

### 3.3.4. PUSH HAND GESTURE PROCEDURE

For recognizing *Push Hand Gesture*, we trace the *Z*-axis value on time frame at first.




*Pz* indicates previous hand's *Z* position value, while *Cz* indicates current hand's *Z* position value.

$$(P_z - C_z > Threshold)\ AND\ (areCloseXY()) \quad (3)$$

In the *Z*-axis coordinates, *Pz* is assumed to be 0, *Cz* is a negative number. The system will check the value change of *Z* which is based on a certain threshold value. *Push Hand Gesture* checks the boolean function *areCloseXY()* while comparing with *Swing Hand Gesture*. The *areCloseXY()* is defined to check the change of *X*,*Y* value in the time frame. In other words, it will compare the *X*,*Y* value of the previous position and current position when *Z* value is changed. The *areCloseXY()* returns "True" only when *X*,*Y* value is changed in a specified range. When both conditions of (3) are met, the *Push Hand Gesture* will be detected. On the other hand, the Swing Hand Gesture may also cause the Z value to change in some cases. If the value of *areCloseXY()* is "False", the *Swing Hand Gesture* will be detected.

*Double Push Gesture* is determined through the same process with *Single Push Gesture*. This is done by checking left hand and right hand value simultaneously.

## 4. OPERATION FLOW

### 4.1. SELECT MODE

While user enters the camera detection area, the system state will automatically change to the *Selecting Mode*, and start to capture user's hand gestures. In addition, *Select Mode* will be displayed in the title column so that it can provide real-time feedback to the user. In the current state, the user can simply use *Swing Gesture* to flip the preview images of virtual objects. The current object being browsed will be set in the center of the defined array.

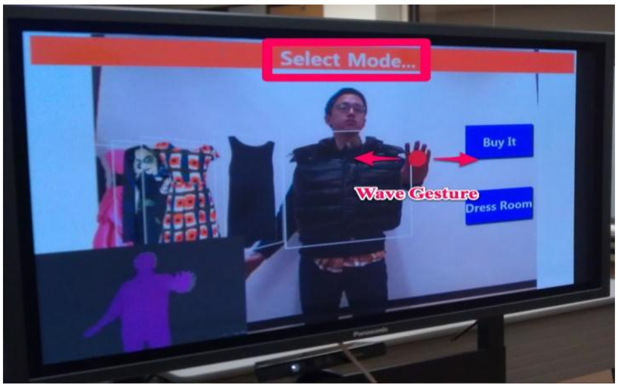



Figure4. Select Mode: User uses *Wave Gesture* to flip Virtual objects.

### 4.2. CONTROL MODE

After succeeding in finding the appropriate object, the user is able to use *Single Push Gesture* to select the desired virtual object accurately. At this stage, *Select Mode* will switch to *Control Mode* seamlessly. In the current state, the selected object will start to move along with the hand's movement. The user can control the object and move it to the appropriate position on the big display.


Figure5. Control Mode: *Single Push Gesture* to select the Virtual object.

### 4.3. FITTING MODE

Through the hand's movement, the user can place the selected object in the preferred location. In this state of affairs, the user can use *Double Push Gesture* to switch to the *Fitting Mode*. The virtual object will fit the user's body position and movement.

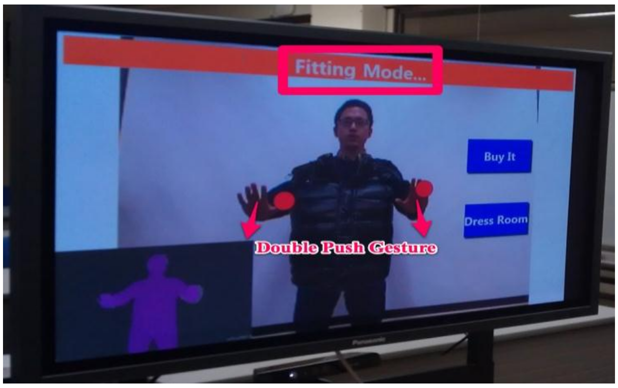



Figure6. Fitting Mode: User uses *Double Push Gesture* to fit the virtual object with the body.

