2007

# Methods for Augmented Reality E-commerce

Yuzhu Lu
*Iowa State University*

## Recommended Citation

Lu, Yuzhu, "Methods for Augmented Reality E-commerce" (2007). *Retrospective Theses and Dissertations*. 15845.
https://lib.dr.iastate.edu/rtd/15845

# Methods for Augmented Reality E-commerce

by

**Yuzhu Lu**

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Shana S. Smith, Major Professor
Frederick O. Lorenz
Derrick J. Parkhurst
Viren R. Amin
Julie A. Dickerson

Iowa State University

Ames, Iowa

2007

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

# TABLE OF CONTENTS

# LIST OF FIGURES

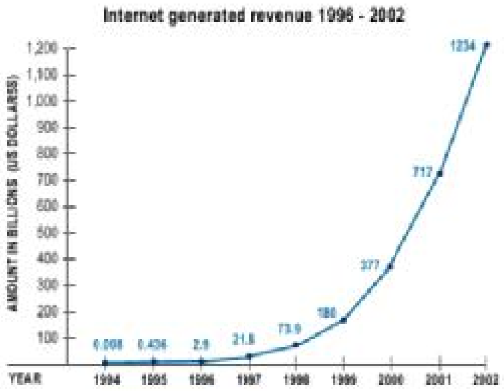
# LIST OF TABLES

# ABSTRACT

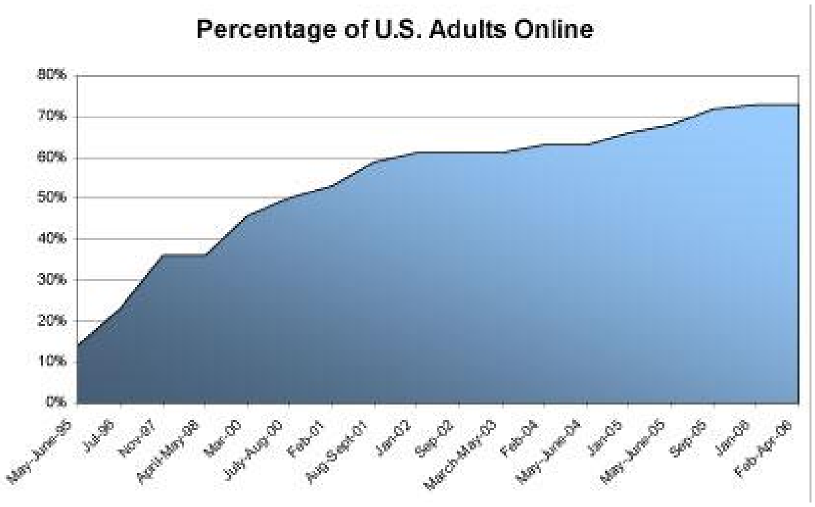
A new type of e-commerce system and related techniques are presented in this dissertation that customers of this type of e-commerce could visually bring product into their physical environment for interaction. The development and user study of this e-commerce system are provided. A new modeling method, which recovers 3D model directly from 2D photos without knowing camera information, is also presented to reduce the modeling cost of this new type of e-commerce. Also an immersive AR environment with GPU based occlusion is also presented to improve the rendering and usability of AR applications. Experiment results and data show the validity of these new technologies.

# CHAPTER 1. GENERAL INTRODUCTION

Electronic commerce (e-commerce) is defined as the online exchange of goods, services, and money within firms or between firms and their customers [1]. According to this definition, there are two main kinds of e-commerce: business-to-business (B2B), and business-to-consumer (B2C). Currently, e-commerce and online shopping are rapidly progressing, because of the convenience that was made available with the development of computer and internet technology as shown by the data from ActiveMedia (Figure 1.1), which became common in most households within recent years. E-commerce and online shopping make peoples' lives easier, especially for individuals with disabilities and for others who have difficulty engaging in onsite shopping. Amazon.com, Dell.com, and Ebay.com have become a part of our lives. Lefebvre [2] argued that e-commerce is growing faster than expected, and it is likely to have a dominant position in the future economy.



**(a) Internet-generated revenue** (Source: ActiveMedia)



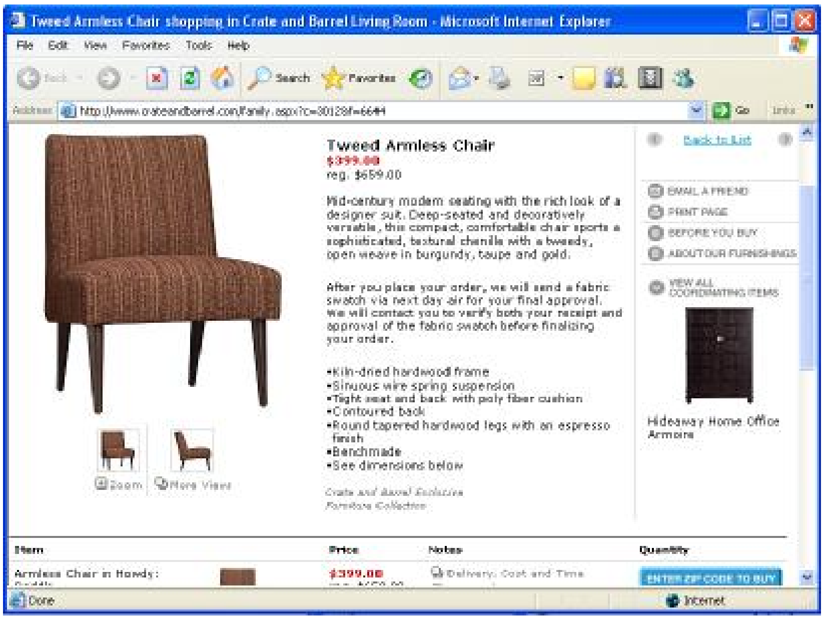
**(b) Percentage of US adult online**

**Figure 1.1 Internet-generated revenue and Percentage of US adult online**

However, according to our experiences, e-commerce and online shopping are still not able to fully replace onsite shopping, especially for products like clothing, shoes, jewelry, and furniture. For many products, onsite shopping has many distinct advantages when compared to online shopping. One of the main advantages is that online shopping does not usually provide enough information about a product for the customer to make an informed decision before checkout. Onsite shoppers frequently engage in some sort of interaction with their potential purchase to discover the scent, texture, appearance, and/or sound before buying it. This experience is often impossible with online purchases.

Cho, Im, and Hilts [3] conducted research to analyze sources and causes of online shoppers' complaints. Based on their research, the two biggest factors that caused complaints were product- and customer service-related. The authors also found that, for online purchases, there was a higher percentage of complaints about clothing and shoe products than about other kinds of products because of the lack of additional information that onsite shopping could have provided.

What kinds of information can traditional e-commerce systems provide for their customers? What kinds of information are they unable to provide? Figure 1.2 shows a B2C e-commerce example, a furniture website that provides some product-related information—such as a picture, feature description, product details, and customer reviews—to give as much information as possible to the customers. However, a picture and a description cannot always provide enough information for consumers to make a good decision. Many times people only have a rough idea about the size of a product from the information provided on the website and finally find it too large after buying it. Also, customers may also have a difficult time deciding if a product's color or design will match other objects in their home.


Tweed Armless Chair shopping in Crate and Barrel Living Room - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address
http://www.crateandbarrel.com/family.aspx?c=3012&f=6644
Go
Links
Back to List
EMAIL A FRIEND
PRINT PAGE
BEFORE YOU BUY
ABOUT OUR FURNISHINGS
VIEW ALL COORDINATING ITEMS
Hideaway Home Office Armoire
Zoom
More Views
ENTER ZIP CODE TO BUY
Done
Internet

Tweed Armless Chair
$399.00
reg. $659.00

Mid-century modern seating with the rich look of a designer suit. Deep-seated and decoratively versatile, this compact, comfortable chair sports a sophisticated, textural chenille with a tweedy, open weave in burgundy, taupe and gold.

After you place your order, we will send a fabric swatch via next day air for your final approval. We will contact you to verify both your receipt and approval of the fabric swatch before finalizing your order.

•Kiln-dried hardwood frame
•Sinuous wire spring suspension
•Tight seat and back with poly fiber cushion
•Contoured back
•Round tapered hardwood legs with an espresso finish
•Benchmade
•See dimensions below

*Crate and Barrel Exclusive Furniture Collection*

| Item | Price | Notes | Quantity |
|---|---|---|---|
| Armless Chair in Howdy: | $399.00 | Delivery: Cost and Time | |

**Figure 1.2 An e-commerce web page for furniture** (source: crateandbarrel.com)

Generally, many products such as books, software, tickets, music, and computers are suitable for online shopping because their features and related important information are easy to gain from written descriptions. However, for others—like clothing, shoes, jewelry, and furniture—people are more likely to prefer onsite shopping, because online shopping cannot provide adequate information to customers. The development of e-commerce for products like clothing, shoes, jewelry, and furniture is far behind what is needed in the e-commerce world.

Is there any technology that could improve e-commerce and provide more intuitive information so that customers could make proper buying decisions? Many researchers have been using virtual reality (VR) in e-commerce to provide consumers with a new type of mediated experience—a virtual consumer experience. “A virtual consumer experience is a psychological and emotional state that consumers undergo when they interact with products in a 3D environment or within intelligent agents” [4]. It has the potential to rich consumers’ experience [4]. Hughes, Brusilovsky, and Lewis [5] presented an adaptive navigation support system for using a virtual environment for online shopping, to help ease the consumer’s navigation and to help the consumer focus on important product features. They used some detailed technologies—such as direct guidance, hiding, sorting, and annotation—for the adaptive navigation. Cássia, and Fernando [6] presented an adaptive 3D virtual environment with structure and content that changed according to the user’s interest and preference and according to its application in e-commerce. They integrated intelligent agents, user models, and automatic content categorizations together in their system. Chittaro and Ranon [7] presented two design guidelines (massive and walking product) to improve the usability of VR stores. The authors also analyzed and discussed how to apply the walking product to personalized VR store for guidance. Their usability experiment results showed that the “massive” did affect the shoppers’ purchasing of products, and that the “walking product” significantly improved the shoppers’ convenience when finding a product [8]. Shen and Georganas [9] presented a multi-user collaborative virtual environment system and described application of the system in industry training and e-commerce. In their e-commerce system, users could communicate with each other about the product. Bogdanovych, Berger, and Simoff [10] focused on developing the social interface of VR e-commerce by combining electronic institutions and virtual worlds to make use of their advantages. Communication problems in their 3D electronic institution were analyzed and discussed as an important topic. In Sanna, Zunino, and Lamberti [11]’s presented VR e-commerce system which base on

Virtual Reality Modeling Language (VRML), an animated virtual human was introduced. The virtual human was to be used as a virtual shopping assistant for helping online shoppers navigate through an e-commerce environment. The researchers used Quick 3D to generate 360-degree image-based backgrounds for immersion. Daugherty, Li, and Biocca [4] conducted five experiments to study the usability of VR in e-commerce. Their results showed that using their system designed for testing, users gained virtual experiences and acquired significantly more information with their virtual experiences about the product than from their pre-existing indirect experiences.

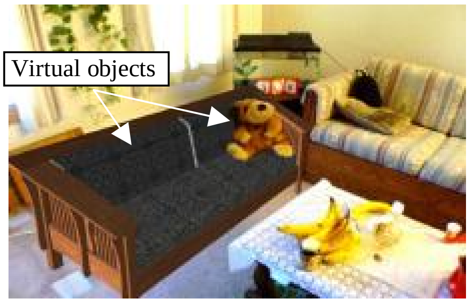


**Figure 1.3 Augmented Reality**

Although prior studies show that VR can enhance e-commerce by providing virtual experiences and product interactions, VR technology can still only provide a virtual, not a real, experience. It is important to provide consumers with “real experiences” and “real object interactions”. Augmented Reality (AR) is a technology which aims to mix or overlap computer-generated 2D or 3D virtual objects with the real world as shown in Figure 1.3. Unlike VR, which replaces the physical world, AR enhances physical reality by integrating virtual objects into the physical world. The virtual object becomes, in a sense, an equal part of the natural environment. In recent years, many researchers have focused on AR applications. In 2001, Azuma, Baillot, and Behringer [12] reviewed advances in AR, since 1997, including display devices and methods, indoor and outdoor tracking, model rendering, and interaction technologies. They indicated several problems that could be improved in the future, such as occlusion, broader sensing, advanced rendering, and users' perception issues. However, there has been minimal research conducted regarding the use of AR to enhance e-commerce.

There are two methods for tracking in existing AR research. The first method tracks both the camera and the users, including their head directions and gestures. This tracking method usually uses general tracking devices for indoor tracking and global positioning system (GPS) for outdoor tracking. However, AR that uses tracking systems is not the best solution to enhance e-commerce because it is inconvenient and most online shoppers can not afford it. The other tracking method recognizes and tracks markers or objects in the real scene by using computer vision and image-processing technology. Despite some limitations, this is a feasible solution to be used for enhancing e-commerce and visually bringing the product into online shoppers' home. Zhang, Fronz, and Navab [13] compared four existing AR marker systems—ARToolKit(ATK), Hoffman marker system (HOM), Institute Graphische Datenverarbeitung (IGD), and Siemens Corporate Research (SCR)—based on usability, efficiency, accuracy, reliability, and qualitative measures. From their experiment results, they found that there was no marker system that was obviously better than the rest.

Among the limited research regarding the use of AR in e-commerce, Zhang, Navab, and Liou [14] proposed a prototype direct marketing system that uses AR technology. Sales people could use the system to show the main features of a product by manually holding a plate with specially designed markers. A 3D virtual product mixed with a real scene could be video taped and sent to interested customers by email. However the researchers' method of combining AR with e-commerce did not fully use the advantages of AR. With their method, online shoppers would still not know whether a product was suitable for them and suitable to their real physical environment.

Motivated by the need to enhance current e-commerce systems, chapter two of this dissertation presents a new AR e-commerce system that “visually” puts a product model into the online shopper’s physical environment and gives the customer a chance to “realistically” interact with the product. A pilot user study and a formal user experiment were carried out. Experiment results show that AR e-commerce is a good solution that provides more direct information and experience to online customers and helps them make better buying decisions, even though the interaction, rendering, and modeling methods still need improving.

Making models for all products of e-commerce website would cost a lot and take a long time, which might be a core problem hindering AR e-commerce from being widely used. In chapter three of this dissertation, we also developed a convenient and low-cost way to recover 3D models directly from 2D images without knowing more information. A hierarchical matching method and a camera parameter estimation method were developed. Experiment results show that this model recovery method is quite feasible.

To improve the rendering method of AR e-commerce, make users feel more comfortable, and make the products seem more real, chapter four of this dissertation, presented an immersive AR environment with GPU based occlusion. This environment makes use of the advantages of CAVE-based facilities, uses remote stereo cameras to capture the background, and renders virtual objects in the stereo background. With stereo images, a real-time real-virtual object occlusion approach was developed and programmed into a GPU. The GPU-based method determines occlusion in real time by calculating depth information from the background of a real scene. As a result, the application intelligently knows which virtual object is in front of objects from video and which is behind them.

In chapter five of this dissertation, a case study is presented to show the process of recovering a product model via the method presented in chapter three, the process of model normalization, the process of its application in the AR e-commerce presented in chapter two, and also in the immersive AR environment presented in chapter four, which shows the strong feasibility of AR e-commerce.

## References

[1] Standing, C., 2000, “Internet Commerce Development”, Artech House, Hardcover, Published February 2000.

[2] Lefebvre, L.A., & Lefebvre, E., 2002, “E-commerce and Virtual Enterprises: Issues and Challenges for Transition Economies”, *Technovation*, 2002, 22(5), pp.313-323.

[3] Cho, Y., Im, I., Hiltz, R., & Fjermestad, J., 2002, “An Analysis of Online Customer Complaints: Implications for Web Complaint Management”, *Proc. of the 35th Hawaii International Conference on System Sciences*, January, Big Island, Hawaii.

[4] Daugherty, T., Li, H., & Biocca, F., 2005, “Experiential commerce: A summary of research investigating the impact of virtual experience on consumer learning”, Society of Consumer Psychology: *Online Advertising*. Mahwah, NJ: Lawrence Erlbaum Associates.

[5] Hughes, S., Brusilovsky, P., & Lewis, M., 2002, “Adaptive navigation support in 3D e-commerce activities”, *Proc. of Workshop on Recommendation and Personalization in E-Commerce at the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2002)* Malaga, Spain, May 28, 2002, pp. 132-139.

[6] Santos, C.T., & Osorio, F.S., 2004, “AdapTIVE: An Intelligent Virtual Environment and Its Application in E-Commerce”, *COMPSAC 2004*, pp. 468-473.

[7] Chittaro L., & Ranon R., 2000, "Virtual Reality stores for 1-to-1 e-commerce", *Proc. of the CHI2000 Workshop on Designing Interactive Systems for 1-to-1 E-Commerce*, The Hague, The Netherlands, 2000.

[8] Chittaro L., & Ranon R., 2002, “New Directions for the Design of Virtual Reality Interfaces to E-Commerce Sites”, *Proc. of AVI 2002: 5th International Conference on*

*Advanced Visual Interfaces*, ACM Press, New York, May 2002, pp. 308-315.

[9] Oliveira, C. Shen, X., & Georganas, N., 2000, "Collaborative Virtual Environment for Industrial Training and e-Commerce", *Workshop on Application of Virtual Reality Technologies for Future Telecommunication Systems, IEEE Globecom 2000 Conference*, Nov.-Dec. 2000.

[10] Bogdanovych, A., Berger, H., Simoff, S., & Sierra, C., 2004, "3D Electronic Institutions: Social Interfaces for E-Commerce", *In 2nd European Workshop on Multi-Agent Systems*, Barcelona, Spain, December 16-17 2004.

[11] Sanna, A., Zunino, C., & Lamberti, F., 2002, "HAVS: a human animated VRML-based virtual shop for e-commerce", *In SCI'02 Proc.*, vol. XII, pp.24-29.

[12] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B., 2001, "Recent Advances in Augmented Reality", *IEEE Comp. Graph. & App*, vol. 21, no. 6 (Nov/Dec 2001), pp. 34-47.

[13] Zhang, X., Fronz, S., & Navab N., 2002, "Visual Marker Detection and Decoding in AR Systems: A Comparative Study", *ISMAR 2002*, pp. 97-106.

[14] Zhang, X., Navab, N., & Liou S.P., 2000, "E-Commerce Direct Marketing using Augmented Reality", *IEEE International Conference on Multimedia and Expo (I)* 2000, pp. 88-91.

# CHAPTER 2. AUGMENTED REALITY E-COMMERCE SYSTEM: DEVELOPMENT AND USER STUDIES

A paper submitted to *the International Journal of Human Computer Studies*

Yuzhu Lu Shana Smith

1620 Howe Hall, 2274

Human Computer Interaction Program

Iowa State University, Ames, IA 50011-2274

yuzhu@iastate.edu sssmith@iastate.edu

**Abstract.** Traditional electronic commerce (e-commerce) is limited because it cannot provide enough direct information about products to online consumers. Online shoppers are often unhappy with the products and related customer service they receive because of the lack of interaction and try, which, on the other hand, onsite shopping can provide. In this study, an augmented reality (AR) e-commerce system was developed, using user-centered design principles. The tool was developed as an Internet plug-in, so it can be used on different kinds of computers and handheld devices. A usability study was also conducted to compare the developed AR e-commerce system with traditional e-commerce and virtual reality (VR) e-commerce systems. Study results show that the AR e-commerce system provides more information and more direct experiences to online customers, by combining physical environment information with virtual product models. As a result, the AR system can help customers make better purchasing decisions.

**Keywords**: Augmented Reality, Electronic Commerce, Usability Study.

## 1 Introduction

Standing (2000) defined e-commerce as the online exchange of goods, services, and money within firms and between firms and their customers. In the past decade, e-commerce and online shopping have become popular because they make life easier, especially for individuals with disabilities and for others who have difficulty engaging in onsite shopping. In 2002, Lefebvre (2002) showed that e-commerce was growing faster than expected, and that it was likely to have a dominant position in the future economy.

However, e-commerce and online shopping still cannot fully replace onsite shopping, especially for products like clothing, shoes, jewelry, and furniture. For such products, onsite shoppers frequently engage in some sort of interaction with their potential purchase before buying it to discover the product's scent, texture, appearance, fit, or sound. Unfortunately, such interaction is often impossible for online purchases. As a result, online shoppers, particularly when shopping for clothing and shoe products, are often unhappy with the products and related customer service they receive (Cho et al., 2002).

2D pictures or written descriptions used in traditional e-commerce systems often cannot provide enough product information. Thus, to improve e-commerce systems, it is important to develop systems which provide more sensory information, to help online shoppers make better purchasing decisions and, thus, improve customer satisfaction.

## 2 Background

### 2.1 VR in E-commerce

Virtual reality (VR) is a computer-simulated environment that allows users to manipulate 3D virtual models online. Recently, researchers have been using VR in e-commerce to provide consumers with a new type of shopping experience using virtual product models. Hughes et al (2002) presented an adaptive navigation support system for using a virtual environment for online shopping. Sanna et al. (2002) presented a VR e-commerce system based on VRML. They used QuickTime 3D to generate 360-degree image-based immersive backgrounds and an animated virtual human to help online shoppers navigate through their e-commerce environment. Bhatt (2004) analyzed the interactivity, immersion, and connectivity of several major VR-ecommerce websites, such as amazon.com, ebay.com, and schwab.com.

Daugherty et al. (2005) conducted five experiments to study the usability of VR for e-commerce. Their results showed that users acquired more information about products when using a VR-based e-commerce system than when using traditional tools. Fomenko (2006) developed a tool for creating online VR shops, which also gives domain experts more control during the website development process. With Fomenko's tool, developers can use high-level concepts to model and semi-automatically generate a complete VR shop.

## 2.2 Moving from VR to AR

Although prior studies show that VR can enhance e-commerce by providing more product information, through enhanced human-computer interaction, current VR methods for e-commerce still only provide scaled virtual product models displayed on traditional computer screens. New, more advanced, methods are needed to provide consumers with more realistic product models, with respect to size, customer experience, and user interaction.

AR is a technology which can mix or overlap computer-generated virtual objects with real-world scenes or objects. Unlike VR, which experientially replaces the physical world, AR enhances physical reality by integrating virtual objects into a physical scene. Generated virtual objects become, in a sense, an equal part of the natural environment.

In recent years, much research has focused on developing AR applications, which could be generally classified into two types based on the different devices used: optical see-through AR, and video see-through AR. Optical see-though AR uses semi-transparent screens to project computer generated objects, by which user could also see-through it to gain the integrated AR scene. Video see-through AR uses cameras to capture the live scene as videos. At each frame, video image is processed and computer generated object are added. The mixed scene of video see-through AR could be displayed on different devices. Markers are often used for tracking with computer vision technology in video see-through AR. Among these recent AR applications, video-based AR has attracted the most attention from researchers.

However, there has been little research conducted related to using AR to enhance e-commerce. In 2001, Azuma et al. reviewed new advances in AR which, after 1997, included display devices and methods, indoor and outdoor tracking, model rendering, and interaction technologies. At that time, they identified several problems that still needed to be addressed, such as occlusion, broader sensing, advanced rendering, and user perception issues. In addition, in 2005, Swan et al's survey showed that, although there were an increasing number of AR applications, research which considered usability was only a small part (less than 8%) of the total, and most of the usability studies were neither formal nor systematic.