### 4.4. MULTI-USER OPERATION


Figure7. Multi-user cooperation with interactions based on hand gestures. (*Giving and Receiving -> Operating*)

The scenario to be proposed assumes that user A and user B do apparel shopping together. User A uses the *Wave Gesture* and *Single Push Gesture* in the *Select Mode* and selects the virtual object for user B, then drags it to the appropriate location. Through the depth image highlighted in red in Figure 7, users can confirm that user A completely transfers the selected object to user B.


Figure8. Interaction with the object received from user A.

User B controls the object location appropriately and fits the virtual object on his/her body by using the *Double Push Gesture*.

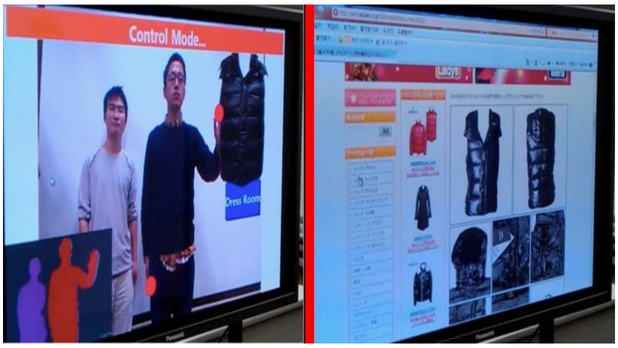



Figure9. Browse the details of the object on the Internet.

User B decides to buy the selected object, and drags the virtual object to the "But It" button, which starts the Internet browser to view the details of the product.

## 5. EVALUATION

We perform an evaluation to observe the practical situation where multiple users make use of the system.

We asked five people to use our system and answer a questionnaire that includes questions about system satisfaction, improvement, and user's feedback.

The participants are all university students, aged 23~25; three people are right-handed and two people are left-handed.

### 5.1. QUESTIONNAIRE

At first, we allow users to become familiar with the hand gestures before performing multi-user interactions. For this purpose, 5 users practiced the whole set of gesture (*Wave Gesture*, *Single Push Gesture*, *Double Push Gesture*). Then, we proposed an evaluation experiment, and asked users to evaluate the system through usefulness, intuition, ease of operation and overall satisfaction.

| Type / User | Usefulness | Intuition | Ease of Operation | Overall Satisfaction |
|---|---|---|---|---|
| A | 3 | 5 | 5 | 4 |
| B | 4 | 5 | 5 | 5 |
| C | 3 | 5 | 4 | 4 |
| D | 3 | 5 | 5 | 3 |
| E | 4 | 5 | 5 | 5 |

1 : Very Bad 2 : Bad 3: Normal 4:Good 5: Very Good

Table1. Questionnaire for System

We obtained feedback from the 5 users and most of them appeared satisfied with our system. We particularly obtained a high score in the intuition and ease of operation parts.

### 5.2. RESULTS

Through the questionnaire, we have learned that the users found our system interesting and with an intuitive interface. Comparing with previous shopping systems, the users found it is very easy to use as well.

However, users did not find our system very smooth yet. The reason is that while users use the fitting function, it does not seem very realistic. Users want to try virtual objects on the front, rear, and side of their body. In our system, the supported virtual objects are provided only for the front. Therefore, users find this kind of fitting not flexible compared to fitting real clothes.

Some users also require the system to adjust the graphic objects to fit their shoulder width automatically, as well as try on more graphic objects simultaneously. In our future work we aim to fit our system to these requirements and enhance the flexibility of our system.