Among the limited number of prior related studies, Zhu et al. (2006) proposed AR in-store shopping assistant devices, which provided personalized advertising and dynamic contextualization. Their study was aimed at using AR technology to enhance in-store shopping. Zhang et al. (2000) proposed and developed a prototype direct marketing system that used AR technology. Salespeople could use the system to show the main features of a product by manually holding a plate with specially designed markers. With their marker-based system, they could mix a 3D virtual product with a real scene, videotape the resulting scene, and then send the video tape to interested customers by email. However, their method of using AR in e-commerce did not make full use of the advantages of AR. With their method, online shoppers had no direct interaction with either physical objects or virtual product models. With only video recordings of AR scenes, customers still might not know whether products are suitable for them in their real physical environments. Two industry companies: metaio and bitmanagement (http://www.ar-live.de/main.php)(2007), are also trying to cooperate and extend e-commerce systems with AR technology. Users are asked to upload a photo of the personal environment with markers. The mixed scene could be visualized through their online tool. With their application, online users can visually see how the model fit their personal environment.  However static picture greatly limits uses' direct interaction with virtual product models in a natural way, and the flexible try in their environment.

In this study, a new AR e-commerce system was developed using video see-through AR considering that devices used by this type of AR are more available for online consumers, and that this type of AR is more flexible because the mixed AR scene could be displayed on different device in steady of optical see-through devices only. This system integrates a full-sized virtual product model into an online shopper's physical environment and provides the customer methods for "realistically" interacting with the virtual product. By this system, online shoppers can directly interact with the product model in their environment freely in a more nature way. For example, they can move around to see how the product fit their space from different viewpoint, and they can also move around markers as moving products. This paper presents both the design of the AR e-commerce assistant system and related usability studies. Several key issues related to using AR to enhance e-commerce are also discussed and analyzed.

## 3 System and User Interface Design

In this study, an AR e-commerce assistant system was designed to provide consumers with more realistic product experiences and interactions. With the developed AR e-commerce assistant, online consumers can bring a product into their physical environment and even try out and visualize the product in their physical environments while shopping from their computers.

### 3.1 Structure

Like traditional e-commerce systems, our AR e-commerce system uses the internet as the primary user interaction platform. However, with our AR e-commerce system, a camera is needed to capture the consumer's physical environment and then integrate it with virtual objects.

The system was developed as an Active X plug-in for an e-commerce web page. Online users can use the web page navigation to search for and view pictures and product related information, just as they would on a traditional e-commerce website. However, online shoppers can also use the plug-in to bring virtual products into their physical environment and then interact with the products to determine if the products are suitable.

The plug-in was made using the MFC and OpenGL libraries. The plug-in works between clients and an e-commerce assistant server through an Internet Explorer interface, so that online consumers can easily log onto the Internet, using different hardware, like a computer, cell phone, or Personal Digital Assistant (PDA) to access it as shown in Figure 2.1. In this system, an extra camera is needed, so that consumers can bring product models into their home, auto, outdoor, or other scenes.  ARToolkit (Kato and Billinghurst, 1999) was used for tracking, and Open VRML was used for rendering models. The complete structure of the system is shown in Figure 2.2.

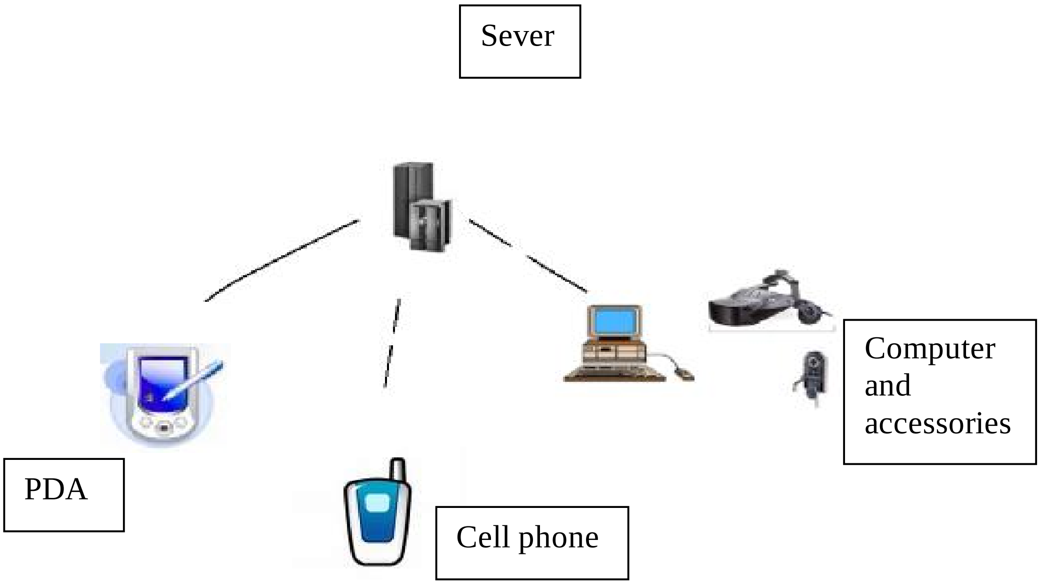


**Fig. 2.1** AR e-commerce assistant system working model

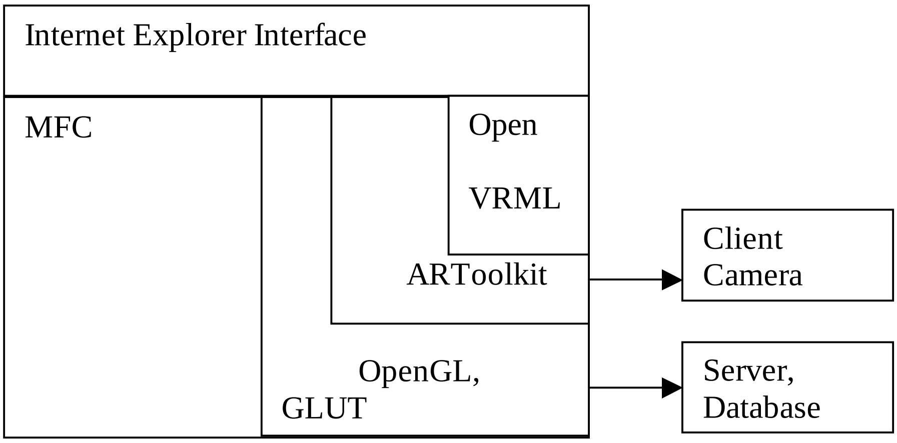


**Fig. 2.2** The structure of the AR e-commerce assistant system

### 3.2 Interfaces

Primary users of the system are expected to be common computer users, with minimal computer experience. As a result, the user interface of the system was made as simple and user-friendly as possible. In the study, we determined that consumer shopping typically includes three main tasks according to our analysis:

1. Searching for products.
2. Interacting with products.
3. Acquiring product information.

As a result, the user interface was designed to facilitate the three primary shopping tasks. The three tasks were combined into a two-level menu system within the AR window as shown in Figure 2.3 considering 2D menu system is still the most intuitive interaction way with computer for users because of their previous computer experience. Through the menu user could access the full interaction designed for AR e-commerce. Shortcut keys are also available to avoid the interruption between the user and the AR scene.

To provide convenient searching for products, a product search interface in AR window is provided as shown in Figure 2.4 so that user do not have to exit the AR application every time to find another product at web page level and reopen another AR application for comparison. Several capabilities were also developed to make product searching efficient, such as searching by keywords, sorting by properties, image viewing, listing operations, and price displays. With the tool, users can recursively search for and switch product models in an AR display, to compare products and thus gain enough direct information to make purchasing decisions. For tracking purposes, within the system, different markers correspond to types of products. Online shoppers can also combine different types of products together when shopping. For example, a shopper can combine a table with different chairs or sofas to check the appearance of different combinations in their home.

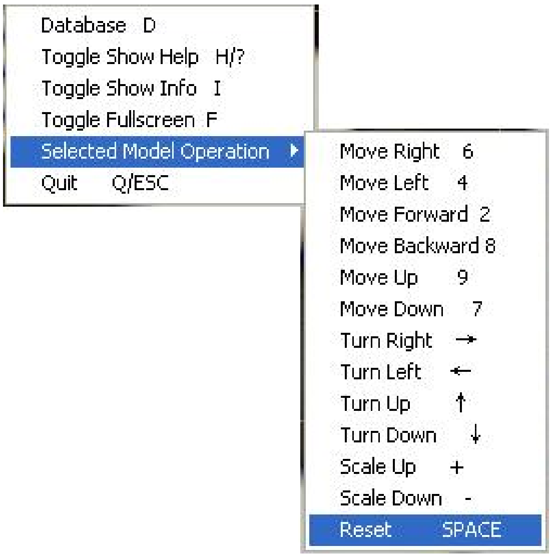
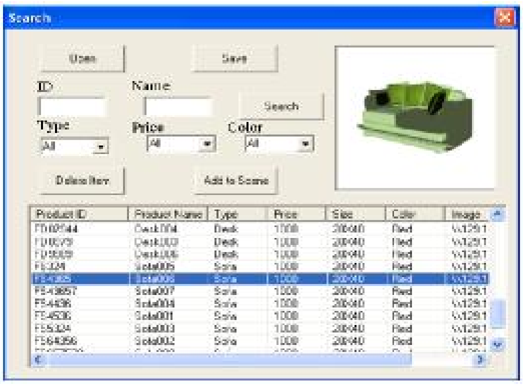


**Fig. 2.3** User interface menu system

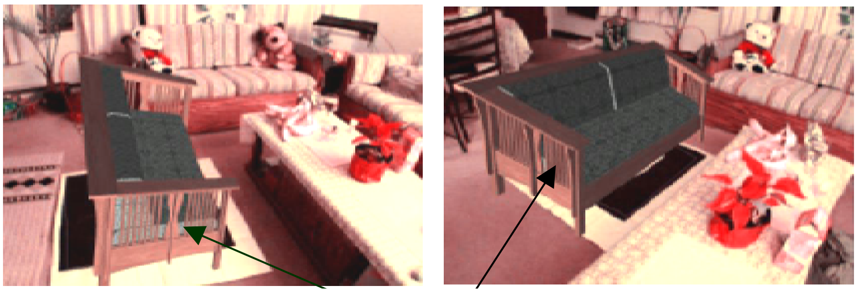


| Product ID | Product Name | Type | Price | Size | Color | Image |
|---|---|---|---|---|---|---|
| FD82544 | Desk004 | Desk | 1000 | 200x40 | Red | \\129.1 |
| FD10579 | Desk003 | Desk | 1000 | 200x40 | Red | \\129.1 |
| FD15889 | Desk006 | Desk | 1000 | 200x40 | Red | \\129.1 |
| FE3324 | Sofa005 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| [illegible] | Sofa006 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| FE43857 | Sofa007 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| FE4436 | Sofa004 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| FE4536 | Sofa001 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| FE55324 | Sofa003 | Sofa | 1000 | 200x40 | Red | \\129.1 |
| FE64356 | Sofa002 | Sofa | 1000 | 200x40 | Red | \\129.1 |

**Fig. 2.4** Product search interface

With the well-built and normalized product models selected and loaded to the AR scene, the products can be visualized with the actual size in the live background environment which is captured by the local camera. Users can also pick one of virtual products and manipulate it, for example, move or rotate the model, and view specific information about the selected product, such as name, price, size, and color, to help them make their decision.

In AR e-commerce, user could have special interactions, which is not available in other applications. User can walk around the environment with the laptop and camera to see how the product fit the environment from different viewpoint as shown in Figure 2.5. User can also interact with the AR scene by moving or rotating markers used for tracking.



Virtual sofa at different angles

**Fig. 2.5** A virtual model in a real scene

As mentioned above, ARToolkit library is used for marker-based tracking in real scenes (Kato and Billinghurst, 1999). Large markers are used for large virtual objects, such as furniture, as shown in Figure 2.6. Using large markers makes the recognition and registration easier and more reliable. With large markers, online consumers can bring virtual furniture or other large virtual products into their homes, and view it in a farther distance. Otherwise it will bring more instability since marker tracking is based on computer vision technology. Product model needed to be normalized according to the marker size so that user see the actual size to help with their decision.

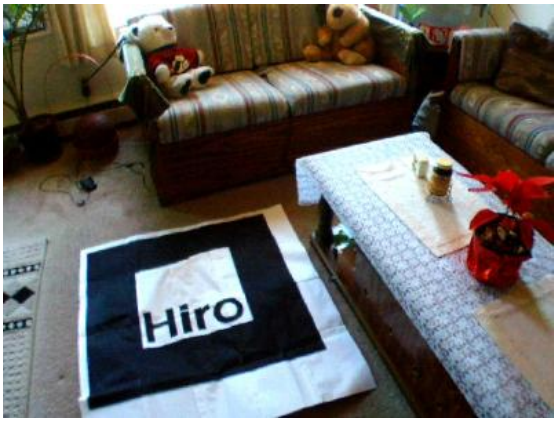


**Fig. 2.6** A Big marker was used

## 4 Usability Study

A usability study was conducted to compare the developed AR enhanced e-commerce system with a traditional e-commerce system and a VR-enhanced e-commerce system. To avoid web page design bias, all three web pages were designed using the same design template, which included a word description of the product and a visualization of the product, as shown in Figures 2.7-2.9. The word description parts of the three e-commerce web pages were the same. The only difference among the three types of e-commerce systems was in the visualization component.

For visualization, traditional e-commerce web pages typically use several static 2D pictures of a product, from different perspectives, as shown in Figure 2.7. With a traditional e-commerce web page, users can visually examine the static 2D product pictures before they buy the product. They can also usually interactively switch between the images. The traditional method is the most commonly used e-commerce approach generally used today.

VR-enhanced e-commerce web pages typically use JAVA applets for visualization. The JAVA applets dynamically download 3D product models in real-time and provide different manipulation capabilities (translate, rotate, zoom) to users, as shown in Figure 2.8. With VR-enhanced e-commerce web pages, users can easily control and select viewpoints for looking at virtual product models. There might be different designs of VR e-commerce. But this type of design is more representative since similar type of design has been taken for user studies of VR e-commerce (Daugherty 2005) and also for commercial use in like Compaq.com and Dell.com.

AR-enhanced e-commerce web pages use ActiveX controls for visualization as described above. System users can visually bring products into their actual physical environments, as shown in Figure 2.9. With the developed AR-enhanced system, users can hold a laptop, which has a camera, and move around their environment to see how a virtual product model looks corresponding to the translation, rotation, zoom interaction in VR e-commerce, and pick operation in traditional e-commerce, and then decide if they want to buy the product. They can also move markers to position the virtual products at different locations to help them make their buying decisions.  Figure 2.10 shows an example of our AR e-commerce system running on a laptop. To control different interaction bias with VR e-commerce and traditional e-commerce, the developed AR e-commerce menu system is not asked to use in the user study.

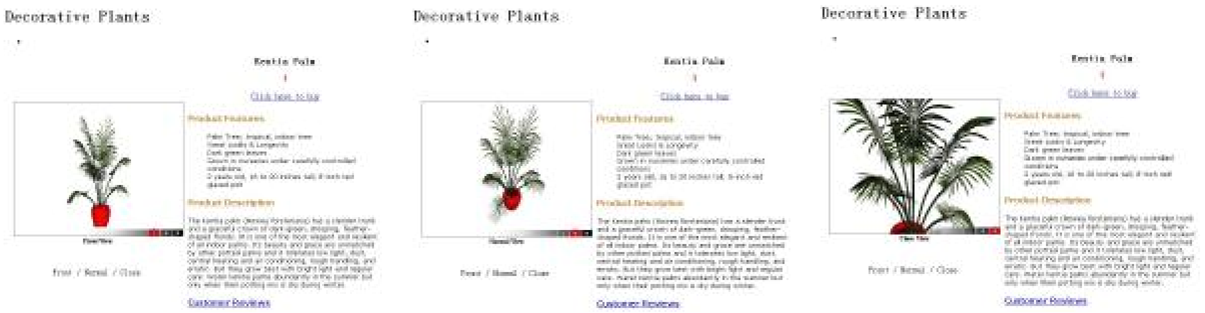
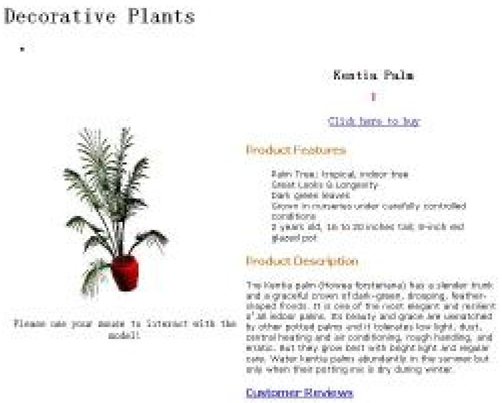


**Fig. 2.7** Traditional e-commerce with three static 2D images



Decorative Plants

Kentia Palm

$

Click here to buy

Product Features

Palm Tree; tropical, indoor tree
Great Looks & Longevity
Dark green leaves
Grown in nurseries under carefully controlled conditions
2 years old, 18 to 20 inches tall; 8-inch red glazed pot

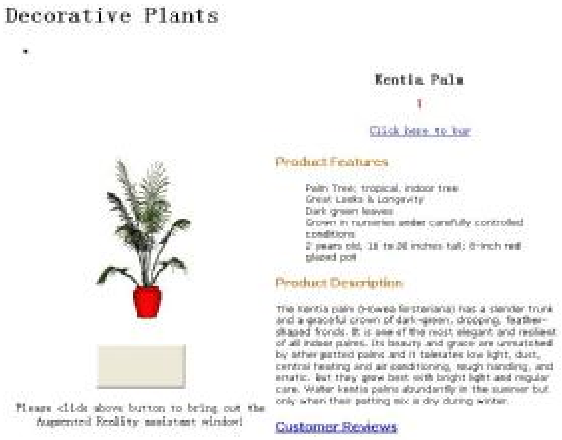
Product Description

The Kentia palm (Howea forsteriana) has a slender trunk and a graceful crown of dark-green, drooping, feather-shaped fronds. It is one of the most elegant and resilient of all indoor palms. Its beauty and grace are unmatched by other potted palms and it tolerates low light, dust, central heating and air conditioning, rough handling, and erratic. But they grow best with bright light and regular care. Water kentia palms abundantly in the summer but only when their potting mix is dry during winter.

Customer Reviews

Please use your mouse to interact with the model!

**Fig. 2.8** VR e-commerce with interactive 3D model



Decorative Plants

Kentia Palm

$

Click here to buy

Product Features

Palm Tree; tropical, indoor tree
Great Looks & Longevity
Dark green leaves
Grown in nurseries under carefully controlled conditions
2 years old, 18 to 20 inches tall; 8-inch red glazed pot

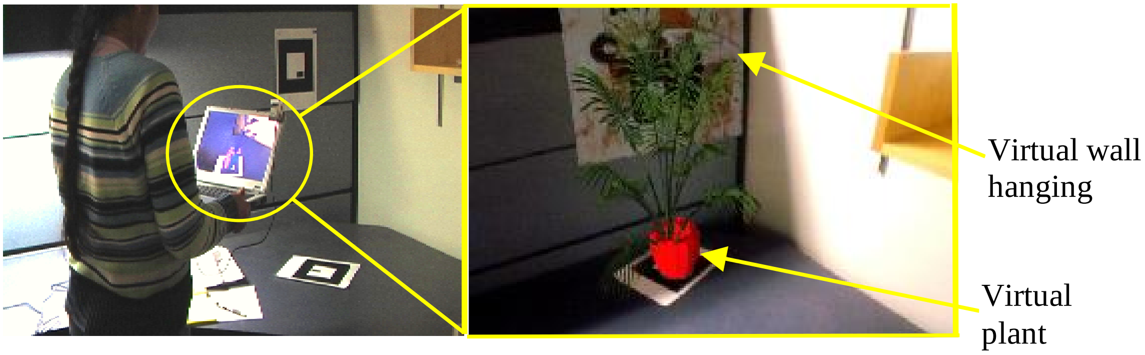
Product Description

The Kentia palm (Howea forsteriana) has a slender trunk and a graceful crown of dark-green, drooping, feather-shaped fronds. It is one of the most elegant and resilient of all indoor palms. Its beauty and grace are unmatched by other potted palms and it tolerates low light, dust, central heating and air conditioning, rough handling, and erratic. But they grow best with bright light and regular care. Water kentia palms abundantly in the summer but only when their potting mix is dry during winter.

Customer Reviews

Please click above button to bring out the Augmented Reality assistant window!

**Fig. 2.9** AR e-commerce interface



(a) AR application

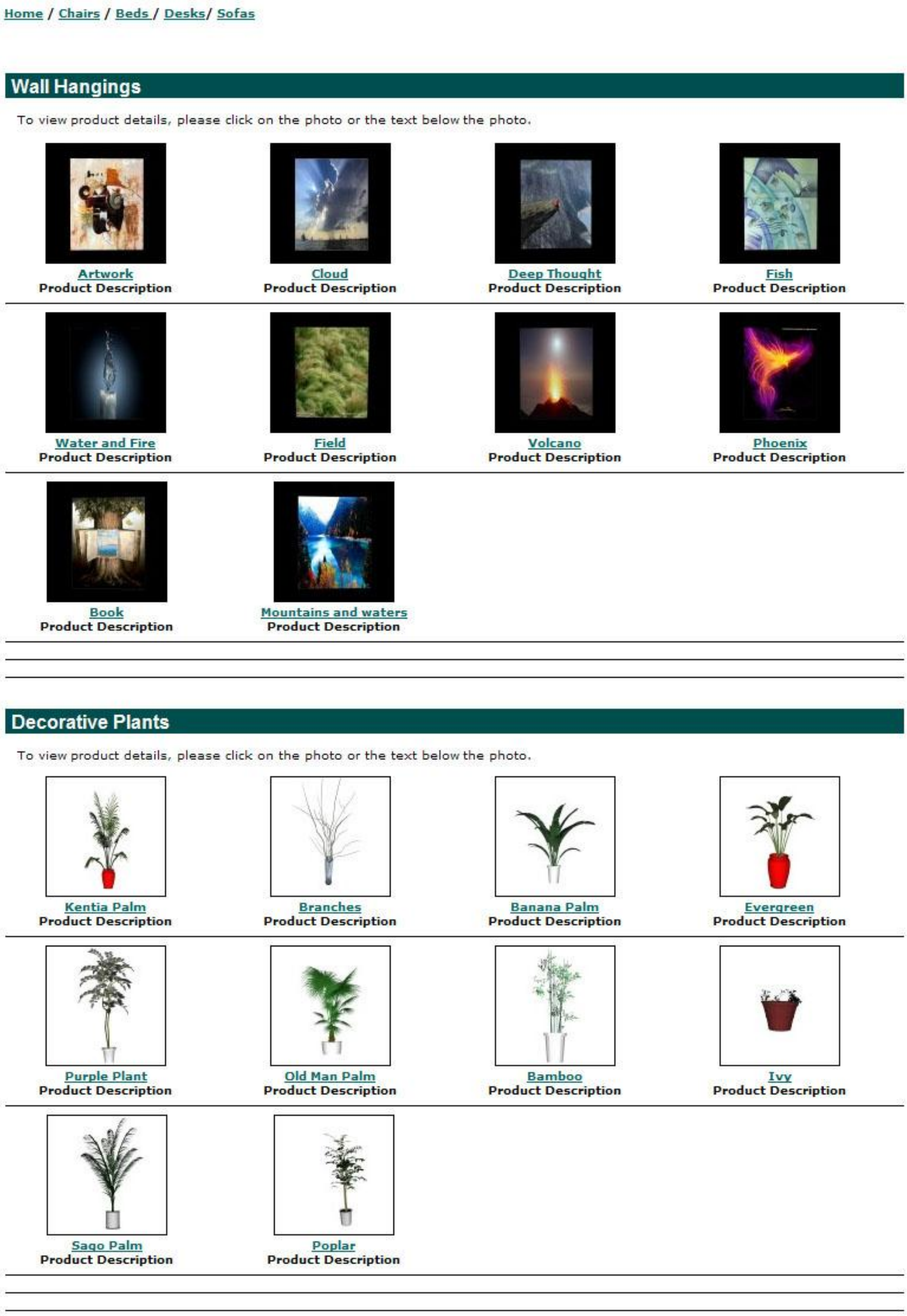
(b) AR scene on computer screen

**Fig. 2.10** AR application running on a laptop computer

## 4.1 Experiment Design

Based on a pilot user study for home furniture products (Lu and Smith, 2006), a formal user study was designed and conducted to test the usability of the developed AR e-commerce system. In the full study, different types of e-commerce web pages were designed for office products (wall hangings and decorative plants) to avoid product-based bias, as shown in Figure 2.11.