## 6. RELATED WORK

Operating with virtual objects using hand gestures is a well-studied area of research in the recent works of human computer interaction. As related work, Head Mount Display (HMD) 7) is used to interact with 3D objects in a 3D virtual environment. The HMD can be used to locate the user's exposed facial skin. Using this information, a skin model is built and combined with the depth information obtained from a stereo camera. The information when used in tandem allows the position of the user's hands to be detected and tracked in real time. Once both hands are located, the system allows the user to manipulate the object with five degrees of freedom (translation in x, y, and z axis with roll and yaw rotations) in a three-dimensional virtual space using a series of intuitive hand gestures.

In our implementation, the processing of hand gesture is based on the depth data and skeleton data that is tracked by the depth camera. Users can directly operate the virtual object by using hand gestures without any hardware devices or complex arithmetic calculations.

This is the main difference between our system and other related systems.

## 7. CONCLUSIONS AND FUTURE WORK

In this paper, we presented Coordinated Shopping, an interactive shopping system that allows two users to select, pass and fit the same virtual object through hand gestures in a room-sized environment. Our work contributes a novel interactive shopping experience with multiple users.

The merits of our system are as follows.

First, we use "Title Column" to display the current operation mode and "Panel Zone" to support the user to switch to other functions frequently. This provides a vivid feedback to the users. Second, we mainly designed two types of hand gestures, which the users can remember with little effort. Third, the gestures that we used are natural and it is comfortable for users to perform interactions with virtual objects. Also, there is no physical stress while using the Swing or Push gestures. Finally, the gesture recognition is rapid and steady in a fixed range, with either one or two users.

Through the evaluation experiments, we obtained a positive feedback from the users. However, the results also show that we have the potential to enhance the interactive ability with multi-user operation.

In the next step, we plan to include more intuitive gestures using some other body regions (head, legs, etc.), to perform a more functional interaction with the system.


## ACKNOWLEDGEMENTS

This research was supported by the program for fostering advanced ICT professionals in the era of massive information contents, which was conducted by the University of Tsukuba and Rakuten, Inc.


## 8. REFERENCES


1) Kinect for Windows, http://kinectforwindows.org

2) J. Sullivan and S. Carlsson. "Recognizing and Tracking Human Action". Lecture Notes in Computer Science, vol. 2350, pp.629-644. ( 2002)

3) L. Bretzner, I. Laptev, T Lindeberg. "Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Partical Filtering". Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp.1-6. (2002)

4) R. Urtasun and P. Fua. "3D Human Body Tracking Using Deterministic Temporal Motion Models", Lecture Notes in Computer Science, Vol. 3023, pp.92-106. (2004)

5) N. Ukita, M. Kidode. "Wearable virtual tablet: fingertip drawing on a portable plane-object using an active-infrared camera". Proceedings of the 9th International Conference on Intelligent User Interfaces, IUI '04, pp.169-176. (2004)

6) Carl A. Pickering, Keith J. Burnham, and Michael J. Richardson. "Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction". Automotive Electronics, 2007 3rd Institution of Engineering and Technology Conference, pp.1-15. (2007)

7) C. Manders, F. Farbiz, T. Kayin, Y. Miaolong, B.Chong, and C. G.. Guan. "Interacting with 3D objects in a virtual environment using an intuitive gesture system". Proceeding VRCAI '08 Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, pp.1-5. (2008)

8) X. Zabulis, H. Baltzakis and A. Argyros. "Vision-based Hand Gesture Recognition for Human-Computer Interaction.". Computer Vision Techniques for Hand Gesture Recognition, pp. 25-35. (2009)

9) J. Wachs, M. Kölsch, H. Stern, and Y. Edan. "Vision-based hand-gesture applications". Communications of the ACM, Vol. 54 Issue 2, pp.60-70. (2011)

10) S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, J. Shotton, S. Hodges, and D. Freeman. "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera". ACM UIST Proceedings ACM User Interface and Software Technologies, UIST'11, ACM, pp.1-9. (2011)