The experiment was designed as within-subjects for types of e-commerce, so that each subject would access all three e-commerce system. Because subjects inevitably differ from one another. In between-subject designs, these differences among subjects are uncontrolled and are treated as error. In within-subject designs, the same subjects are tested in each condition. Therefore, differences among subjects can be measured and separated from error (Howell 2007). Removing variance due to differences between subjects from the error variance greatly increases the power of significance tests. Therefore, within-subjects designs are almost always more powerful than between-subject designs. Since power is such an important consideration in the design of experiments, this study was designed as within-subjects experiment to compare user's subjective satisfaction level of using three different types e-commerce system, by which different participant's rating standard will not affect the comparison. Tests were carried out with six volunteer participants in each of the four office environments. In total, twenty-four participants were tested in the experiment. At the beginning of the experiment, participants were trained to use the three types of e-commerce systems. During the experiment, real-time help concerning how to use the systems was also provided. In the test, participants were asked to use the three types of e-commerce system to buy different office products for the different environments, without considering the budget. Users were asked to select wall hangings and decorative plants and then compare the three types of e-commerce systems. During the experiment, the process was recorded and observed. After the experiment, participants were asked to fill out a questionnaire and to give their evaluations of usability. Four main variables (overall evaluation, information provided, ease of use, and confidence level in the final decision) were measured for each type of e-commerce system for each participant. In the study, the independent variables were the three different types of e-commerce systems, four different environments (an open space office, a cubical, a single-user single-room office, and a multi-user single-room shared office). Within each environment, presentation of the e-commerce systems was systematically varied to control the "carryover" effects of within-subjects design. Since we assigned 6 subjects to each environment, we were able to test all possible presentation orders of the three e-commerce systems (3 choose 1 * 2 choose 1 * 1 choose 1) = 6 different testing orders: (T, VR, AR), (T, AR, VR), (VR, T, AR), (VR, AR, T), (AR, T, VR), and (AR, VR, T). The dependent variables in the research question were four main variables: overall evaluation, information provided, ease of use, and confidence level in the final decision.



Home / Chairs / Beds / Desks/ Sofas

Wall Hangings

To view product details, please click on the photo or the text below the photo.

Artwork
Product Description

Cloud
Product Description

Deep Thought
Product Description

Fish
Product Description

Water and Fire
Product Description

Field
Product Description

Volcano
Product Description

Phoenix
Product Description

Book
Product Description

Mountains and waters
Product Description

Decorative Plants

To view product details, please click on the photo or the text below the photo.

Kentia Palm
Product Description

Branches
Product Description

Banana Palm
Product Description

Evergreen
Product Description

Purple Plant
Product Description

Old Man Palm
Product Description

Bamboo
Product Description

Ivy
Product Description

Sago Palm
Product Description

Poplar
Product Description

**Fig. 2.11** Office products

To test whether the usability results were affected by experience order, the six user study participants in each of the four environments were randomly assigned to one of the six orders. Evaluations of the four main variables were also compared for the different orders. The formal study addressed the following hypotheses:

*Hypothesis 1*: The overall evaluation and satisfaction level of using AR e-commerce system is higher than using the other two e-commerce systems.

*Hypothesis 2*: The AR e-commerce system provides more visualized information to online shoppers than the other two e-commerce systems.

*Hypothesis 3*: The ease of use rating for the AR e-commerce system is lower than the other two e-commerce systems.

*Hypothesis 4*: Users of AR e-commerce system have a higher confidence level in their final decision than users of the other two e-commerce systems.

*Hypothesis 5*: User performance in the different e-commerce systems will not be affected by locations.

To test the 5 hypotheses, different ratings given by the participants, after using the three types of e-commerce systems, were compared.

## 4.2 Experiment Participants

All participants for the study were individuals from Iowa State University who responded to an invitation email. They represented students, staff, and faculty. Figure 2.12 shows the composition of subjects for the study.

Figure 2.12 shows that the gender of participants was equally distributed. Since most of the participants were students, the age distribution of participants was skewed toward lower age groups, and computer experience level was skewed toward high levels (“A little” mean little computer experience while “Pro” means professional computer experience), which might have caused some sample bias.

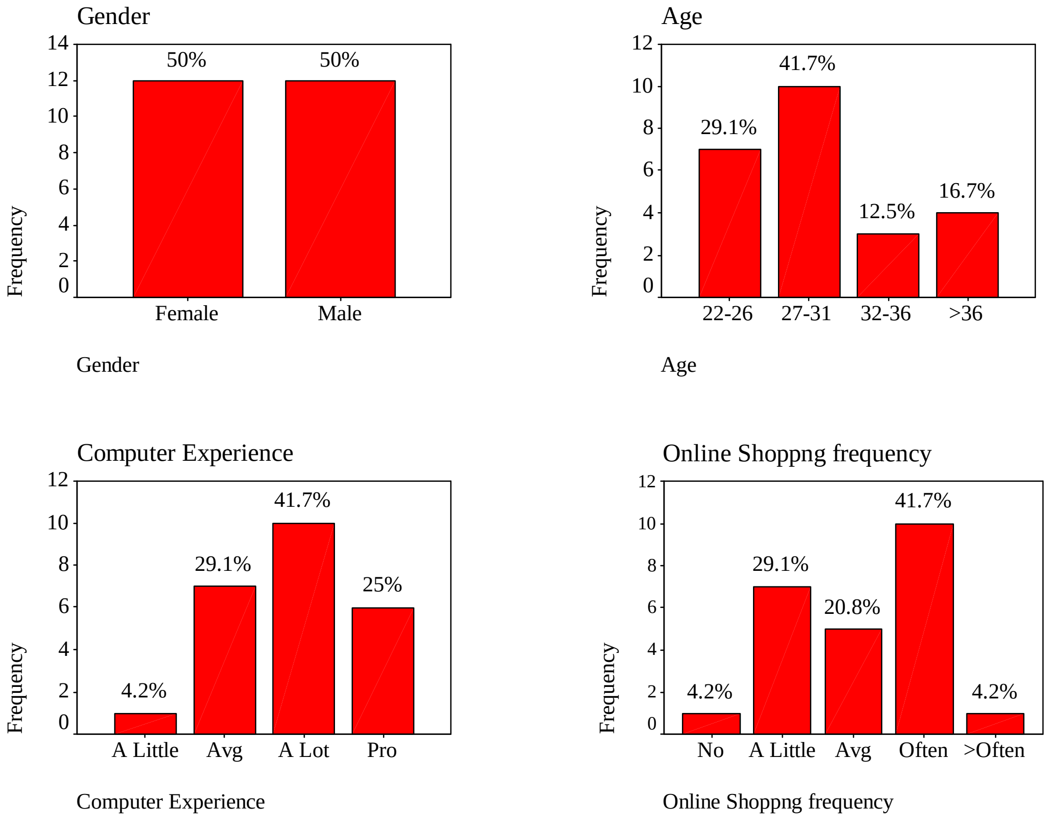


**Fig. 2.12** Participants' self description

## 4.3 Results

### 4.3.1 Overall Evaluation

The first research question in the questionnaire was designed to capture overall feelings about the three different types of e-commerce systems, without being affected or guided by later questions. The participants' overall evaluations are listed in Table 2.1, by locations and by experience orders, which were also separately tested using Factorial ANOVA.

**Table 2.1** Overall evaluation (1=lowest 5=highest)

| LOCATION | SUBJECTS | PARTICIPANT RATING | | |
|---|---|---|---|---|
| | | T | VR | AR |
| Open space office (1) | 1 | 2 | 4 | 5 |
| | 2 | 1 | 5 | 5 |
| | 3 | 1 | 3 | 4 |
| | 4 | 2 | 3 | 5 |
| | 5 | 1 | 3 | 4 |

| | | | | |
|---|---|---|---|---|
| | 6 | 2 | 4 | 5 |
| | Mean/Std. Dev | 1.5/0.548 | 3.667/0.816 | 4.667/0.516 |
| Cubical office (2) | 7 | 3 | 5 | 4 |
| | 8 | 2 | 5 | 4 |
| | 9 | 2 | 3 | 4 |
| | 10 | 1 | 5 | 5 |
| | 11 | 3 | 4 | 5 |
| | 12 | 1 | 3 | 5 |
| | Mean/Std. Dev | 2/0.894 | 4.167/0.983 | 4.5/0.548 |
| Single-user single-room office (3) | 13 | 3 | 5 | 5 |
| | 14 | 1 | 3 | 5 |
| | 15 | 3 | 3 | 5 |
| | 16 | 5 | 4 | 4 |
| | 17 | 1 | 3 | 5 |
| | 18 | 1 | 3 | 5 |
| | Mean/Std. Dev | 2.333/1.633 | 3.5/0.837 | 4.833/0.408 |
| Multi-user single-room shared office (4) | 19 | 3 | 4 | 5 |
| | 20 | 3 | 4 | 5 |
| | 21 | 2 | 3 | 4 |
| | 22 | 1 | 2 | 4 |
| | 23 | 3 | 4 | 4 |
| | 24 | 5 | 5 | 4 |
| | Mean/Std. Dev | 2.833/1.329 | 3.667/1.033 | 4.333/0.516 |
| Mean | | 2.167 | 3.75 | 4.583 |
| Std. Dev. | | 1.204 | 0.897 | 0.504 |

As shown in Table 2.1, the mean overall evaluation for traditional e-commerce was 2.167, the mean overall evaluation for VR enhanced e-commerce was 3.75, and the mean overall evaluation for AR enhanced e-commerce was 4.583. As shown in the between-subjects effects and within-Subjects effects analysis of Table 2.2, the p-value for the effect of the type of e-commerce system is very small (<0.05), which indicates that there is a statistically significant difference in mean overall evaluations between the three types of e-commerce systems. In contrast, the p-values for the effect of location is 0.7913, which indicates that there is no statistically significant difference in mean overall evaluations for different locations.

Figures 2.13, clearly shows that the main effect for different types of e-commerce systems is obvious and that the overall evaluation for the AR e-commerce system is higher than the ratings for the traditional and VR e-commerce systems. The p-value for interaction between types and locations is 0.1407, which indicate that there are no statistically significant interaction effects for types and locations. Thus, interaction effects, and location effects were neglected in the refined analysis model shown in Table 2.3.

**Table 2.2** Tests of Between-Subjects Effects and Within-Subjects Effects (Dependent Variable: Overall evaluation)

| SOURCE | DF | SUM OF SQUARES | MEAN SQUARE | F | SIG. |
|---|---|---|---|---|---|
| Location | 3 | 1.2222 | .4074 | .3476 | .7913 |
| Error | 20 | 23.4444 | 1.1722 | | |
| | | | | | |
| Type | 2 | 72.3333 | 36.1667 | 55.1695** | .000** |
| Location*Type | 6 | 6.7778 | 1.1296 | 1.7232 | .1407 |
| Error | 40 | 26.2222 | .6556 | | |

**p<0.05

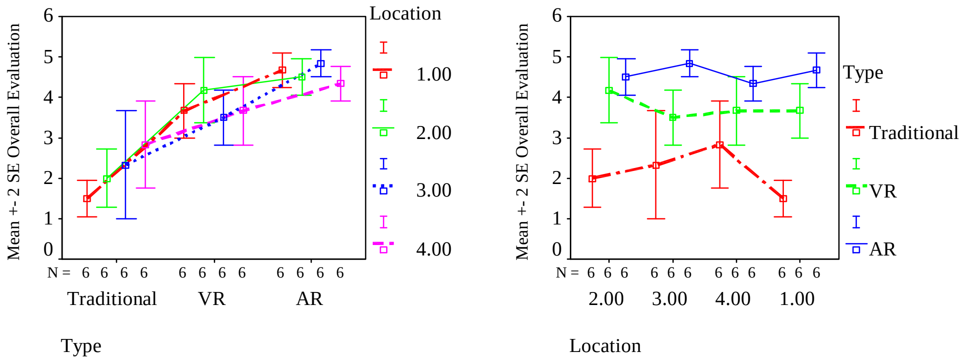


**Fig. 2.13** Interaction between type and location for Overall evaluation

**Table 2.3** Homogeneous Subsets Tukey HSD (Dependent Variable: Overall evaluation)

| | N | Subset | | |
|---|---|---|---|---|
| **Type** | | 1 | 2 | 3 |
| **Traditional** | 24 | 2.1667 | | |
| **VR** | 24 | | 3.7500 | |
| **AR** | 24 | | | 4.5833 |
| **SIG.** | | 1.000 | 1.000 | 1.000 |

To determine differences in overall evaluations for the three types of e-commerce systems, multiple mean comparisons (Tukey HSD) was used, without considering. The analysis results in Table 2.3 show that each pair of mean overall evaluations for the three types is significantly different.

In comparing the three e-commerce systems, the AR enhanced e-commerce was rated highest by users, which indicates that users preferred the AR enhanced e-commerce system more than the other two for office decoration. So research hypothesis 1 is accepted. Based on the strength and weakness of AR e-commerce comparing to the other two type of e-commerce, costumers still prefer AR e-commerce. “It is a very potential method, especially for products like furniture,” as mentioned by one of the participant. There was also no significant evidence that location had any effect on users’ overall evaluations from the statistics.

### 4.3.2 Visualized Information Provided

In the questionnaire, users were asked to rate how much information they gained from the three different types of e-commerce systems. Participants’ ratings for information provided are listed in Table 2.4, by locations and by experience orders, which were also tested separately using Factorial ANOVA.

**Table 2.4** Information provided (1=lowest 5=highest)

| LOCATION | SUBJECTS | PARTICIPANT RATING | | |
|---|---|---|---|---|
| | | T | VR | AR |
| Open space office (1) | 1 | 3 | 3 | 3 |
| | 2 | 1 | 3 | 5 |
| | 3 | 1 | 3 | 5 |
| | 4 | 3 | 4 | 5 |
| | 5 | 1 | 4 | 4.5 |
| | 6 | 3 | 4 | 5 |
| | Mean/Std. Dev | 2/1.095 | 3.5/0.548 | 4.583/0.801 |
| Cubical office (2) | 7 | 3 | 4 | 4.5 |
| | 8 | 2 | 4 | 4 |
| | 9 | 2 | 2 | 5 |

| | | | | |
|---|---|---|---|---|
| | 10 | 1 | 3 | 4 |
| | 11 | 3 | 5 | 4 |
| | 12 | 1 | 2 | 5 |
| | Mean/Std. Dev | 2/0.894 | 3.333/1.211 | 4.417/0.492 |
| Single-user single-room office (3) | 13 | 3 | 5 | 5 |
| | 14 | 1 | 3 | 5 |
| | 15 | 1 | 4 | 4 |
| | 16 | 4 | 5 | 4 |
| | 17 | 1 | 3 | 5 |
| | 18 | 1 | 3 | 5 |
| | Mean/Std. Dev | 1.833/1.329 | 3.833/0.983 | 4.667/0.516 |
| Multi-user single-room shared office (4) | 19 | 3 | 4 | 5 |
| | 20 | 2 | 4 | 5 |
| | 21 | 1 | 3 | 4 |
| | 22 | 1 | 3 | 4 |
| | 23 | 2 | 3 | 4 |
| | 24 | 3 | 4 | 5 |
| | Mean/Std. Dev | 2/0.894 | 3.5/0.548 | 4.5/0.548 |
| Mean | | 1.958 | 3.542 | 4.542 |
| Std. Dev. | | 1.000 | 0.833 | 0.569 |

From Table 2.4, the mean rating for information provided by the traditional e-commerce system was 1.958, the mean information provided by the VR-enhanced e-commerce system was 3.542, and the mean rating for information provided by the AR-enhanced e-commerce system was 4.542. As shown in the between-subjects effects and within-Subjects effects analysis of Table 2.5, the p-value for the effect of type of e-commerce system is very small (<0.05), which indicates that there is a statistically significant difference in mean information provided between the three types of e-commerce. However, the p-value for the effect of location is 0.9555, which indicates that there is no statistically significant difference in mean information provided for different locations and different experience orders.

Figures 2.14, clearly shows that the information users gained from the AR e-commerce system was more than the information they gained from the traditional and VR e-commerce systems. The p-value for the interaction between type and location is 0.9677, which indicates that there was no statistically significant interaction effect between type and location. Thus, the location effect, and interaction effects on information provided were neglected in the refined analysis model shown in Table 2.6.

**Table 2.5** Tests of Between-Subjects Effects and Within-Subjects Effects (Dependent Variable: Information Provided)

| SOURCE | DF | SUM OF SQUARES | MEAN SQUARE | F | SIG. |
|---|---|---|---|---|---|
| Location | 3 | .3472 | .1157 | .1062 | .9555 |
| Error | 20 | 21.8056 | 1.0903 | | |
| | | | | | |
| Type | 2 | 81.4444 | 40.7222 | 69.4787** | .000** |
| Location*Type | 6 | .7778 | .1296 | .2212 | .9677 |
| Error | 40 | 23.4444 | .5861 | | |

**p<0.05

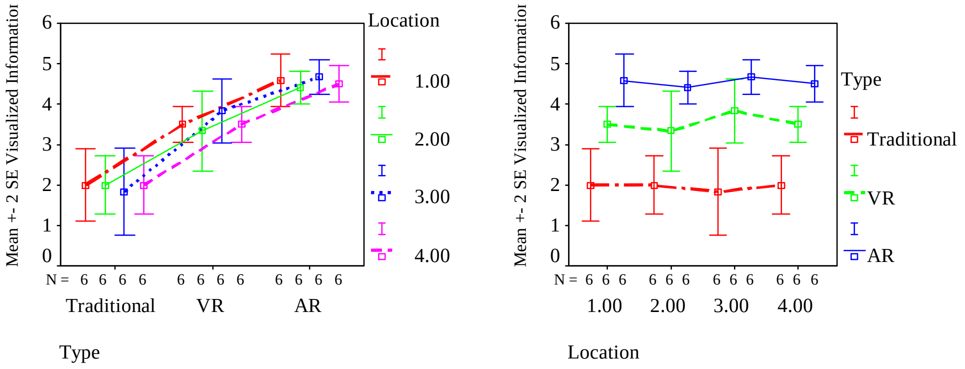


**Fig. 2.14** Interaction between type and location for Information Provided

**Table 2.6** Homogeneous Subsets Tukey HSD (Dependent Variable: Information Provided)

| | N | Subset | | |
|---|---|---|---|---|
| **Type** | | 1 | 2 | 3 |
| **Traditional** | 24 | 1.9583 | | |
| **VR** | 24 | | 3.5417 | |
| **AR** | 24 | | | 4.5417 |
| **SIG.** | | 1.000 | 1.000 | 1.000 |

To determine the differences between the information users gained for the three types of e-commerce system, Tukey HSD was used, without considering location or order. With an experiment-wise error rate of 0.05, Table 2.6 shows that the differences in information provided between the AR e-commerce system and both the traditional e-commerce and VR enhanced e-commerce system are both statistically significant. So the research hypothesis 2 is accepted. As participants mentioned in their feedbacks that the AR e-commerce system provides the capability to see how products fit in the physical space, so that they can gain more visualized information. “It is very vivid, as if you put a real product into the place where you want. You can efficiently evaluate product information, such as color and size, and determine whether it can match with the scene very well.” “It can provide people an interesting experience and help people gain more information and a much more correct judgment.” Besides, there was also no significant evidence showing that location had an effect on information provided from the statistics.

### 4.3.3 Ease of Use

Participants’ ratings concerning ease of use for the three different types of e-commerce systems are listed in Table 2.7, by location and by experience order, which were also tested separately using Factorial ANOVA.

**Table 2.7** Ease of use (1=lowest 5=highest)

| LOCATION | SUBJECTS | PARTICIPANT RATING | | |
|---|---|---|---|---|
| | | T | VR | AR |
| Open space office (1) | 1 | 5 | 4 | 2 |
| | 2 | 5 | 1 | 5 |
| | 3 | 5 | 4 | 3 |
| | 4 | 2 | 3 | 5 |
| | 5 | 5 | 4.5 | 4.5 |
| | 6 | 2 | 5 | 4 |
| | Mean/Std. Dev | 4/1.549 | 3.583/1.429 | 3.917/1.201 |
| Cubical office (2) | 7 | 5 | 5 | 3 |
| | 8 | 4 | 4 | 4 |
| | 9 | 5 | 3 | 2 |
| | 10 | 5 | 5 | 4 |
| | 11 | 4 | 5 | 3 |
| | 12 | 5 | 4 | 3 |

| | | | | |
|---|---|---|---|---|
| | Mean/Std. Dev | 4.667/0.516 | 4.333/0.816 | 3.167/0.753 |
| Single-user single-room office (3) | 13 | 5 | 4 | 4 |
| | 14 | 5 | 5 | 5 |
| | 15 | 5 | 4 | 5 |
| | 16 | 5 | 4 | 3 |
| | 17 | 5 | 4 | 3 |
| | 18 | 5 | 4 | 2 |
| | Mean/Std. Dev | 5/0 | 4.167/0.408 | 3.667/1.211 |
| Multi-user single-room shared office (4) | 19 | 5 | 5 | 3 |
| | 20 | 5 | 3 | 5 |
| | 21 | 4 | 4 | 3 |
| | 22 | 5 | 3 | 2 |
| | 23 | 4 | 4 | 3 |
| | 24 | 5 | 5 | 3 |
| | Mean/Std. Dev | 4.667/0.516 | 4/0.894 | 3.167/0.983 |
| Mean | | 4.583 | 4.021 | 3.479 |
| Std. Dev. | | 0.881 | 0.938 | 1.037 |

The mean ease of use for the traditional e-commerce system was 4.583, the mean ease of use for the VR enhanced e-commerce system  was 4.021, and the mean ease of use for the AR enhanced e-commerce system was 3.479.  As shown in the between-subjects effects and within-Subjects effects analysis of Table 2.8, the p-value for the effect of type of e-commerce system is 0.0027 (<0.05), which indicates that there is a statistically significant difference in mean ease of use between the three types of e-commerce systems. In contrast, the p-value for the effect of location is 0.4033, which indicates that there is no statistically significant difference in mean ease of use for different locations.

Figures 2.15 shows the main effect of different types of e-commerce systems. Ease of use for the AR e-commerce system is much lower than ease of use for the traditional and for the VR e-commerce systems. The p-value for the interaction effect between type and location is 0.5186, which indicate that there are also no statistically significant interaction effects for type and location or type. Thus, the interaction effects, for ease of use were neglected in the refined analysis model shown in Table 2.9.

**Table 2.8** Tests of Between-Subjects Effects and Within-Subjects Effects (Dependent Variable: Easiness to Use)

| SOURCE | DF | SUM OF SQUARES | MEAN SQUARE | F | SIG. |
|---|---|---|---|---|---|
| Location | 3 | 1.9444 | .6481 | 1.0234 | .4033 |
| Error | 20 | 12.6777 | .6333 | | |
| | | | | | |
| Type | 2 | 14.6319 | 7.3160 | 6.8721** | .0027** |
| Location*Type | 6 | 5.6181 | .9363 | .8795 | .5186 |
| Error | 40 | 42.5833 | 1.0646 | | |

**p<0.05

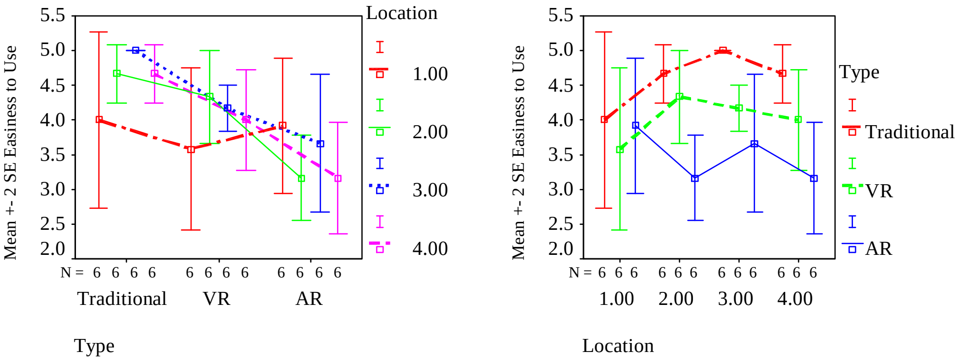


**Fig. 2.15** Interaction between type and location for Easiness to use

**Table 2.9** Homogeneous Subsets Tukey HSD

| | N | Subset | |
|---|---|---|---|
| **Type** | | 1 | 2 |
| **AR** | 24 | 3.4792 | |
| **VR** | 24 | 4.0208 | 4.0208 |
| **Traditional** | 24 | | 4.5833 |
| **SIG.** | | .128 | .110 |

To determine the differences between ease of use for the three types of e-commerce systems, Tukey HSD was used, without considering location or order. With an experiment-wise error rate of 0.05, Table 2.9 shows that the difference in ease of use between the traditional e-commerce system and the VR enhanced e-commerce system is not statistically significant. The difference between the VR enhanced e-commerce system and the AR enhance e-commerce system is also not statistically significant. However, ease of use for the traditional e-commerce system is significantly better than ease of use for the AR enhanced e-commerce system.

So the research hypothesis that ease of use for the AR e-commerce system is lower than for the traditional e-commerce systems is accepted. Participants mentioned in their feedback that the AR e-commerce system needs more high-end hardware equipment, and that it is inconvenient to use. “It is not very convenient to hold the laptop with your hands all the time.” There are two explanations about this, the first one is that AR e-commerce use more devices and need more computer skills, the second one is that users are still not familiar with AR and its interaction. Meanwhile, there is also no significant evidence that location has significant effects on ease of use.

### 4.3.4 User Confidence Level for Decision

The final main dependent variable measured in the questionnaire was the user’s confidence level in their decision (buy or not buy). Participants’ ratings are listed in Table 2.10, by location and by experience order, which were also tested using Factorial ANOVA.

**Table 2.10** User **c**onfidence level for decision (1=lowest 5=highest)

| LOCATION | SUBJECTS | PARTICIPANT RATING | | |
|---|---|---|---|---|
| | | T | VR | AR |
| Open space office (1) | 1 | 2 | 2 | 4 |
| | 2 | 1 | 3 | 5 |
| | 3 | 1 | 3 | 4 |
| | 4 | 2 | 4 | 5 |
| | 5 | 1 | 4 | 4.5 |
| | 6 | 2 | 3 | 5 |
| | Mean/Std. Dev | 1.5/0.548 | 3.167/0.752 | 4.583/0.491 |
| Cubical office (2) | 7 | 2 | 5 | 4 |
| | 8 | 2 | 4 | 5 |
| | 9 | 3 | 4 | 5 |
| | 10 | 2 | 4 | 3 |
| | 11 | 3 | 4 | 5 |
| | 12 | 1 | 2 | 5 |

| | | | | |
|---|---|---|---|---|
| | Mean/Std. Dev | 2.167/0.752 | 3.833/0.983 | 4.5/0.837 |
| Single-user single-room office (3) | 13 | 3 | 4 | 5 |
| | 14 | 3 | 5 | 5 |
| | 15 | 4 | 3 | 5 |
| | 16 | 4 | 4 | 3 |
| | 17 | 1 | 3 | 5 |
| | 18 | 2 | 3 | 5 |
| | Mean/Std. Dev | 2.833/1.169 | 3.667/0.816 | 4.667/0.816 |
| Multi-user single-room shared office (4) | 19 | 3 | 4 | 5 |
| | 20 | 2 | 3 | 5 |
| | 21 | 3 | 4 | 5 |
| | 22 | 2 | 4 | 5 |
| | 23 | 2 | 3 | 4 |
| | 24 | 3 | 3 | 5 |
| | Mean/Std. Dev | 2.5/0.548 | 3.5/0.548 | 4.833/0.408 |
| Mean | | 2.25 | 3.542 | 4.646 |
| Std. Dev. | | 0.897 | 0.779 | 0.634 |

**Table 2.11** Tests of Between-Subjects Effects and Within-Subjects Effects (Dependent Variable: User Confidence Level for Decision)

| SOURCE | DF | SUM OF SQUARES | MEAN SQUARE | F | SIG. |
|---|---|---|---|---|---|
| Location | 3 | 4.2049 | 1.4016 | 2.2107 | .1184 |
| Error | 20 | 12.6806 | .6340 | | |
| | | | | | |
| Type | 2 | 69.0208 | 34.5104 | 64.6229** | .0000** |
| Location*Type | 6 | 3.4514 | .5752 | 1.0772 | .3923 |
| Error | 40 | 21.3611 | .5340 | | |

**p<0.05

The mean user confidence level for the Traditional e-commerce system was 2.25, the mean user confidence level for the VR enhanced e-commerce system was 3.542, and the mean user confidence for the AR enhanced e-commerce system was 4.646. As shown in the between-subjects effects and within-Subjects effects analysis of Table 2.11, the p-value for the effect of type of e-commerce system is very small (<0.05), which indicates that there is a statistically significant difference in user confidence level between the three types of e-commerce systems. However, the p-value of the effect of location is 0.1184, which indicates that there is no statistically significant difference in user confidence level for different locations.

Figures 2.16 clearly shows the main effect for different types. User confidence level for the AR e-commerce is much higher than user confidence level for either the traditional or the VR e-commerce systems. The p-value for the interaction effect of type and location is 0.3923, which indicate that there is no statistically significant interaction effect for type and location, Thus, location effect, and interaction effects on user confidence level were neglected in the refined analysis model as shown in Table 2.12.

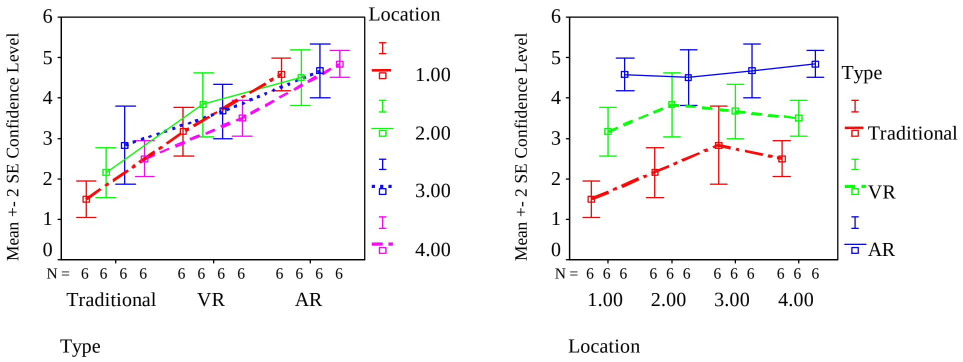


**Fig. 2.16** Interaction between type and location for Confidence Level for Decision

**Table 2.12** Homogeneous Subsets Tukey HSD (Dependent Variable: User Confidence Level in Decision)

| | N | Subset | | |
|---|---|---|---|---|
| **Type** | | 1 | 2 | 3 |
| **Traditional** | 24 | 2.2500 | | |
| **VR** | 24 | | 3.5417 | |
| **AR** | 24 | | | 4.6458 |
| **SIG.** | | 1.000 | 1.000 | 1.000 |

To determine the differences in user confidence level for the three types of e-commerce systems, Tukey HSD was used, without considering location or order. With an experiment-wise error rate of 0.05, Table 2.12 shows that the difference in user confidence level between the AR e-commerce system and both the traditional e-commerce system and the VR enhanced e-commerce system was statistically significant.

The results show that users had a higher confidence level in their shopping decisions when using the AR enhanced e-commerce system, rather than the other two e-commerce systems, for purchasing office decoration products. The research hypothesis 4 is accepted. AR e-commerce makes shopping more “visually intuitive”. “The user naturally sees what will happen before actually buying”. “It gives you a real-time experience in your own environment so that you can instantly tell whether or not the product is a good fit”. Meanwhile there was also no significant evidence that either location had an effect on user confidence level.

### 4.4 Observations and Users’ Comments

#### 4.4.1 “As Is” View

95.8% of participants mentioned in their feedbacks that the AR e-commerce system provides the capability to see how products fit in the physical space. Users’ comments included: It is “visually intuitive”. “The user naturally sees what will happen before actually buying”. “It gives you a real-time experience in your own environment so that you can instantly tell whether or not the product is a good fit”. “It presents products in a real scale relative to the environment, and is able to show views from several perspectives”. “AR makes shopping more confident”. “It is cool and helpful for making the decision”. “It is very vivid, as if you put a real product into the place where you want. You can efficiently evaluate product information, such as color and size, and determine whether it can match with the scene very well.” “It can provide people an interesting experience and help people gain more information and a much more correct judgment.”

#### 4.4.2 Ease of Use

87.5% of participants mentioned in their feedback that the AR e-commerce system needs more high-end hardware equipment, and that it is inconvenient to use. Users' comments included: "You have to have a laptop or mobile device." "It is not very convenient to hold the laptop with your hands all the time." "It is constrained to a marker." "It is limited to certain viewing areas". "If the designer could use a small device (like a cell phone) to replace the laptop, it would be more convenient for customers." "It is slower for the user and more complicated." "If it was more user friendly and more easy to use, it would be widely used." "Not as convenient as VR and traditional e-commerce."

However, 12.5% of participants believed that the AR e-commerce system was convenient to use. Users' comments included: "It is very easy". "There is not much I have to learn to dive right in." "It is friendly and looks real." "It is easy to manipulate. It is a more natural interactive method than mouse interaction". "It is more convenient, and otherwise, it is difficult to shop at onsite stores that are far away."

### 4.4.3 Unstable

29.2% of participants mentioned in their feedback that the AR e-commerce system is unstable. "The images on the screen are not stable, and sometimes disappear due to problems with light intensity." "If people could easily change the position of the target, without considering light problems, it would be better." "The smoothness of motion tracking needs to be improved." "There are limited spots where you can see the product." "Sometimes I cannot see the virtual image."

### 4.4.4 Real Modeling and Rendering

25% of participants said that the virtual objects in the AR e-commerce display were not very real. "If it looked more realistic, it would be better." "If the models looked the same as the real objects, it would be better." "The model should be designed more accurately." "It needs some easy way to directly transfer real things into 3D virtual models." "It needs accurate illumination." "It would be great if I could feel the texture of a product".

### 4.4.5 Internet Speed

25% of participants felt that the Internet wireless connection speed was not fast enough for AR e-commerce. They considered the process of downloading models slow. However, they believed this problem would be solved with further development of technology. One user said: “While I thought that the quality of the graphics of the product would be an issue, I found that the AR system provided me with an excellent sensation of the product. The lack of a very high graphical representation of the product did not bother me at all.”

## 5 Discussion and Conclusions

Traditional e-commerce systems have reached a limitation that needs to be overcome, because they do not provide enough direct information for online shoppers, especially when they are shopping for products like furniture, clothing, shoes, jewelry, and other decorative products. In this paper, we developed an AR e-commerce system and studied the effectiveness of AR for enhancing e-commerce.

A formal usability study was designed and conducted. Usability experiments results verified that the developed AR e-commerce system could be used to provide more direct product information to online shoppers and thereby help them make better purchasing decisions. Additionally, in the study, users preferred the AR e-commerce system more than traditional e-commerce and VR e-commerce systems.

Although the AR e-commerce system provides more information and interaction capability than the other e-commerce systems, it is also evident that some limitations still exist in the proposed approach. According to the study participants, the major limitation of using the AR e-commerce system is that it is currently not as easy to use as the traditional or VR e-commerce systems. The AR e-commerce system’s interaction method still needs to be improved, to make it more convenient for users. For example, online shopper could have multiple choices to use this system like upload static picture with markers, or videos pre-made so that users do not have to take the computer around for each product every time. The application could also be specifically implemented on PDA and cell phone, which is available for most consumers and is also light to carry. The rendering methods used also need to be improved to help integrate virtual models into real scenes more seamlessly. For example, large amount of texture mapping should be used to improve the realness of the virtual product. Real time occlusion could also be implemented to help with consumers' depth perception about the virtual product. The computer vision algorithm used in the AR system needs to be improved, to make the marker tracking more stable, even in a poor lighting condition. New smarter algorithm should also be studied and developed for partial marker tracking so that user do not need to worry about that the virtual product will disappear because the marker is partially occluded. However, even though wireless Internet access is currently still not fast enough to transfer high-resolution product models in real-time, it might not be a problem in the near future.

## References

ARToolkit document, Retrieved March 31, 2007 from http://www.hitl.washington.edu/artoolkit/documentation/

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B., 2001. Recent Advances in Augmented Reality. IEEE Comp. Graph. & App 21 (6), 34-47.

Bhatt, G., 2004. Bringing virtual reality for commercial Web sites. International Journal of Human-Computer Studies 60 (1), 1-15.

Cho, Y., Im, I., Hiltz, R., Fjermestad, J., 2002. An Analysis of Online Customer Complaints: Implications for Web Complaint Management. 35th Hawaii International Conference on System Sciences (HISS '02) 7, 176b.

Daugherty, T., Li, H., Biocca, F., 2005. Experiential commerce: A summary of research investigating the impact of virtual experience on consumer learning, Society of Consumer Psychology: Online Advertising. Mahwah, NJ: Lawrence Erlbaum Associates.

Fomenko, V., 2006. Generating Virtual Reality Shops for E-commerce. Ph.D. Dissertation, Vrije Universiteit Brussel.

Hughes, S., Brusilovsky, P., Lewis, M., 2002. Adaptive navigation support in 3D e-commerce activities. AH'2002, Malaga, Spain, 132-139.

Kato, H. and Billinghurst, M., 1999. Marker Tracking and HMD Calibration for a Video based Augmented Reality Conferencing System. Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, San Francisco, CA, 85-94

Lefebvre, L.A., Lefebvre, E. 2002. E-commerce and Virtual Enterprises: Issues and Challenges for Transition Economies. Technovation 22 (5), 313-323.

Lu, Y., Smith, S. 2006. Augmented Reality E-Commerce Assistant System: Designing While Shopping. Proceedings of IDETC/CIE’06, Philadelphia, PA, paper number DETC2006-99401.

Sanna, A., Zunino, C., Lamberti, F. 2002. HAVS: a human animated VRML-based virtual shop for e-commerce. In SCI'02 Proc., vol. XII, 24-29.

Standing, C. 2000. Internet Commerce Development,  Artech House Computing Libraty, Hardcover.

Swan II, J.E., Gabbard, J.L. 2005. Survey of User-Based Experimentation in Augmented Reality. Proceedings of 1st International Conference on Virtual Reality, Las Vegas, Nevada.

Zhang, X., Navab, N., Liou S.P. 2000. E-Commerce Direct Marketing using Augmented Reality. IEEE International Conference on Multimedia and Expo (I), 88-91.

Zhu, W., Owen, C.B., Li, H., Lee, J.H. 2006. Design of the PromoPad: an Automated

Augmented Reality Shopping Assistant. 12th Americas Conference on Information Systems, Acapulco, Mexico.

Howell, D.C. Statistical Methods for Psychology. 6th Edition.  Thomson Wadsworth, 2007.

# CHAPTER 3. A COMPREHENSIVE TOOL FOR RECOVERING 3D MODELS FROM 2D PHOTOS WITH WIDE BASELINES

A paper published in *the Transactions of The ASME, Journal of Computing and Information Science in Engineering*

Yuzhu Lu   Shana Smith

Virtual Reality Applications Center, Human Computer Interaction Program, ISU

1620 Howe Hall, Ames, IA 50011-2274

yuzhu@iatate.edu, sssmith@iastate.edu

## Abstract

Recovering 3D objects from 2D photos is an important application in the areas of computer vision, computer intelligence, feature recognition, and virtual reality. This paper describes an innovative and systematic method, which integrates automatic feature extraction, automatic feature matching, manual revision, feature recovery, and model reconstruction, into an effective and integrated 3D object recovery tool. The proposed method is a convenient and inexpensive way to recover 3D scenes and models directly from 2D photos. New automatic key point selection and hierarchical matching algorithms were developed for matching 2D photos with wide baselines. The method uses a universal camera intrinsic matrix estimation technique to eliminate the need for camera calibration experiments. A new automatic texture-mapping algorithm was also developed for finding the best textures in 2D photos. The paper includes some examples and results to show the capabilities of the new method.

**Keywords:** 3D recovery, Stereo matching, Computer modeling, Wide baseline

## 1. Introduction

With the rapid and widespread application of virtual models in many areas, methods for creating 3D models from real scenes are greatly needed. Traditional manual model building is labor-intensive and expensive. Thus, methods for automatically constructing 3D computer models have recently received much attention [1,2].

Much prior work has been conducted concerning recovering existing 3D environments. The resulting recovery methods can be classified into two categories: those that use scanning devices [3,4], and those that use cameras [5,6,7,8,9]. Scanning devices can automatically reconstruct objects precisely, but they are very expensive and inconvenient to transport and use, especially in an outdoor environment. Thus, the proposed method uses 3D model recovery from 2D images, which were captured using cameras.

The proposed method uses two or more photos of the same objects to recover 3D information from the overlapped areas of the photos and, subsequently, to reconstruct the model. The process includes four steps: key feature selection, feature matching, recovery computation, and model reconstruction.

The features of an image are often expressed as discontinuities in image signals. In prior related research, image discontinuities were extracted, from first or second derivatives of the image signal information, as corner points [1,8,9,10,11,12], edges [13,14], or regions [7]. Common corner detection methods include the Kitchen-Rosenfeld [12], Harris [12], KLT [12], and Smith [12] corner detectors. Popular edge detection methods include the Sobel [14], Prewizz [14], Laplacian of the Gaussian [14], zero-crossing [14], Hough transform [14], and Canny [13,14] methods. Some of the more widespread region extraction methods include the snake [15], split and merge [16], and level set [16] methods.

Among the different feature extraction algorithms, the most widely used are the Harris corner detection method and the Canny edge detection method. The Harris corner detection method is less sensitive to noise in the image than most other corner detection algorithms. Consequently, the Harris method’s high reliability and stability have made it quite popular [12]. Canny [13], on the other hand, aimed to design an “optimal” edge detector by formally specifying an objective function to be optimized and then designing or deriving method operations from the objective function definition. The resulting Canny method is, therefore, less likely to be “fooled” by noise than other edge detectors, and also more likely to detect true weak edges than other edge detectors.

Feature matching has been both the focus of, and the bottleneck in, recent research related to recovering 3D information from 2D images. Feature matching processes are typically applied to different attributes of detected features: corner points [1,8,9,11,12,17], line edges [10,18], curved edges [19,20], and regions [5,7,21]. Point matching methods have been most widely used, in stereovision research, because corners are easy to detect and they are more stable and robust when viewer perspective changes. Most prior point-matching algorithms were designed based upon image similarity, uniqueness, continuity, and epipolar information [1,2,5,9,11,17,20,22].

Zhang et al. [11] used a classical cross-correlation method to get an initial point correspondence set and then used the relaxation method to optimize point correspondence. Next, they used a robust algorithm to find the eight best point correspondences and the epipolar geometry, and then, finally, they used the epipolar geometry as a constraint for refining their correlation matching and, in turn, to obtain final matching results. The Zhang et al. method is popular and has been used by many related studies, although the method’s performance is weaker when applied to images with a wide baseline [11,21].

Several region-based matching algorithms have been designed [21,23], but existing region-based stereo techniques are still unstable and often locally inaccurate, due to deformation that occurs with images taken from different perspectives.

Recovery computation (also called stereo triangulation) is relatively stable and sophisticated, when camera parameters are known. However, when camera parameters are not known, the camera must be calibrated [1,2,22,24,25,26], an operation which is inconvenient for many common users. Thus, camera calibration and, particularly, self-calibration research has emerged as another major research area [22,25,27]. Most calibration methods provide both intrinsic and extrinsic parameters. However, they require accurate knowledge of the 3D coordinates of scene points to obtain the parameters [25].

Guermeur and Louchet [25] presented a method, using a genetic algorithm, to estimate intrinsic camera parameters, without knowing the 3D coordinates of points in the scene. However, to be effective, their method depends upon finding a five by five matrix of regularly spaced points in the image, which is most often quite difficult. Similarly, Cipolla, Robertson, and Boyer [27] used architectural perspective information and three vanishing points to extract camera intrinsic parameters; however their method can only be used in architectural images.

After obtaining 3D point information from a pair of camera images, many prior studies have used the Delaunay triangulation method to automatically rebuild 3D scenes [1,9,17], although most prior studies did not describe, in detail, how they reconstructed 3D models from the 3D points which were found. Delaunay triangulation is very convenient, in some situations, but, like other prior methods mentioned above, it too has known problems, particularly when recovering an object which has more than one surface, special shapes, or holes.

Although many prior studies exist, related to recovering 3D models from 2D camera images, practically, problems arise when trying to use prior methods, due to their lack of sophistication. Problems with prior methods include, primarily, inability to complete 3D recovery automatically and difficulty dealing with images having wide baselines.

Prior research studies have commonly determined that the baselines of cameras used to take images greatly influences feature point correspondence and recovery results [1,2,6,28]. Matching features from 2D images with more narrow baselines is generally easier than matching features from images with wide baselines, but the recovered 3D information is less accurate. Using images with small baselines, when the images are similar, leads to large errors in the reconstructed model, due to the low signal-to-noise ratio (SNR) [1]. At the same time, using images with large baselines, when the images are less similar, makes it difficult to find stereo correspondences, but the recovered results are generally more accurate.

In conclusion, wide baseline photos must be used, to obtain an accurate reconstructed model from 2D photos. From our research study, we also found that finding a 100% correct correspondence for all key image features is still a challenge, which makes full automation difficult to achieve.

Another problem, which was discovered in prior related research studies, is that overemphasizing correspondence accuracy has a negative impact on feature completeness, in other words, how much recovered feature correspondences cover all key features of the object. Image completeness is most important when using edge or planar information to recover the shape of an object.

In addition, most prior work has focused on reconstructing objects with one continual surface [1,2,28]. The Delaunay triangulation method can be used to reconstruct continual surfaces without holes [1,2]. However, most scenes and objects in the real world have more than one continual surface [7,8], which makes automatic object reconstruction even more challenging.

Finally, as mentioned earlier, camera calibration is usually a necessary step in a stereo matching and recovery process [1,2,22,24,25,26,27]. Camera calibration can be used to determine a camera’s intrinsic parameter matrix, which is needed during image recovery computations. Although camera self-calibration has been studied to reduce image recovery complexity, proposed self-calibration methods are still very inconvenient for common users [2,22]. In addition, for reconstructing 3D objects from historical or existing images, it is not possible to carry out camera calibration experiments. Thus, a camera parameter estimation model needs to be developed, which is convenient and easy to use, for common users.

Several prior studies have completed research which is very similar to our study. However, there are also important differences. Debevec, Taylor, and Malik [26] presented a very capable and complete system for recovering 3D architectural models from 2D images, which used a three-step process composed of photogrammetric modeling, view-dependent texture mapping, and model-based stereopsis. However their system can only be used to recover architectural building models because they used many architectural constraints. Their system also requires a lot of human interaction and considerable time is needed to decompose the scene into blocks, calibrate the camera, and compute depth maps for each surface. In addition, their view-dependent texture mapping method makes it almost impossible to output results in a widely used model format, such as VRML or 3DS.

To improve Debevec et al.’s system, Cipolla, Robertson, and Boyer [27] proposed a method for extracting camera intrinsic parameters from uncalibrated architectural images and for finding a resulting projection matrix. Their proposed method can help to determine better feature correspondence from architectural perspective information. However, their method, once again, can only be used to recover architectural building models because their method also depends upon important architectural constraints, such as parallelism and orthogonality. In their approach, all selected edges must be parallel or perpendicular to each other, to find three vanishing points. Most objects, other than architectural buildings, do not meet such constraints, in real-world scenes.

Finally, Strecha, Tuytelaars, and Gool [28] presented a partial differential equation (PDE) based algorithm for extracting dense depth information from multiple wide baseline images. However, their method can only be used to recover 3D scenes as one connected surface.

To address existing problems related to reconstructing 3D objects from 2D camera images, the investigators developed a systematic partially automated method for recovering 3D models directly from 2D photos with wide baselines. The investigators also developed automatic feature information extraction and hierarchical matching algorithms, as well as a tool which users can use to edit key points, revise possible mismatches, and select triangles for reconstructing a model with multiple surfaces. A new method was developed which uses statistical analysis to estimate universal camera intrinsic parameter matrices, without camera calibration. Finally, a new texture-mapping algorithm was developed for automatically selecting the best textures from different photos.

## 2. Epipolar Geometry

Epipolar geometry is the intrinsic projective geometry which provides the constrain between two views of the same scene. Epipolar geometry is independent of scene structure, since it only depends on the camera's internal parameters and the relative poses used to capture images. As a result, epipolar geometry can be used as a constraint for finding corresponding points during stereo matching. Most recent related research studies have used epipolar geometry for stereo matching [1,2,5,6,9,11,17,20,22,27].

The epipolar geometry for two camera views is the geometry which lies on the intersection of the two image planes with the pencil of planes having the baseline (i.e. the line joining the camera centers) as an axis. As shown in Figure 3.1, point P, in 3D space, is viewed by two cameras: COP1 and COP2. P1 and P2 are the image positions of point P on the two image planes. From the figure, we can see that the image points P1 and P2, 3D point P, and the two camera centers (COP1, COP2) are coplanar. This property is most significant when searching for a point correspondence in stereo views. The plane defined by P and the two camera centers (COP1, COP2) is called the epipolar plane. The two intersections of the epipolar plane with the two camera image planes are called the epipolar lines. The line connecting COP1 COP2 is called the baseline of the two cameras. The baseline intersects the image planes at the conjugate points e1 and e2, which are called epipoles.

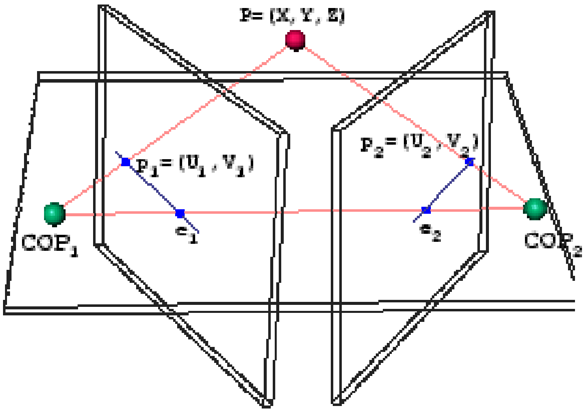


**Figure 3.1 Epipolar Geometry**

If both camera positions and one image point P1 are known, then the epipolar plane is also known and, as a known constraint, image point P2 must lie on the epipolar plane and one epipolar line. Thus, when searching for P2, the point on the second image plane which corresponds with point P1, the search can be restricted to the epipolar line, rather than the entire image plane.

After some manipulation, the epipolar geometry can be expressed as shown in Equation 3.1 [6,11,22]. In Equation 1, ($u_1$, $v_1$) and ($u_2$, $v_2$) are the coordinates of P1 and P2, within their respective image planes.

$$[u_1, v_1, 1]\mathbf{F}\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = 0 \tag{3.1}$$

In Equation 3.1, **F**, the fundamental matrix, is a three by three matrix that contains nine parameters, which include both cameras’ parameters and the rotation and translation information between them. The fundamental matrix can include an arbitrary scale factor [11].

With a fixed scale factor, there are eight degrees of freedom in **F**. Thus, Equation 1 is a typical linear algorithm [6].

As a result, Equation 1 can also be expressed as a linear equation:

$$u_1u_2f_{11}+u_1v_2f_{12}+u_1f_{13}+v_1u_2f_{21}+v_1v_2f_{22}+v_1f_{23}+u_2f_{31}+v_2f_{32}+f_{33}=\mathbf{0} \tag{3.2}$$

From Equation 3.2, it is clear that the fundamental matrix can be determined if at least eight points are known. From a set of n point matches, a set of linear equations of the form **A F** = 0 can be obtained.

After calculating the fundamental matrix (i.e. after the epipolar geometry is known), Equation 3.2, in which all $f_{mn}$ are known, can be used as epipolar constraints to search for the point that corresponds with any point P1.

## 3. Methodology

The method proposed in this paper includes five steps:

1. Extracting key feature points from 2D photo images.
2. Feature matching from wide baseline stereo images, including initial hierarchical feature matching followed by robust feature matching using epipolar contraints.
3. Human interaction: revising mismatched points, adding missing features, and selecting triangles for reconstruction.
4. Recovering 3D information from feature correspondences.
5. Reconstructing the 3D object from the recovered features and applying texture mapping.

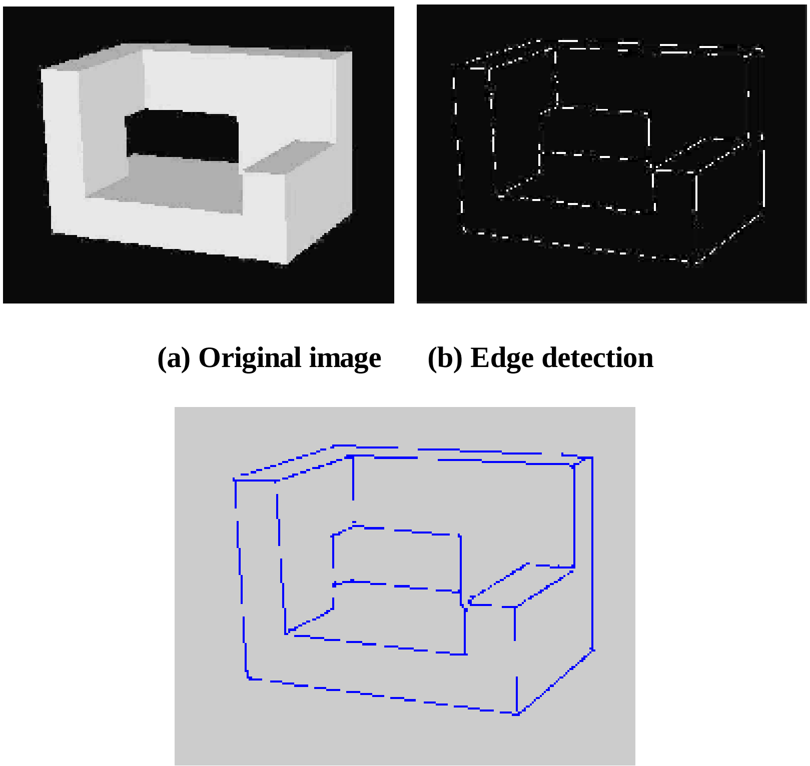
One of the investigators' primary goals was to accomplish feature detection and feature matching using as few key feature points as possible. They determined that detecting edges and then using segment information from the detected edges was the best choice for meeting their goal. As a result, the investigators developed a new hierarchical feature point matching method for determining initial stereo correspondences between key points and edge segments from dissimilar images. The method was designed, in particular, for determining stereo correspondences from wide baseline stereo images. Other known technologies were used to find eight (or more) best-matched points from the given initial matching, from which the epipolar geometry was also found.

As discussed earlier, it is difficult to fully automate the 3D object recovery process. Therefore, in the proposed method, a small amount of human interaction is still used. As a result, a user interface was designed, which allows users to add missing key feature points, revise poor feature correspondences, and select triangles for reconstruction. Triangulation is carried out to produce 3D information from matched feature points derived from epipolar information. The manual tool takes matching results, as input, and provides output which is integrated into the remaining automated matching process. After completing manual operations, the 3D model is automatically reconstructed and texture mapped with the best textures from the photos. Since the proposed method currently depends upon a limited amount of human interaction, developing a user-friendly interface was also a primary consideration when designing and developing the overall method.

## 3.1. Extracting Key Points

Feature points are points that define the main characteristics of a 2D image. In the proposed method, geometric information is the primary image characteristic that needs to be recovered. After considering both the Harris corner detection method [1,9,10] and the Canny edge detection method [8,13], the investigators chose to use the Canny method for extracting segment information. The investigators chose to use the Canny edge detection method for two primary reasons. First, edge detection gives more complete geometric information than corner detection. Second, edge segments can be displayed using only two end points and, therefore, they are easy to edit or revise, using manual methods.

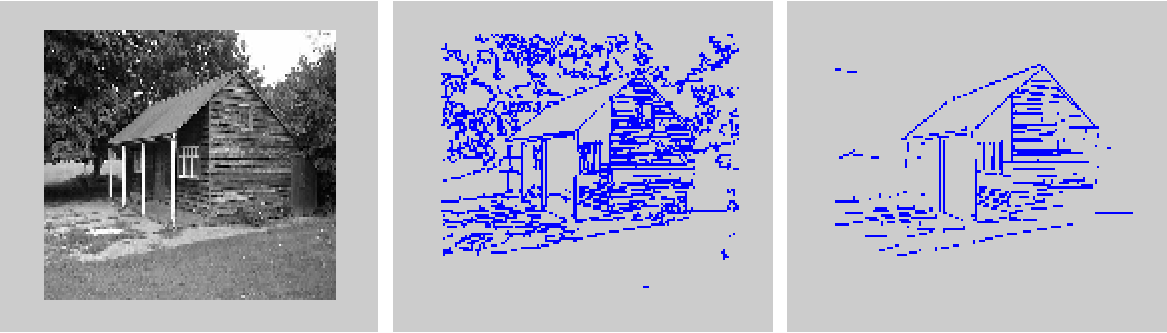
The proposed method currently uses Peter’s MATLAB Toolbox to implement Canny’s algorithm and, thereby, to detect edge segments from end points in stereo camera images [13]. Figure 3.2 illustrates the process of segment extraction, for one camera image. The key feature points, which were extracted, were used as input for the subsequent automatic feature matching process.



**(a) Original image** **(b) Edge detection**

**(c) Segment extraction**

**Figure 3.2 Process of feature segment extraction**



**(a) Original image** **(b) Segment extraction** **(c) Features kept after noise filtering**

**Figure 3.3 Process of background noise filtering**

In the proposed method, background noise is filtered out of the feature set, using segment length as a filtering parameter, before passing segment features to the automatic feature matching process,. With segment length as a threshold parameter, smaller segments are discarded and only main image edge segments are kept (as shown in Figure 3.3). Users can set the length threshold according to how much detail they want to keep, as main features, and how large the image is. For the example shown in Figure 3.3, the image size was 511 by 383 pixels, and the threshold value was set to 10 pixels.

## 3.2. Feature Matching Using a Hierarchical Matching Algorithm

Prior findings related to using epipolar constraints have contributed greatly to stereo matching research [1,2,9,10,11]. Compared to other matching methods, methods which use epipolar constraints are typically more robust.

As shown in Section 2, to implement a feature matching method based upon epipolar constraints, it is only necessary to find eight well-matched points, from which the epipolar geometry may then be determined [11]. However, in general, finding eight well-matched points is a challenging problem. Typically, it is almost impossible to check all possible combinations of extracted feature points. Therefore, an initial seed matching is needed for providing an initial candidate matching for epipolar geometry computation. The quality of the initial seed matching affects the final matching results, especially when key feature points are not contained in a large-scale space.

The most widely used method to obtain an initial matching set is a classical cross correlation method [11,20]. However, classical cross correlation methods and other commonly used methods for finding an initial seed matching usually do not work very well when they are applied to images with wide baselines, because most commonly used methods for finding an initial seed matching require high similarity between the two images.

Therefore, the investigators developed a new hierarchical algorithm for obtaining an initial correspondence set, as shown in Figure 3.4. In the proposed method, first, feature segments are matched. Since edge segments have more attributes than corner points (such as length, position, direction, and background color information), matching accuracy is increased, particularly when using large baseline images, which have a low degree of similarity. Second, end points of the edge segments are matched based upon segment matching results from the first step. If edge segments from the first step are well matched, accuracy in the second step is very high.

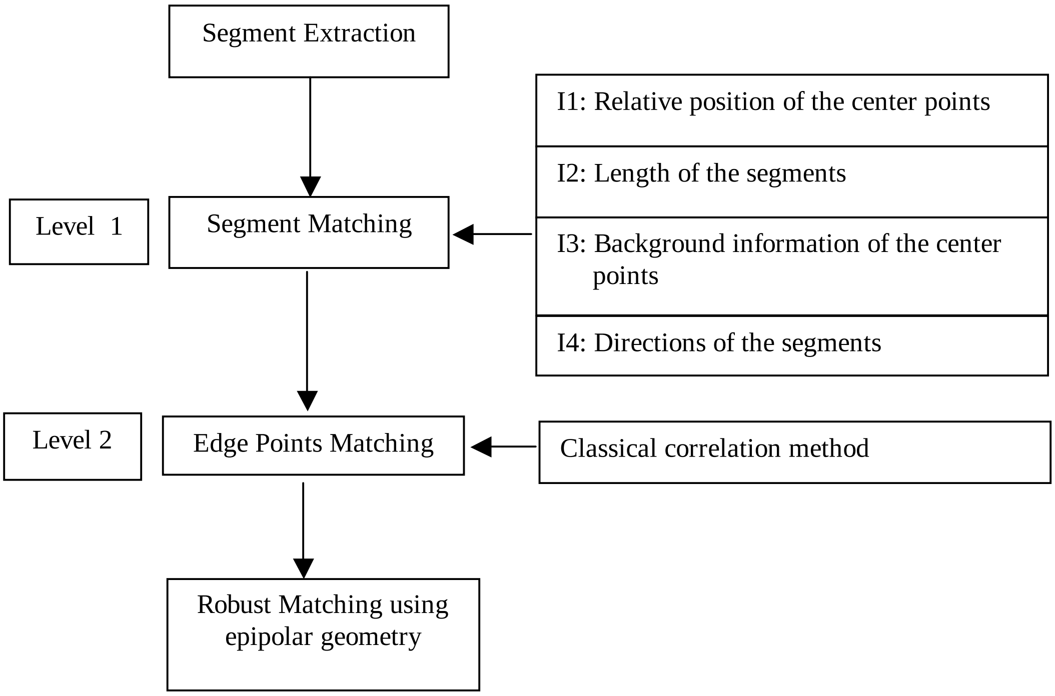
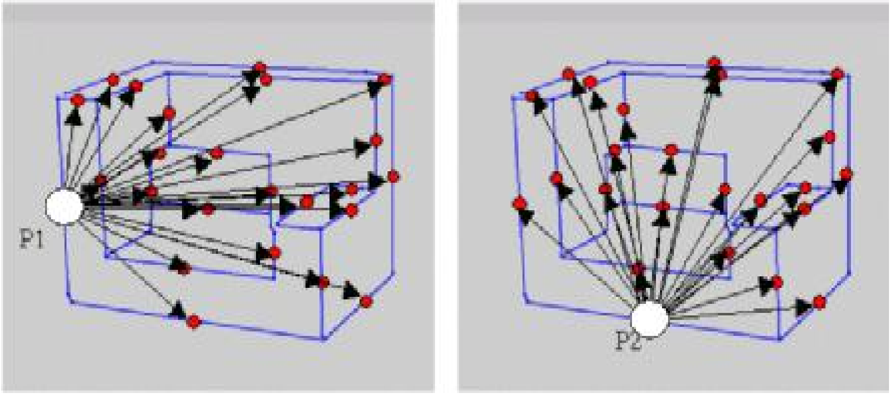

Segment Extraction
I1: Relative position of the center points
I2: Length of the segments
Level 1
Segment Matching
I3: Background information of the center points
I4: Directions of the segments
Level 2
Edge Points Matching
Classical correlation method
Robust Matching using epipolar geometry

**Figure 3.4 Hierarchical matching algorithm**

The new algorithm uses four indices for segment matching. The first index considers the relative positions of the center points of the extracted segments, which is represented by the summation of the vectors from the segment centers to all other segment centers. For example, as shown in Figure 3.5 (a), the index vector for P1 is found by adding up all the vectors from P1 to all other segment center points, and, as shown in Figure 3.5 (b), the index vector for P2 is found by adding up all the vectors from P2 to all other segment center points. The second index considers the length of the segment, and is represented by the distance between the segment's two end points. The third index considers the background information for each segment, which is represented by a 7 by 7 intensity matrix about a neighborhood around the center point of the segment. The fourth index considers the direction of the segment, which is represented by the angle of the segment vector.



**(a) Relative position of P1  (b) Relative position of P2**

**Figure 3.5  First indices for P1 and P2 - relative position**

In the proposed method, the four indices for each edge segment in the two camera images are compared, and the difference between each index value for each pair of segments is computed and added together, as shown in Equation 3.3. Potentially matched segments are chosen to be the segments that have the smallest differences between index values.

$$\mathbf{CP}=\boldsymbol{a}*(\boldsymbol{I_{1x}} - \alpha* \boldsymbol{I_{1x'}})+\boldsymbol{b}*(\boldsymbol{I_{2x}}- \alpha* \boldsymbol{I_{2x'}})+\boldsymbol{c}*(\boldsymbol{I_{3x}}- \boldsymbol{I_{3x'}})+\boldsymbol{d}*(\boldsymbol{I_{4x}}- \boldsymbol{I_{4x'}}) \qquad (3.3)$$

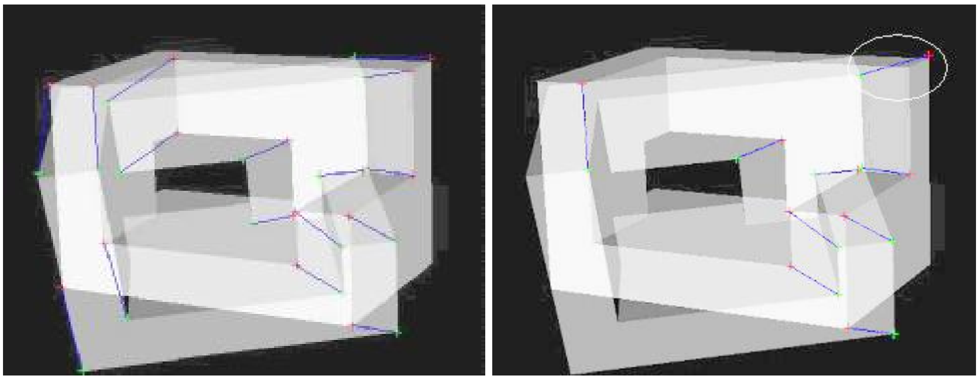
In Equation 3.3, $I_{1x}$, $I_{2x}$, $I_{3x}$, $I_{4x}$, $I_{1x'}$, $I_{2x'}$, $I_{3x'}$, and $I_{4x'}$ are the four index values for a pair of segments in the two images. As also shown in Equation 3.3, four weights, *a, b, c,* and *d,* can be defined, one for each of the four indexes, based on the relative importance of each of the four indices. In this study, the four normalized weights ***a*** = 1, ***b*** = 1, ***c*** = 2, ***d*** = 0.12 were used.

In addition, in Equation 3.3, an estimated scale parameter, α, was used to remove scaling problems due to size differences between the two images used. In the proposed method, the scale parameter $\alpha$ is automatically estimated and calculated every time a 3D scene is created from a set of 2D images, from the bounding box size ratio for the same object in the two images.

In the first level of the matching process, potential matched segments can be found, but matching directions for the segments cannot be determined. Consequently, in the second level of the matching process, a classical cross correlation method is used to match the end points of matched segments. If segments have been correctly matched in level one, then level two matching is easier and faster than in prior methods, because there are only a few candidate points with which each point can be matched.

As a result, the proposed hierarchical matching algorithm finds an initial correspondence (seed matching) for each point in the initial set of key feature points. A least squares method is then used to find the eight key feature points which are best matched. Eight matches are randomly and recursively selected, from which a fundamental matrix is calculated and then used to estimate all the matches and the least squares error, until a minimal least squares error value is found. The process results give the eight best-matched points, the exact fundamental matrix, and the algebraic representation of the epipolar geometry.

After the fundamental matrix is calculated, it is used to find inliers, among the seed matches found in the level-two matching, and reject the outliers. Figure 3.6 shows a comparison between a final matching result using the proposed hierarchical matching algorithm and a final matching result using the classical cross correlation algorithm [1,8,9,11,12,17,20]. Results for the given example show that the proposed hierarchical matching algorithm, for the given example, can produce more correct corresponding key points (16 matches) than the classical cross correlation algorithm (9 matches), and that the quality of the initial correspondence set can greatly affect the final matching result.



**(a) Proposed matching algorithm** **(b) Classical cross correlation algorithm**

**Figure 3.6 Matching result comparison**

**Table 3.1 Matching Experiment results**

| | Total Features | Matched Features | | Average Completeness | Mismatches | | Average Accuracy |
|---|---|---|---|---|---|---|---|
| | | Average | St.d | | Average | St.d | |
| Hierarchical algorithm | 21 | 15.4 | 0.966 | 73.3% | 0.5 | 0.527 | 96.8% |
| Classical cross correlation method | 21 | 11.4 | 0.843 | 54.2% | 1.3 | 0.675 | 88.6% |

An experiment was carried out to compare the two algorithms using MATLAB on a PC with a Pentium (R) 4 CPU 2G and 512 RAM. In the experiment, correct features and edges were used as the input to the two algorithms. Each algorithm was run ten times and the matching results were recorded and compared. Results of the experiment are shown in Table 3.1. From Table 3.1, the proposed hierarchical algorithm gave more matches and had better accuracy than the classical cross correlation algorithm.

Since the proposed hierarchical algorithm is used to generate an initial correspondence set, before finding the epipolar information, various robust strategies can also be used, as a following step, to further improve the completeness and accuracy of matching results, such as a relaxation process, searching point correspondences a second time using the epipolar geometry as a constraint, as suggested by Zhang et al. [11], or the epipolar geometry based contour matching algorithm suggested by Han and Park [20]. However, additional processing would cost more in both time and resources.

## 3.3. Human Interaction

Fully and automatically recovering a 3D model from 2D photos with wide baselines is difficult [1,2,21], and, at the current state of the art, some manual interaction is still needed. In addition, the Delaunay triangulation method is not suitable for 3D model recovery, in most cases, because it can only been used for objects composed of one surface, as discussed earlier. As a result, the investigators designed a user-friendly tool for allowing users to remove remaining noise, to add or delete corresponding key feature points from the matching results, and to create triangles for reconstructing the 3D model. The tool was fully integrated with both an automatic matching tool and a 3D recovery tool.

## 3.4. Recovering 3D Information

The relationship between a 3D point's coordinates and its corresponding image plane coordinates, when viewed through a camera, is shown in Equations 3.4 and 3.5. *S* is a scaling factor, while (*x*, *y*, *z*) and (*u*, *v*) are the corresponding 3D point coordinates and camera image coordinates. **P** is a three by four perspective projection matrix, which can be decomposed into an intrinsic camera matrix **A** and an extrinsic matrix, which contains rotation and translation information (**R**, **T**), as shown in Equation 3.5.

$$S\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad (3.4)$$

$$\mathbf{P} = \mathbf{A[RT]} \qquad (3.5)$$

As a result, the relationships between the coordinates of a 3D point and its coordinates on two image planes, when viewed through two cameras, can be expressed by Equation 3.6, where **A**$_1$ and **A**$_2$ are the two cameras' intrinsic matrices. In Equation 3.6, the first camera is chosen to be an original reference camera, and the second camera is chosen to be a transformed camera, with respect to the first camera. The two parts of Equation 3.6 can be used, by the principle of stereo triangulation, to recover 3D coordinates when all parameters are known.

$$
\begin{aligned}
s_1\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} &= \mathbf{A}_1[\mathbf{I0}]\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \\
s_2\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} &= \mathbf{A}_2[\mathbf{RT}]\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}
\end{aligned}
\tag{3.6}
$$

In the investigators' proposed method, after matching corresponding key points and constructing triangular surfaces, prior methods are used to carry out stereo triangulation, using Equation 6, to recover corresponding 3D information [17,22]. Prior research has also shown that camera calibration (or self-calibration) should be completed to obtain the camera's intrinsic parameter matrix [1,2,6]. The intrinsic matrix A in Equation 7 and the fundamental matrix F, obtained during the matching process, can then be used to calculate the rotation and translation parameters between the two cameras, which, in turn, can be used to calculate the 3D information of the object. However, completing camera calibration experiments is very inconvenient and time consuming or, sometimes, even impossible, especially when recovering historical scenes from old photographs.

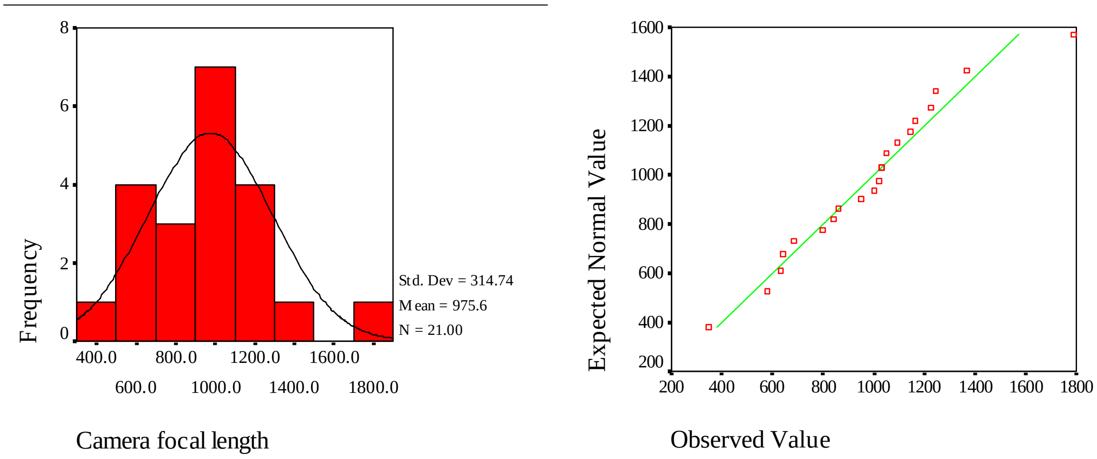
$$
\mathbf{A}=\begin{pmatrix} fk_u & fk_u cot\theta & u_0 \\ 0 & fk_v/sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}
\tag{3.7}
$$

Equation 3.7 shows that there are six important intrinsic camera parameters: focal length $f$ of the cameras, aspect ratios $k_u$ and $k_v$, angle $q$ between the retinal axes, and coordinates of the principal camera points $u_0$ and $v_{0.}$ [1,2,22]. Xu, Terai, and Shum [22] found that, if high precision is not required, most of the intrinsic camera parameters can be assumed; angle $q$ between the retinal axes of the two cameras can be set to $p/2$ ($q = p/2$), aspect ratios for both cameras can set to 1 ($k_u = k_v = 1$), and principal points for both cameras can be set to their respective image centers. Thus, the only unknown parameter, which cannot be assumed, is camera focal length *f*. The intrinsic camera matrix can then be rewritten as shown in Equation 3.8, in which focal length *f* for both cameras is assumed to be the same:

$$\mathbf{A}=\begin{pmatrix} f & 0 & pixel_x/2 \\ 0 & f & pixel_y/2 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.8}$$

To estimate the focal length *f* of an unknown camera, the investigators conducted a study of *f* value distributions for common cameras. They surveyed 21 different research studies to determine typical camera focal length. A Shapiro-Wilk normality test was conducted to test the survey data, and the results showed that the statistic is 0.968 with a p value of 0.665. So it is safe to say that actual camera focal lengths are normally distributed, N (975.6, 314.7), while 70% of focal lengths from the sample varied within a narrow range [700-1300], as shown in the histogram and Q-Q plot in Figure 3.7. In the proposed method, survey results are used to estimate the intrinsic camera parameter matrix, when actual intrinsic parameters are unknown. The proposed 3D recovery tool also allows users to adjust camera focal length to find the best 3D object recovery results. Using different camera focal lengths to recover 3D information leads to minor differences in the recovered 3D object model. The new tool allows users to adjust camera focal length to find the best results.

An experiment was carried out to analyze how three suggested focal length values affected algorithm performance, when used in the camera intrinsic matrix during 3D object model recovery. Four important parameters (three principal angles and a depth length ratio) were defined to evaluate the precision of the recovered results, as shown in Figure 3.8. Experimental data, which includes real values, the recovered values, and error rates, are listed in Table 3.2. In the experiment, the investigators considered the recovered object acceptable if errors for all of the four parameters were less than 10%; otherwise the recovered object was considered unacceptable. Table 3 gives final results of the experiment, which show the performance of the camera intrinsic estimation method and the three trial focal lengths.



**(a) Histogram with normal curve** **(b) Q-Q plot**

**Figure 3.7 Camera focal length analysis for 21 research studies**

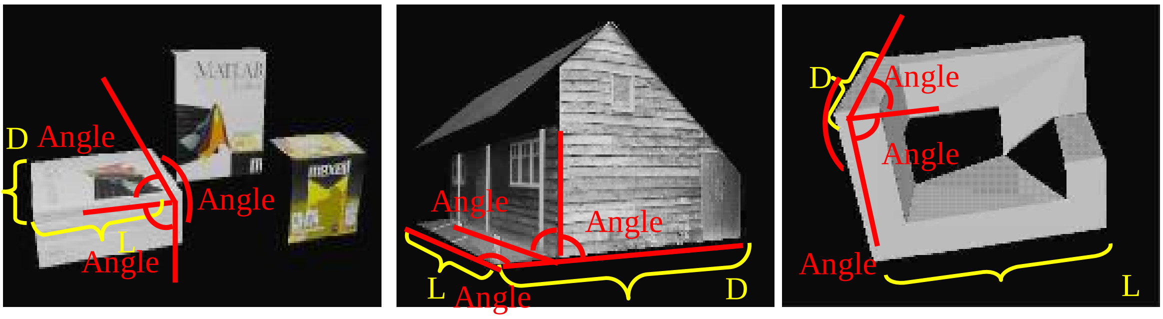


**Figure 3.8 Several parameters were defined to evaluate the precision of recovered results**

**Table 3.2 Experiment results**

| Scenes Recovered in Fig. 8 | | Focal Lengths | | | Real values |
|---|---|---|---|---|---|
| | | 700 | 1000 | 1300 | |
| Books | Ratio (D:L) | 0.9241(15.5%) | 0.8285(3.2%) | 0.7665(4.2%) | 0.8 |
| | Angle 1 | 98.3676(9.3%) | 85.5806(4.9%) | 74.7995(16.9%) | 90 degrees |
| | Angle 2 | 82.5733(8.3%) | 90.8023(0.9%) | 95.4334(6.0%) | 90 degrees |
| | Angle 3 | 94.7521(5.3%) | 88.7101(1.4%) | 84.7634(5.8%) | 90 degrees |
| House | Ratio (D:L) | 0.6695(6.3%) | 0.7233(14.8%) | 0.7569(20.1%) | 0.63 |
| | Angle 1 | 87.2021(3.1%) | 98.9803(10.0%) | 109.9157(22.1%) | 90 degrees |
| | Angle 2 | 87.0966(3.2%) | 85.0006(5.6%) | 83.8777(6.8%) | 90 degrees |
| | Angle 3 | 89.3648(0.7%) | 86.6497(3.7%) | 85.4403(5.1%) | 90 degrees |
| Block | Ratio (D:L) | 1.1632(73.6%) | 0.7344(9.6%) | 0.6416(4.2%) | 0.67 |
| | Angle 1 | 109.4342(21.6%) | 91.6916(1.9%) | 76.9941(14.5%) | 90 degrees |
| | Angle 2 | 79.8519(11.3%) | 81.8136(9.1%) | 79.1281(12.1%) | 90 degrees |
| | Angle 3 | 99.3019(10.3%) | 93.9133(4.3%) | 89.1525(0.9%) | 90 degrees |

**Table 3.3 Recovered results**

| Scenes Recovered in Fig. 8 | Focal Lengths | | |
|---|---|---|---|
| | 700 | 1000 | 1300 |
| Books | Unacceptable | Acceptable | Unacceptable |
| House | Acceptable | Unacceptable | Unacceptable |
| Block | Unacceptable | Acceptable | Unacceptable |

In this study, the investigators propose that three values (700, 1000, 1300) could be used to estimate the real camera focal length. Since focal length values decrease when the distance between the camera and the object increases, 1300 can be used to estimate short distance situations, 1000 can be used to estimate median distance situations, and 700 can be used to estimate long distance situations.

## 3.5. Reconstructing the 3D object

After obtaining all of the 3D feature point information, a new 3D model can be constructed, based upon connectivity information for the triangles created in prior steps of the method. Since the reconstructed solid model does not inherently contain surface texture information, texture mapping can be added to make the 3D model more realistic.

Surface texture information needed for texture mapping can be taken from the original 2D images. However, since the object is recovered from information contained in at least two photos, and since each photo is taken from a different angle (and therefore, typically, shows certain details in a slightly different way), the proposed method needs to determine which photo to use. Normally, a better surface texture should be found on the image created by the camera that captures a larger area of the object surface and which, therefore, most likely contains details that are more clear, for the given surface. Based upon that assumption, the investigators designed a new algorithm for texture mapping that compares the textures from corresponding triangles from the two images and selects the texture from the object surface with the larger area.

Figure 3.9 shows an example of the new texture-mapping method. The proposed texture mapping method is better than prior view-dependent texture mapping methods [26] because the results can be output in general model formats, such as VRML and 3DS, for further editing or other applications.

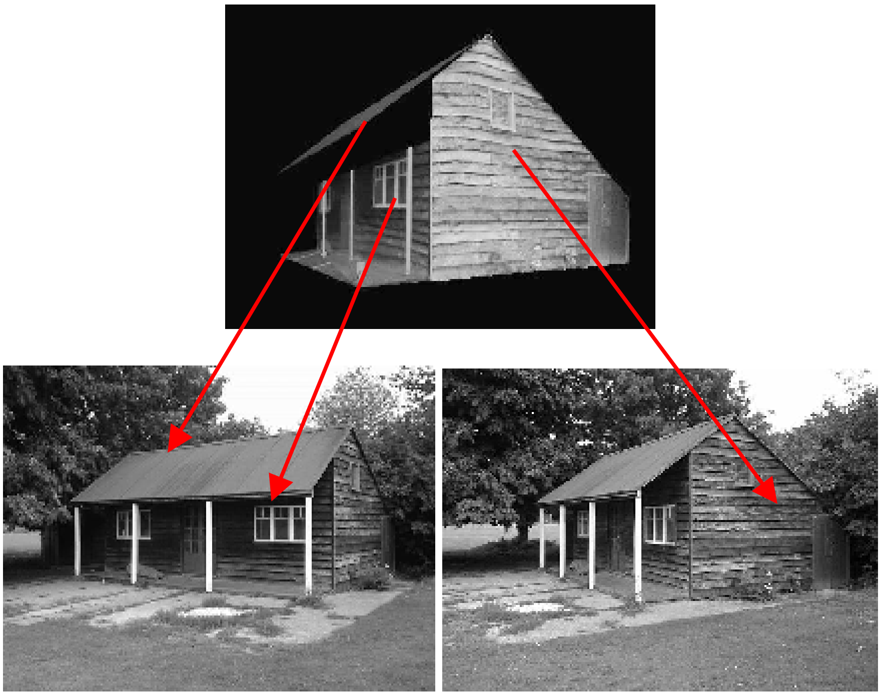


**Figure 3.9 Reconstruction with texture mapping**

## 4. Results

The investigators developed a comprehensive and integrated method, and corresponding software tool, for recovering 3D geometries from 2D images with wide baselines. The tool includes capabilities for:

- Importing images
- Extracting key feature points
- Automatic feature matching
- Editing functions
- Recovering 3D information, with texture mapping

Most steps of the method are completed automatically, apart from manual operations for revising automatically generated corresponding points and selecting triangles for reconstruction. Manual interaction time is generally proportional to the number of feature points and the complexity of the models to be reconstructed. As a result, the proposed method greatly reduces overall time required, over prior methods, by using edge segment matching and epipolar geometry constraints to automatically complete hierarchical feature matching. In particular, with the proposed method, fewer feature points are required than with the Harris corner algorithm [8]. In addition, the proposed method allows users to estimate intrinsic camera parameter matrices, by setting and adjusting a camera focal length parameter, which can relieve users from having to carry out complicated and inconvenient camera calibration procedures. Thus, the proposed method makes 3D object recovery from 2D photos easier and more practical for almost any user.

## 5. Conclusions and Discussion

A comprehensive and systematic method was designed for recovering 3D object models from two 2D images with wide baselines. The method and tool were demonstrated, using practical case studies, and were shown to be an effective and convenient means for rapidly recovering 3D object models directly from photos. Noteworthy and novel contributions the investigators made to research in 3D object recovery from 2D camera images include:

1. A hierarchical feature-matching algorithm for generating an initial correspondence set for wide baseline images, which provides better results than prior feature matching algorithms.
2. A technique for estimating a universal intrinsic camera parameter matrix, by which camera calibration experiments can be eliminated.
3. A texture-mapping algorithm, by which reconstructed results can be output as standard 3D object models with different view textures.
4. A comprehensive and fully integrated process and tool for recovering a 3D model from 2D images, while most previous research only focused on one or two parts of the overall process.

Since the proposed method was only designed for recovering 3D scenes from two 2D photos, non-visible areas in either photo cannot be recovered. Using more images does not contribute much to improving the accuracy of recovered models. However, using more images can help to recover objects more completely. Currently, objects with simple geometric shapes (like buildings) are easier to recover than complicated objects (like trees and grass). Since the quality of the estimated intrinsic camera parameter matrix affects the recovered results, users can adjust camera focal length within a recommended range.

In the near future, for recovering complete 3D models, work will be completed for aligning and fusing model parts retrieved from more than two photos. The software tool will also be enhanced to provide capabilities for an automatic mesh generation.

## References

[1] Cornelis, K., Pollefeys, M., Vergauwen, M., and Van Gool, L., 2000, “Augmented Reality using Uncalibrated Video Sequences,” *2th European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE2000),* Dublin, Irleand, pp. 144-160.

[2] Pollefeys, M., 1999, “Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences,” PH.D thesis, Katholieke Universiteit Leuven, Heverlee, Belgium.

[3] Reed, M., and Allen, P., 1999, “3-D Modeling from Range Imagery: An Incremental Method with a Planning Component,” Image and Vision Computing, **17**, pp. 99-111.

[4] Stamos, I., and Allen, P., 2000, “3-D Model Construction Using Range and Image Data,” *Computer Vision & Pattern Recognition Conf. (CVPR),* pp. 531–536.

[5] Shum, H., and Szeliski, R., 1999, “Stereo Reconstruction from Multiperspective Panoramas,” *7th International Conf. on Computer Vision (ICCV'99),* Kerkyra, Greece, pp. 14-21.

[6] Jebara, T., Azarbayejani, A., and Pentland, A., 1999, "3D Structure from 2D Motion," IEEE Signal Processing Magazine, **16**(3), pp. 66-84.

[7] Baker, S., Szeliski, R., and Anandan, P., 1998, "A layered approach to stereo reconstruction," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR98)*, pp. 434-441.

[8] Baillard, C., and Zisserman, A., 2000, "A plane-sweep strategy for the 3d reconstruction of buildings from multiple images," International Archives of Photogrammetry and Remote Sensing, **32**(2), pp. 56–62

[9] Fitzgibbon, A. W., and Cross, G., 1998, "A. Zisserman, Automatic 3D Model Construction for Turn-Table Sequences", *Proc. European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pp. 155-170.

[10] Harris, C. G., and Stephens, M. J., 1988, "Combined Corner and Edge Detector", *Proc. 4th Alvey Vision Conf*., Manchester, England, pp. 147-151.

[11] Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q., 1995, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artificial Intelligence, **78**, pp. 87-119.

[12] Tissainayagam, P., and Suter, D., 2004, "Assessing the performance of corner detectors for point feature tracking applications," Image and Vision Computing, **22**(8), pp. 663-679.

[13] Canny, J. F. 1983, "Finding edges and lines in images," Master thesis, MIT. AI Lab.

[14] Gonzalez, R., and Woods, R., 2002, *Digital Image Processing*, 2nd edition, Prentice Hall.

[15] Gao, J., kosaka, A., and Kak, A., 1998, "A Deformable Model for Human Organ Extraction," *Proc. IEEE International Conf. on Image Processing*, **3**, pp. 323-327

[16] Pham, D. L., Xu, C., and Prince, J. L., 1998, “A Survey of Current Methods in Medical Image Segmentation,” Annual Review of Biomedical Engineering, **2**, pp. 315-337.

[17] Ma, Y., Soatto, S., Kosecka, J., and Sastry, S., 2003, *An Invitation to 3-D Vision: From Images to Geometric Models,* Springer-Verlag.

[18] Loaiza, H., Triboulet, J., and Lelandais, S., 2001, “Matching segments in stereoscopic vision,” IEEE Instruction & Measurement Magazine, **4**(1), pp. 37-42.

[19] Deriche, R., and Faugeras, O., 1990, “2-D Curve Matching Using High Curvature Points: Application to Stereo Vision,” *Proc. International Conf. on Pattern Recognition.* New Jersey, USA, pp. 240-242.

[20] Han, J. H., and Park, J. S., 2000, “Contour Matching Using Epipolar Geometry,” IEEE Trans. Pattern Anal. Mach. Intell., **22**(4), pp. 358-370.

[21] Tuytelaars, T., Vergauwen, M., Pollefeys, M., and Van Gool, L., 1999, “Image Matching for Wide baseline Stereo,” *Proc. International Conf. on Forensic Human Identification.*

[22] Xu, G., Terai, J., and Shum, H., 2000, “A Linear Algorithm for Camera Self-Calibration, Motion and Structure Recovery for Multi-Planar Scenes from Two Perspective Images, ” *Computer Vision & Pattern Recognition Conf. (CVPR),* PP. 2474-2479

[23] Vira, N., 2003, “Modeling of a Three-dimensional Digital Image from 2D Stereo Paris,” *Proceedings of the IASTED International Conference on Modeling and Simulation (MS 2003),* PP. 482-488.

[24] Zhang, Z., 1995, “Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images,” IEEE Trans. on Pattern Analysis and Machine Intelligence, **17**(12), PP. 1129-1139.

[25] Guermeur, P., and Louchet, J., 2003, “An Evolutionary Algorithm for Camera Calibration,” *ICRODIC*, Rethymnon, Greece, pp. 799-804.

[26] Debevec, P. E., Taylor, C. J., and Malik, J., 1996, “Modeling and Rendering Architecture from Photographs,” *In SIGGRAPH '96*, August 1996, pp. 11-20.

[27] Cipolla., R., Robertson, D., and Boyer, E., 1999, “Photobuilder-3D models of Architectural scenes from Uncalibrated Images,” *Proc. IEEE International Conference on Multimedia Computing and Systems*, Firenze, pp. 25-31.

[28] Strecha, C., Tuytelaars, T. and Gool L. V., 2003, “Dense Matching of Multiple Wide-baseline Views,” *ICCV 2003*, pp. 1194-1201

# CHAPTER 4. GPU-BASED REAL-TIME OCCLUSION IN AN IMMERSIVE AUGMENTED REALITY ENVIROMENT

A paper submitted to *the Transactions of The ASME, Journal of Computing and Information Science in Engineering*

Yuzhu Lu     Shana Smith

Human Computer Interaction Program

Iowa State University, Ames, IA, USA

Email: yuzhu@iatate.edu, sssmith@iastate.edu

## ABSTRACT

In this paper, we present a prototype system which uses CAVE-based virtual reality to enhance immersion in an augmented-reality environment. The system integrates virtual objects into a real scene captured by a set of stereo remote cameras. We also present a GPU-based method for computing occlusion between real and virtual objects, in real time. The method uses information from the captured stereo images to determine depth of objects in the real scene. Results and performance comparisons show that the GPU-based method is much faster than prior CPU-based methods.

**Keywords**: Immersive Augmented Reality, GPU programming, CAVE, real time occlusion.

## 1. INTRODUCTION

Augmented Reality (AR) is a technology which aims to mix or overlap computer-generated 2D or 3D virtual objects with a real world scene. Unlike Virtual Reality (VR), which replaces the physical world, AR enhances physical reality by integrating virtual objects into real-world scenes. The virtual object becomes, in a sense, an equal part of the natural environment. In recent years, a lot of research has focused on AR applications for, for example, computer-enhanced surgery, training programs, tour systems, industrial maintenance, etc.

AR systems can generally be classified into three categories, based on technologies: projector-based AR [17], optical see-through AR [9,15], and video see-through AR [2-8,10]. Projector-based AR uses projection technology, optical see-though AR uses an optical see-through head mounted display (HMD), and video see-through AR uses cameras and a closed-view HMD. Projector-based AR projects (or maps) virtual objects onto a geometrically similar physical object using a projector. Both optical see-through HMDs and closed-view HMDs are helmet-like devices which users wear on their heads. Using an optical see-through HMD, viewers can still see the external physical environment. Optical see-through AR projects virtual objects onto the semi-translucent lens of the HMD. Viewers can then see the virtual objects, within the physical environment, through the optical lens. However, using a closed-view HMD, viewers cannot see the external physical environment. Video see-through AR uses cameras to capture images of the physical environment, mixes them with virtual objects, and renders the integrated images onto small screens inside the closed-view HMD.

Over the past few years, among the three types of AR, video-based AR has attracted the most attention from researchers because video input provides more information about the real scene. However, HMD users often complain about display issues and discomfort [17,18]. More immersive and realistic AR display methods are still not available. On the other hand, CAVEs have been used extensively for fully immersive VR display. However, they currently have limited functionality for visualization. In particular, CAVEs are currently only used in VR applications which only render virtual scenes and virtual models.

In our study, to enhance the immersion, realism, and usability of AR systems, we integrated video-based AR capabilities into a virtual reality CAVE. Rather than use a clumsy

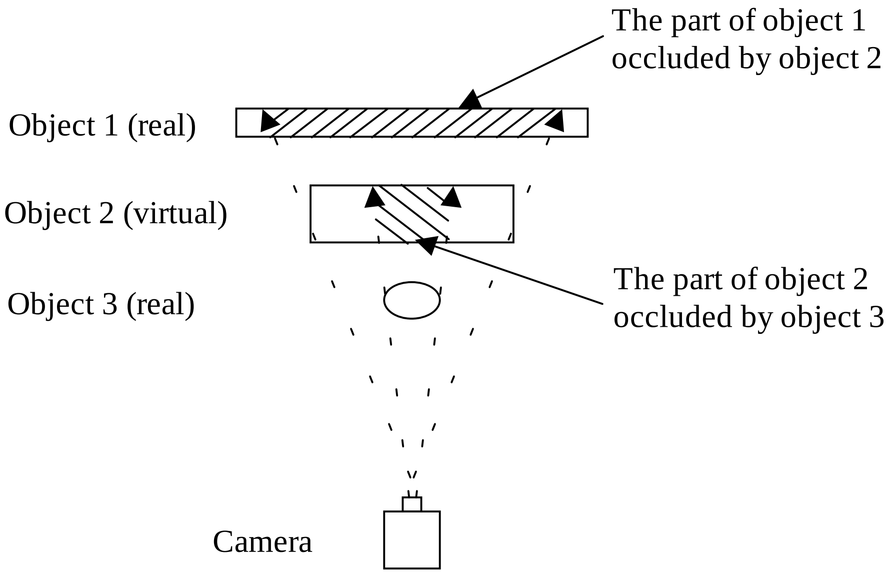


**Fig. 4.1 Top view of occlusion in Augmented Reality**

and uncomfortable HMD, as in a traditional video-based AR system, in our approach, we used simple and light-weight polarized glasses.

For an AR system, handling occlusion is a necessary and important part of the system. Accurately rendering occlusions provides more realistic scenes and helpful depth cue information. For example, Figure 4.1 shows three objects in the scene. If the view point is at the camera position, part of object 2 is occluded by object 3, since object 3 is in front of object 2. Similarly, part of object 1 is be occluded by object 2. Occlusion between real objects takes place naturally. However, if objects 1 and 3 are real objects, but object 2 is a virtual, realistic occlusion relationships between the real objects and the virtual object must be found by computation, and rendered accurately. Realistic mutual occlusion between real and virtual objects enhances users' perception that virtual objects truly exist in the real world scene, whereas incorrect occlusion confuses viewers [1].

To realize correct occlusion in an AR system, an understanding of the background scene is needed. Among the three types of AR systems: projective AR, optical see-through AR, and video see-through AR, video see-through AR has more strengths and potential for gaining depth cue information. Since video see-through systems use true video input, computer vision techniques can be used to extract depth cues from the video input. According to Breen, Whitaker, and Rose [2], there are generally two categories of occlusion methods available for video see-through AR: model registration and scene depth calculation. The primary limitation of the model registration method is that objects in the real scene must be known and modeled, which greatly narrows the method’s application. Thus, over the past few years, most video see-through AR systems have focused on the second method, depth calculation. However, a satisfactory real-time depth calculation method still does not exist because existing methods are still not fast enough, without hardware acceleration [3-10].

Stereo images are generally used for depth calculation. Corresponding pixels/objects from the left eye image and from right eye image can be matched. Based on the matching information and the given camera information, the depth of an object can be calculated. Then, a virtual background model of the scene can be constructed or the depths of real objects can simply be compared with the depths of virtual objects to decide which areas should be occluded.

Gibson, Cook, and Howard [3] presented a complete method to reconstruct a 3D scene from a video sequence. Their method includes improved Kanade-Lucas-Tomasi tracking [3] to keep the number of features stable, computing estimations from a previous tracking to reduce incorrect tracking features, a hierarchical approach for merging sub-sequences together, and Random Sample Consensus (RANSAC) based random sampling for self-calibration and calculation. However, their approach is not a real-time algorithm.

Lepetit and Berger [4] presented a semi-automatic method which makes use of key views for resolving occlusion in AR. After the user outlines occluding objects in key views, their system automatically detects occlusions. The limitation of their method is that it requires human input. Thus, the approach is also not a real-time method.

Schmidt, Niemann, and Vogt [5] presented a more robust method to calculate depth maps for a scene created from stereo views. In their method, first, a similarity accumulator is used to compute similarities and find corresponding points; consistency information is used to optimize their matching algorithm. Then, media filters and morphological operations [5] are used to fill the gaps in calculations. However, their experimental results show that their algorithm is also not a real-time method.

Duchesne and Hervé [6] proposed an innovative rendering method for rendering a virtual model in a real scene, in which a 3D virtual object is considered as a set of 3D points and vertices. Occlusion is realized by matching pixels from stereo images, to determine whether a virtual vertex is shown or not, without explicit reconstruction. Their approach offers a new rendering method for handling the occlusion problem in AR. However, with their method, real-time local stereo matching and depth calculation is still a challenge because point by point matching is time-consuming.

Berger [7] presented a contour-based method to solve the occlusion problem. Berger's method labels each contour point as either "behind" or "in front of" a virtual object and groups the points to avoid 3D reconstruction. However, the required information cannot be determined in real time, since the contour tracking and snake segmentation techniques used in their research are very time-consuming.

Kanbara, Okuma, and Takemura [8] presented a method using stereo video to get the depth information needed for determining occlusion in video see-through AR. In their method, they project the bounding box of a virtual object back to the left- and right-eye images, obtain edge information in the two boxes, and then match the edges. Their algorithm avoids computing depth information, except when there is a possibility of occlusion. Their results show that the enhanced scene, with occlusions, can only be determined after an obvious time latency, and the problem is more serious for high-resolution scenes.

Kiyokawa, Kurata, and Ohno [9] presented a new optical see-through system with occlusion capability. Their system takes advantage of the strengths of video see-through AR and uses a commercial FZ930 stereovision system from Komatsu to capture the depth map of a background scene. To create occlusion effects, the system uses an LCD panel in front of an opaque screen to block specific types of light from coming through the screen. Similarly, Gordon, Billinghurst, and Bell [10] used a different commercial stereovision device from Tyzx to calculate depth information for finding occlusions. However, both commercial stereovision devices use fixed baselines and special hardware processors. As a result, the proposed commercial devices are expensive and inconvenient for users.

Prior depth calculation methods, which all use a CPU to calculate occlusion information, are not fast enough to provide satisfactory real-time AR results, even when commercial stereovision systems are used to compute depth information for the background scene. However, recently, a breakthrough has occurred in graphics hardware. Fixed function pipelines have been replaced with programmable vertex and fragment processors. Graphics hardware is developing general programmable stream processor capability, with much faster speeds than a conventional CPU. As a result, a lot of recent research has been conducted which uses programmable graphics hardware to improve speed in graphics applications. For example, Purcell, Buck, and Mark [11,12] used a modern GPU as a general stream processor and developed a streaming ray tracing algorithm for the programmable GPU. They showed that their algorithm is very competitive, compared to current CPU-based ray casting. Sinha, Frahm, and Pollefeys [19] presented an approach to implement SIFT video feature extraction and KLT video feature tracking in GPU, which is 10 times faster than optimized CPU algorithm.

Fung and Mann [20] developed OpenVIDIA, an open library explores the way of using multiple Graphic Processing Units in parallel to accelerate image analysis and computer vision algorithms which includes stereo correspondence calculation. Their method is to compare the similarity of each pixel from the first texture with pixels in the corresponding search area in the second texture in GPU, and decide the correspondence of each pixel. However this is still not the suitable way to handle AR real-time occlusion because of their performance, which is also affected greatly by the search window size.

In our study, we introduce an innovative real-time stereo matching and depth calculation method for determining occlusions in video-based AR. The method can be used in any video-based rendering system. In our study, we use a CAVE as the rendering system, rather than a clumsy HMD. The method does not require special hardware acceleration, apart from a standard graphics card. Based upon the recommendations given by Purcell [11,12], we developed a streaming algorithm and broke processes into several key steps: preprocessing, segmentation, matching, depth calculation, and occlusion. The key steps are connected by streams of data. The streams of data are stored, temporarily, in texture and frame buffers. We also optimized the entire process to make it suitable for GPU-based computation. We also completed an experiment to compare the performance of our GPU-based method with a CPU-based approach.

## 2. BRINGING AR INTO A CAVE

A CAVE is a room-sized, multi-person, 3D video and audio system used to create an interactive immersive 3D environment, in which users experience a sense of presence, a feeling of "being there". CAVEs use high-performance graphics computers to generate stereo images, which are then rear-projected onto walls. Position information for the users, inside a CAVE, is usually captured from head trackers attached to special 3D glasses, which users wear when they are in the CAVE. Users can usually interact with the graphics environment using hand-held devices called wands.

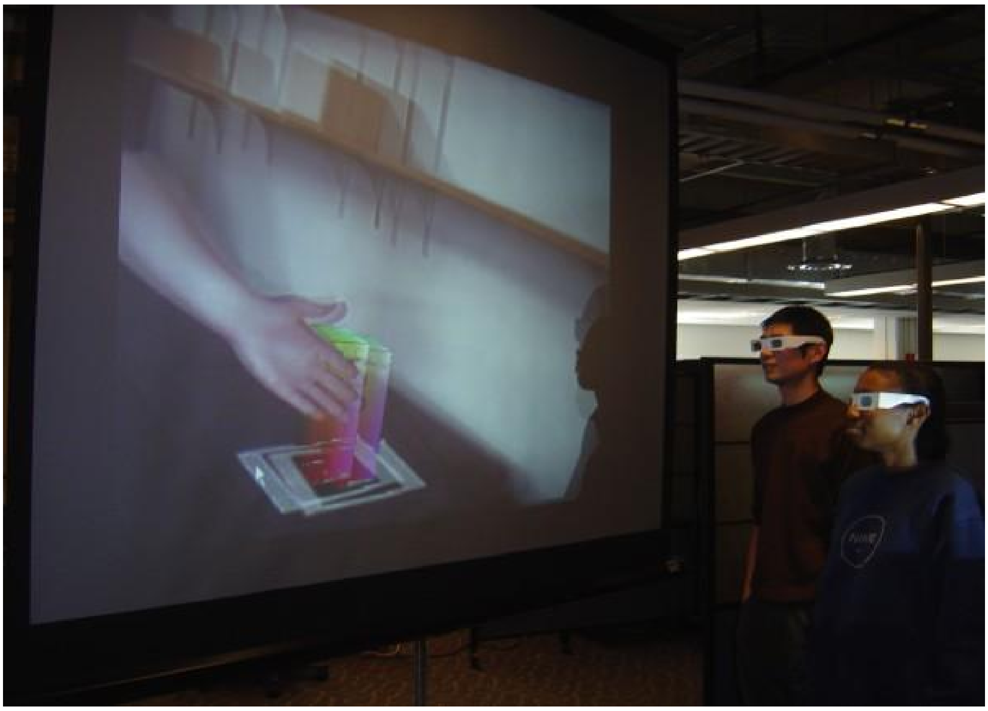


**Fig. 4.2 A portable one-wall BABY CAVE**

Prior to our study, CAVEs have only been used to render computer-generated graphics or virtual models. In our study, we integrated video-based AR into a CAVE, to improve the immersiveness and realism of AR systems. To complete the study, we used a portable single-wall BABY CAVE (as shown in Figure 4.2). The ARToolkit library was used for the tracking system. In the prototype system, to create a stereo AR scene, two images from stereo cameras (as shown in Figure 4.3) are simultaneously sent to the left and right viewpoint of the same application window which is developed based on ARToolkit as shown in Figure 4.5 (a). In each viewpoint, markers are detected and virtual objects are rendered with the video background as shown in Figure 4.5 (b). Then in the full screen mode, two separate well-adjusted projectors are connected to the two graphic card output ports. With NVIDIA horizontal spilt function, the two viewpoints are spitted and sent to the two projectors mounted with polarized filters. These two projectors project encoded images to the same special screen as an overlapped picture. Polarized glasses are needed to decode stereo images and send the left image to the left eye and right image to the right eye so that user wear this type of glasses could see the stereo AR scene.

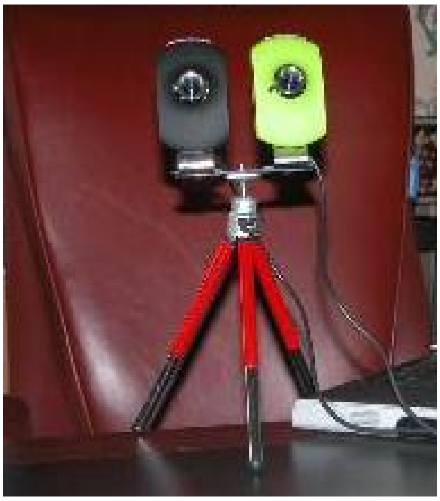


**Fig. 4.3 Stereo cameras**

## 3. DEPTH CALCULATIONS AND OCCLUSION USING A GPU

As discussed in earlier sections, calculating occlusions for AR in real time is still a challenging problem. Our study focused on developing a real-time method for calculating occlusions in a CAVE-based AR system, using GPU programming. The method includes four steps:

1. Find the areas in video-input images which might have occlusion relationships with virtual objects
2. Segment the potentially occluded areas and label them
3. Match the segmented areas
4. Calculate depth

Figure 4.4 shows the structure of the method. The method is implemented as several kernel functions, which are sequentially executed in several programmable fragment processors in different pipelines

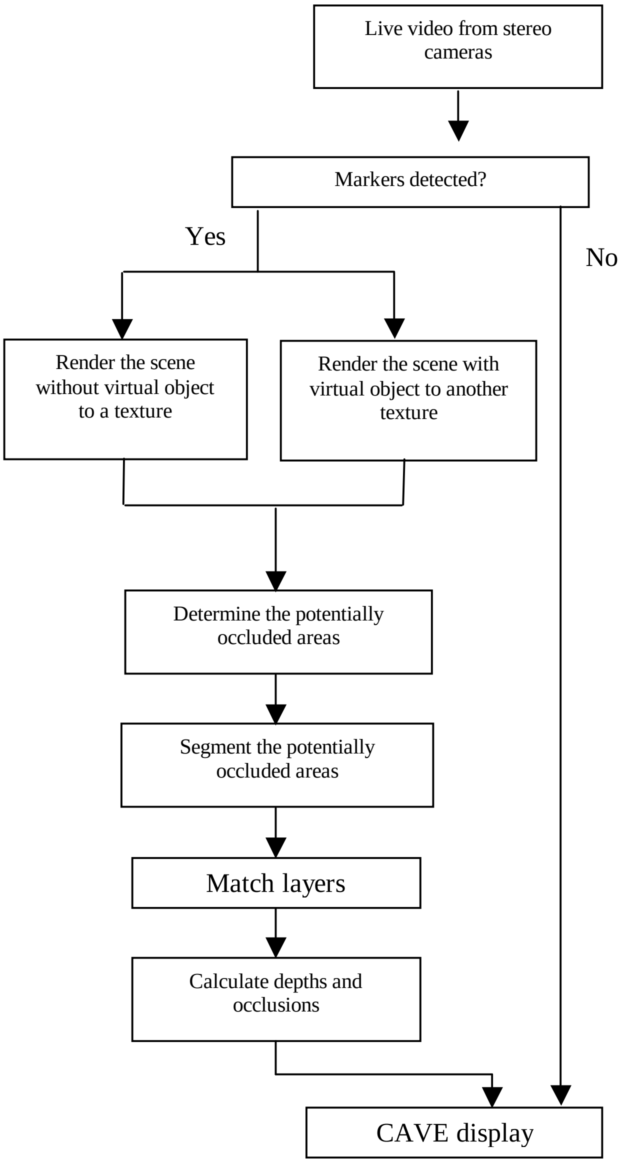

Live video from stereo cameras
Markers detected?
Yes
No
Render the scene without virtual object to a texture
Render the scene with virtual object to another texture
Determine the potentially occluded areas
Segment the potentially occluded areas
Match layers
Calculate depths and occlusions
CAVE display

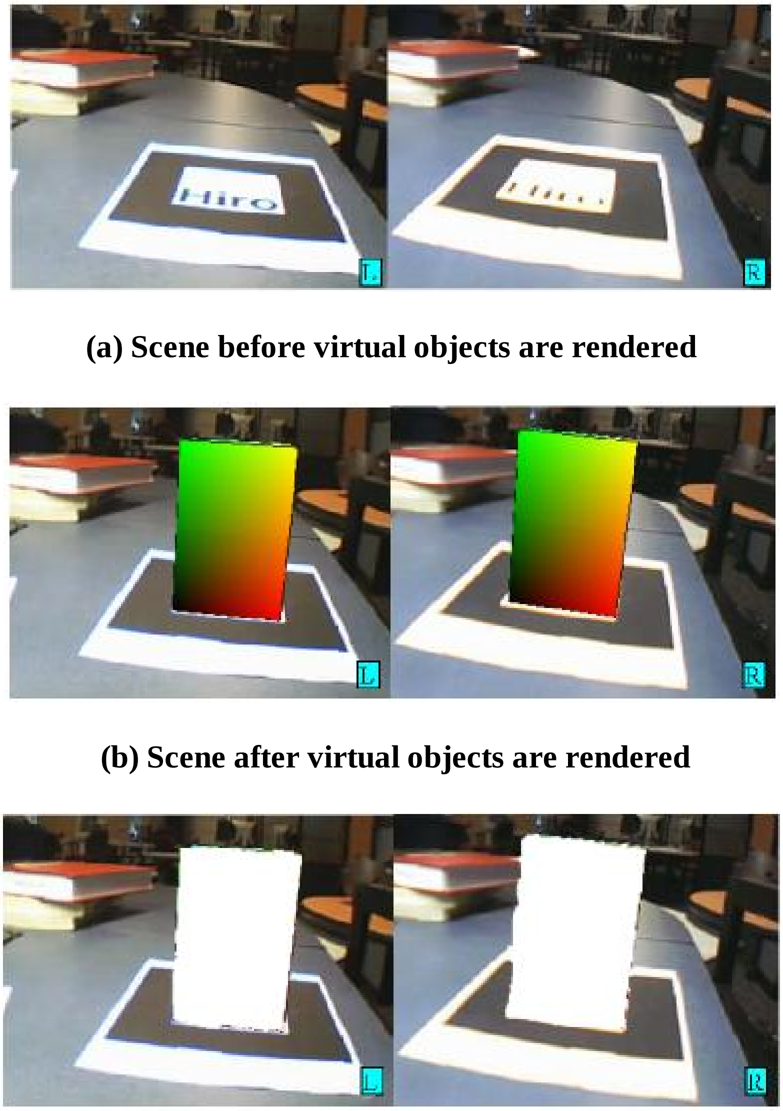
**Fig. 4.4 Method structure**

### 3.1 Preprocessing

The main purpose of the preprocessing step is to find potential areas within the video images that might have occlusion relationships with virtual objects. Only the areas in which virtual objects are placed might have occlusion relationships. In our method, we render the

AR scene twice, once before and once after virtual objects are added to the texture or frame buffers, as shown in figures 4.5 (a) and (b). In the figures, the left eye image and the right eye image are labeled with “L” and “R”, respectively. Next, a pixel-by-pixel comparison is made to find the areas covered by the virtual objects, as shown in figures 4.5 (c) and (d). Then, we segment objects and calculate depths, but only on the extracted potentially occluded areas.

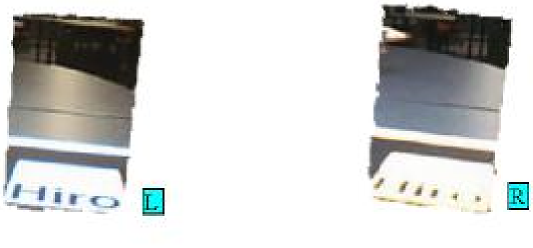
By excluding other areas, the method saves a substantial amount of computational time in subsequent steps. Since computation tasks are processed for each video frame, the method is also robust enough to work with dynamic backgrounds in which virtual objects and real objects move independently.



**(a) Scene before virtual objects are rendered**

**(b) Scene after virtual objects are rendered**

**(c) Potential occlusion areas shown in white**



**(d) Potential occlusion areas extracted**

**Fig. 4.5 Preprocessing**

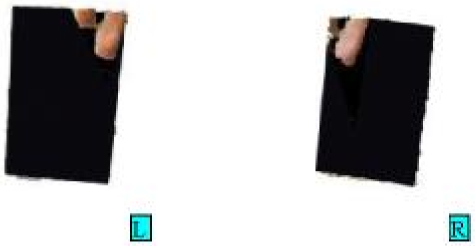


**Fig. 4.6 Segmentation**

### 3.2 Segmentation and Matching

After finding potentially occluded areas, we use a GPU-based segmentation algorithm to segment the potentially occluded areas for matching. The method assumes that a given potentially occluded area is composed of several layers with different depths. For example, in Fig.6, we assume that the depth of the finger and the depth of the background are different, so the finger is detected and segmented as one layer, while other objects are segmented as a background layer. In our example, the skin detection method of explicitly defining skin region described in the reference [21] is implemented in GPU as the segmentation algorithm. However the approach is open to other different segmentation algorithms.

After segmentation, we use a matching process for subsequent depth calculations and occlusion determination. Generally, matching is the most time-consuming part of depth calculations according to the result from OpenVIDIA stereo package [20]. However, in our method, matching is relatively fast because the potential occlusion areas have already been segmented and simplified into several layers, which greatly reduces matching computation time. Layers between two stereo images can be matched by using layer information, such as colors, horizontal positions, shapes, and sizes.

### 3.3 Depth calculations and occlusions

In our method for stereo AR systems, we use two video streams to capture left-eye images and right-eye images (see figure 4.5) Corresponding left-eye and right-eye images are arranged in a horizontal split. As a result, GPU calculations can be simplified by rendering the images once, rather than twice, using the same Fragment Shader. Depth information can then be inferred from the distance information between matched layers. For example, for the three objects with different scene depths, shown in figure 4.7, e1, e2, and e3 are their images on the two camera screens. By lining up the two screens, with the right camera image on the right and the left camera image on the left, and matching the corresponding points e1, e2, and e3 in the two images, the distances between image correspondences will be different if they are at different depths. In our approach, if the distance between the matched layers of a real object is smaller than the matched layers of a virtual object, the real object is in front of the virtual object (as shown in figure 4.8). Otherwise, the real object is behind the virtual object and is therefore occluded by the virtual object.

However, since the approach is based on layers, and each whole layer is assumed to be at the same depth, either in front of or behind the virtual object, mutual occlusions can not be handled.

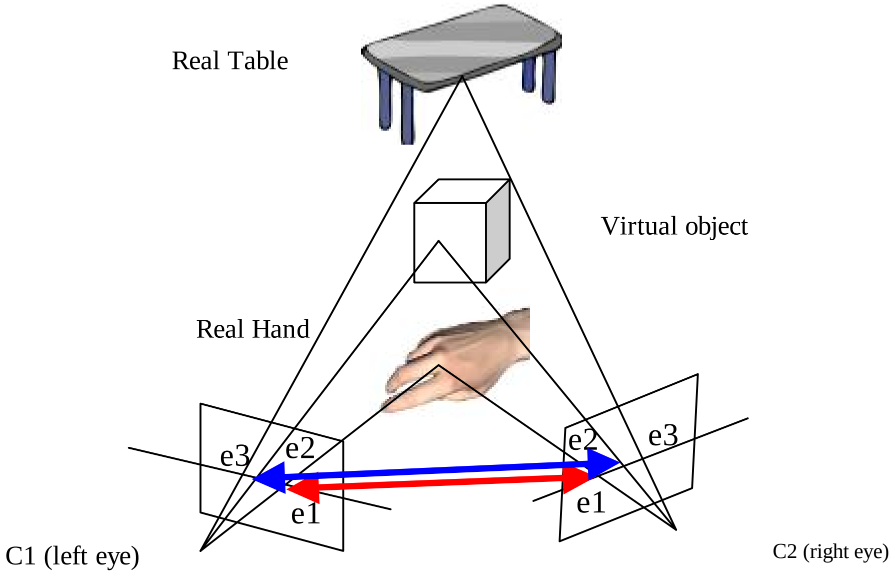


**Fig. 4.7 The relative positions of the real and virtual objects**

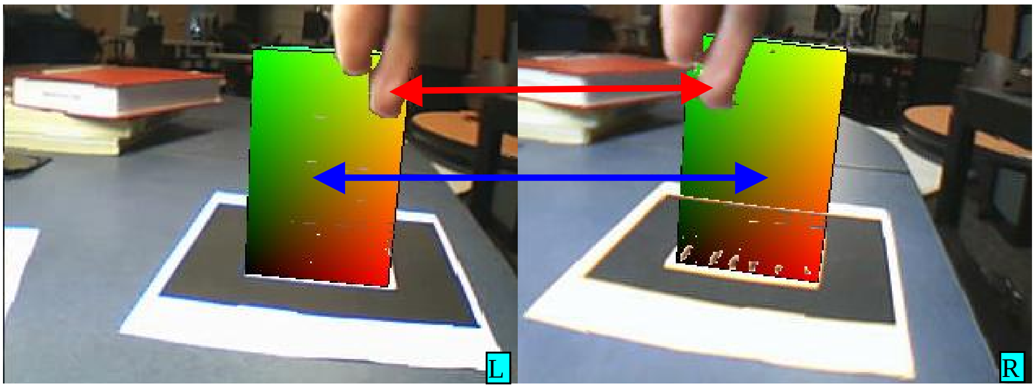


**Fig. 4.8 Occlusion**

### 3.4 Stream Processor and Reduction

As a known issue, although GPU-based calculations intrinsically run faster than corresponding CPU-based calculations, system design for the GPU-based method was somewhat more complex than a corresponding CPU-based method. In particular, the method had to be designed using stream procedures and fragment kernel functions, so that each kernel function could be executed on a GPU during real time AR scene rendering. Results from a given stream procedure can be exported to textures and buffers, which then hold input data needed for the following stream procedure calculation. Thus, the three most important characteristics used to evaluate a GPU algorithm are the complexity of each kernel function, the number of kernel functions, and the number of rendering operations.

Another known issue is about the parallel architecture of a GPU. With a GPU, multiple streaming processors can read from the same variable at the same time without causing problems. However, multiple processors cannot write to the same variable at the same time without causing problems. Thus, every kernel function can accept direct inputs, but must only output results to their corresponding texture or buffer units. Due to GPU architecture, it is not easy to calculate some global information, for example, sums or average colors for an entire texture.

In our method, the centre positions of segmented layers are needed for depth calculations. The position in this research is expressed as the weight centre. As shown in the average calculating equation (1), all pixel positions ($x_i$, $y_i$) must be summed and the number of pixels in each layer must be computed.

$$\begin{aligned} x &= (\sum_{i=1}^{n} x_i)/n \\ y &= (\sum_{i=1}^{n} y_i)/n \end{aligned} \quad n\text{: pixel number in the layer} \qquad (1)$$

As mentioned above, calculating average value with CPU is a very simple and basic operation, but is not a easy task with GPU for the reason mentioned above. To handle this known challenge of using a GPU, we chose to GPU reduction. Reduction is a graphics function that map large size textures into smaller size textures. A well-designed parallel reduction operation helps improve speed and also makes use of the parallel architecture of the GPU. The method iteratively reduces texture size until the whole texture is finally mapped to a 1 x 1 texture. For example, in every iteration, the kernel function computes the local sum of each 2 x 2 group and outputs a 1 x 1 group. So in each iteration, for the input texture of size

M x M, the output texture will become M/2 x M/2.

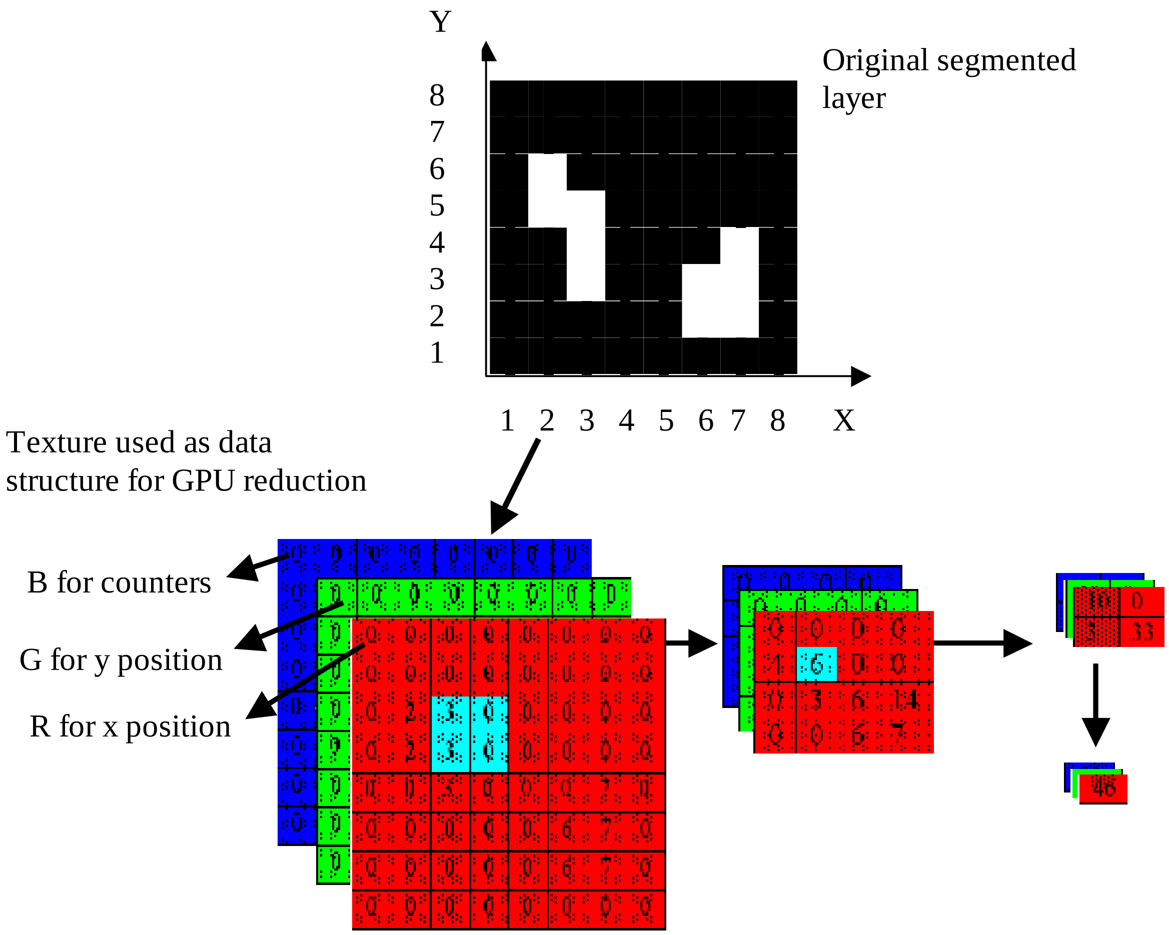
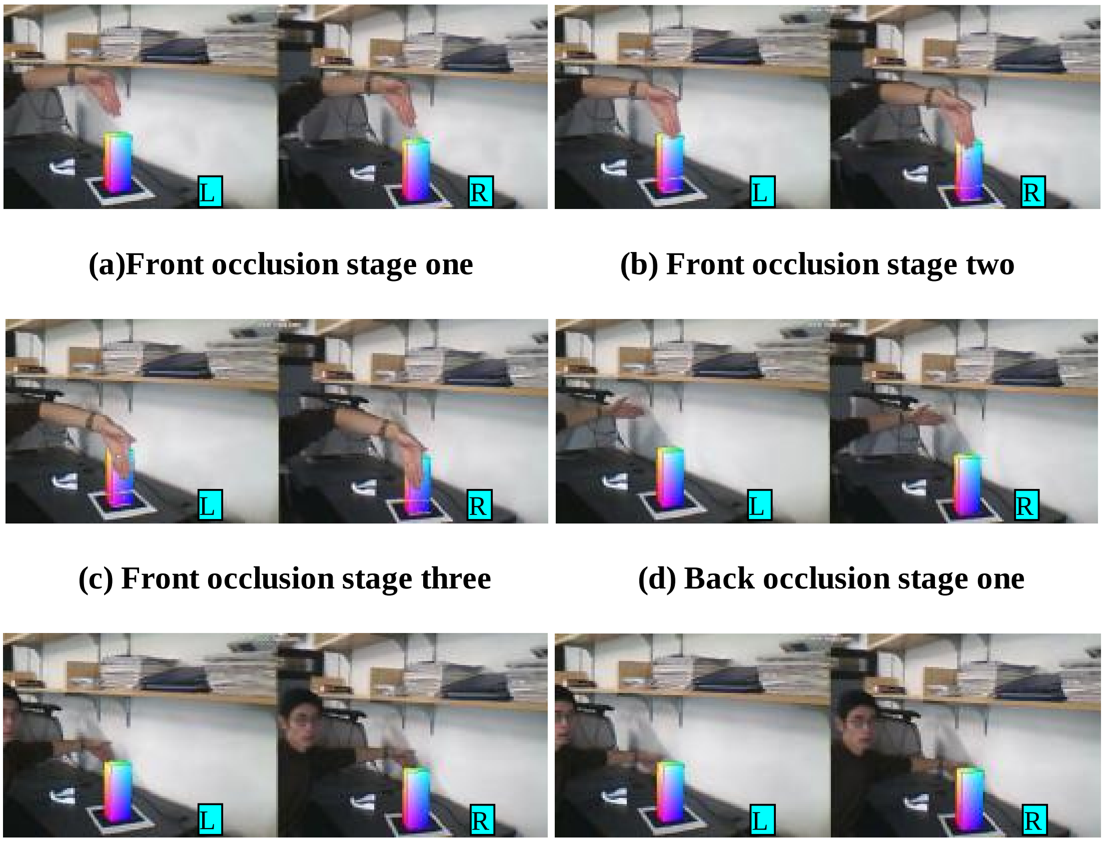


**Fig. 4.9 Examples of GPU reduction to calculate the centre of segmented layer**

If there are *n* matched layers in the stereo images, before conducting the reduction, *n* textures are used to store the layer position information, one for each matched layer. In each texture, the qualified pixel position $x_i$ and $y_i$ in each matched layer is initially stored in the GPU R and G channels, respectively, to replace the color information, and the value 1 is passed to the B channel for counting the number of pixels. Other unqualified pixels in the RGB channels are set to 0. Figure 4.9 summarizes the reduction process we use in our method, with an example of reducing an 8 x 8 input texture to a 1 x 1 texture with RGB channels. The texture on the left shows the initial input texture, with position information. The highlighted corresponding parts show the reduction kernel function for summing four values and outputting the result to the corresponding output texture. Each pixel in the second texture contains the local sum of a corresponding 2 x 2 region in the first texture. All the position information is summed in the last texture.

After finding the position values for a matched layer from the final texture, we use equation (1) to calculate the mean corresponding position of the segmented layer in the two eye images, which we then use to determine the relative depth of the real object with respect to the virtual object, and to decide which object to render. Next, the 1280 x 480 texture is reduced to a 2 x 1 texture that holds the position information for both the left-eye image and the right-eye image in each layer. Since the initial texture size is not a power of 2, 5 x 5, 4 x 3, and 2 x 2 reductions are used.



**(a)Front occlusion stage one** **(b) Front occlusion stage two**

**(c) Front occlusion stage three** **(d) Back occlusion stage one**

**(e) Back occlusion stage two** **(f) Back occlusion stage three**

**Fig. 4.10 Results for the real-time GPU-based occlusion method**

## 4 RESULTS

The series of images in figure 4.10 show experimental results for our real-time GPU occlusion method. The method was implemented on a PC with a Pentium(R) 4 2.0G CPU, 512 RAM, and an NIVIDIA GeForce 6800GS card. Figure 10 shows that the system can accurately determine relative depths between real and virtual objects and perform both front and back occlusion calculations in real time. Figure 4.11 shows an example with more complicated virtual objects.

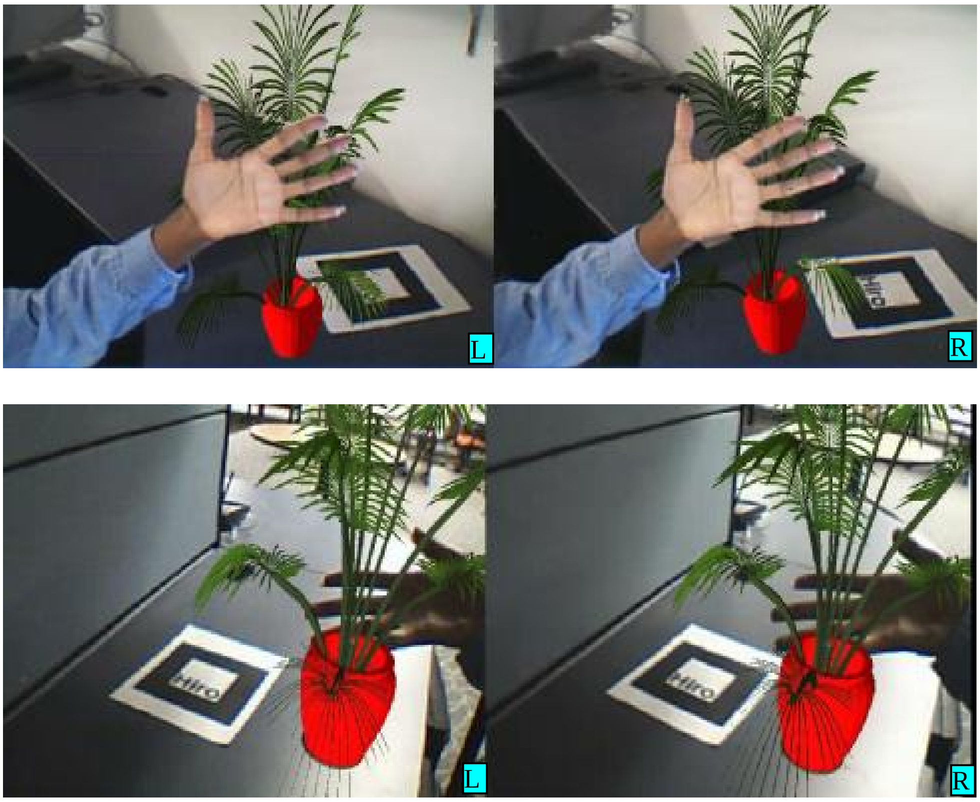


**Fig. 4.11 Occlusion with complicated geometry**

A performance comparison was conducted, on the same computer, without GPU processing, for different window sizes and applications. Results are shown in Table 4.1. The results show that the refresh rate for graphics only was 73.5 fps, while the refresh rate dropped greatly, to 8.5 fps, when real-time video backgrounds were processed and added. The frame rate drop is cause by resource consumption of the data acquisition and marker detection.

The same occlusion example was run on the GPU and the CPU. On the CPU, rendered textures were copied from video buffer to memory for depth and occlusion calculations. Then, the revised textures, with correct occlusions, were written back to the video buffer for display. From the Data in table 1, we can see that approximately GPU method is around (8.5-4.5)/(8.5-8.3) =20 times faster than the CPU method in our case. The results show that the performance of the GPU method was much better than the CPU method, which was also very sensitive to window size and resolution. There was only a very small drop in refresh rate when the GPU based occlusion method was added to the AR system. The experimental results also show that GPU-based occlusion was not affected by window size or resolution.

**Table 4.1 Performance Comparison**

| **Window Size**<br>**Application** | **Original size**<br>(1280*480) | **Full screen**<br>(1280*1024) |
|---|---|---|
| Graphics only | 73.5 fps | 73.5 fps |
| AR without occlusion | 8.5 fps | 8.5 fps |
| AR with CPU occlusion | 4.5 fps | 3 fps |
| AR with GPU occlusion | 8.3 fps | 8.3 fps |

## 5 DISCUSSION

In this study, we developed a GPU-based method for producing stereo AR in a CAVE-based display environment.  The innovative method can greatly improve visualization and usability of an AR system. Users can visually interact with virtual objects. A wide area of applications, e.g., tele-presence, tele-operation, and virtual prototyping, can benefit from the developed method with respect to both immersiveness and occlusion performance. System users can experience higher degrees of “immersiveness” and a greater sense of presence, than in prior AR systems, such as AR systems that use HMD displays.

Our GPU-based method solves the real-time occlusion problem that exists in prior AR systems. Our method introduces a new fast matching approach using layers. The method is not limited to use in a CAVE-based system. It can be used in any AR system that uses stereo video cameras. The fast matching approach depends upon stereo image information, so depth information cannot be determined for objects that only appear in a single camera. Since stereo AR images are the only requirement for the method, mobile applications using PDAs or cell phones can also use the method.

Since the approach is based on layers, each layer is assumed to be at the same depth, either in front of or behind the virtual object. The approach only computes occlusion for full layers. It cannot handle mutual occlusions between each certain layer and virtual object, for example part of the hand is behind the virtual object while others is in front of it, which is the limitation of this approach. The occlusion method also works for dynamic backgrounds, in which virtual objects and real objects move independently, since depth information is dynamically calculated for every frame.

Besides, as mention in the above section that, the frame rate drops a lot from 73.5 to 8.5 when video input and AR function is add, because of the data acquisition and marker tracking. So it will be interesting to implement image processing and marker detection algorithm of ARToolkit in GPU, which might improve the performance of AR.

## REFERENCES

[1] Sekuler, A.B., & Palmer, S.E., 1992, “Perception of Partly Occluded Objects: A Microgenetic Analysis,” Journal of Experimental Psychology: General, 121, pp. 95-111.

[2] Breen, D.E., Whitaker, R.T., Rose, E., & Tuceryan, M., 1996, “Interactive Occlusion and Automatic Object Placement for Augmented Reality”, Computer Graph. Forum 15(3), pp. 11-22.

[3] Gibson, S., Cook, J., Howard, T., Hubbold, R., & Oram, D., 2002, "Accurate camera calibration for off-line, video-based augmented reality," *presented at the Int. Symp. Mixed and Augmented Reality Darmstadt*, Germany, Sep. 2002.

[4] Lepetit, V., & Berger, M.O., 2000, "A Semi-Automatic Method for Resolving Occlusion in Augmented Reality," *CVPR 2000*, pp. 2225-2230.

[5] Schmidt, J., Niemann, H., & Vogt, S, 2002, "Dense Disparity Maps in Real-Time with an Application to Augmented Reality," *WACV 2002*, pp. 225-230.

[6] Duchesne, C., & Hervé, J.Y., 2000, "A Point-Based Approach to the Interposition Problem in Augmented Reality," *ICPR 2000*, pp. 1261-1265.

[7] Berger, M.O., 1997, "Resolving Occlusion in Augmented Reality: a Contour Based Approach without 3D Reconstruction," *CVPR 1997*, pp. 91-96.

[8] Kanbara, M., Okuma, T., Takemura, H., & Yokoya, N., 2000, "A Stereoscopic Video See-Through Augmented Reality System Based on Real-Time Vision-Based Registration," *VR 2000*, pp. 255-262.

[9] Kiyokawa, K., Kurata, Y. & Ohno, H, 2000, "An Optical See-through Display for Mutual Occlusion of Real and Virtual Environments," *Proceedings of IEEE & ACM ISAR 2000*, pp.60-67.

[10] Gordon, G.G., Billinghurst, M., Bell, M., Woodfill, J., Kowalik, B., Erendi, A., & Tilander, J., 2002, "The Use of Dense Stereo Range Data in Augmented Reality," *ISMAR 2002*, pp. 14-

[11] Purcell, T.J., Buck, I., Mark, W., & Hanrahan, P., 2002, "Ray Tracing on Programmable Graphics Hardware," *Proc. SIGGRAPH 2002*, pp. 703 - 712.

[12] Purcell T.J., 2004, "Ray Tracing on a Stream Processor", Ph.D. Thesis, Stanford University, March 2004.

[13] Szirmay-Kalos, L., Aszodi, B., Lazanyi, I., & Premecz, M., 2005, “Approximate ray-tracing on the GPU with distance impostors,” Computer Graphics Forum (Proc. of EG’05), 24(3).

[14] Billinghurst, M., 1999, "Interaction and Collaboration Techniques Using AR," *Web Proceedings of IWAR'99*.

[15] Bimber, O. & Frhlich, B., 2002, "Occlusion Shadows: Using Projected Light to Generate Realistic Occlusion Effects for View-Dependent Optical See-Through Displays," *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR'02)*.

[16] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B., 2001, “Recent Advances in Augmented Reality”, *IEEE Comp. Graph. & App*, vol. 21, no. 6 (Nov/Dec 2001), pp. 34-47.

[17] Raskar, R. & Low, K, 2001, “Interacting with Spatially Augmented Reality,” *ACM International Conference on Computer Graphics, Virtual Reality and Visualization*, Cape Town, South Africa , pp. 101-108.

[18] Steve, B., Zeltzer, D., Bolas, M.T., Chapelle, B., & Bennett, D., 1997, “The Future of Virtual Reality: Head Mounted Displays Versus Spatially Immersive Displays,” *SIGGRAPH 97 Conference Proceedings*, ACM SIGGRAPH, Addison-Wesley, pp. 485-486.

[19] Sinha, S., Frahm, J., Pollefeys, M. & Genc, Y., 2006, "GPU-based video feature tracking and matching," in *EDGE 2006, workshop on Edge Computing Using New Commodity Architectures*, Chapel Hill.

[20] Fung, J., & Mann, S., 2005, “Openvidia: parallel gpu computer vision”, *Proceedings of the 13th annual ACM international conference on Multimedia*, pp 849–852.

[21] Kovac, J., Peer, P., & Solina, F., 2003, “Human Skin Color. Clustering for Face Detection”, *Proc. of Int. Conf. on. Computer as a Tool*, Vol. 2.

# CHAPTER 5. CASE STUDY

## ---From Images to AR E-commerce and Immersive Rendering

This chapter describes a case study of AR e-commerce and related technologies. A 3D model is easily recovered by using the technology presented in chapter 3, which is used in the AR e-commerce web system and Immersive AR environment.

In this scenario, I want to sell my sofa using AR e-commerce, but right now I don't have a 3D model of my sofa, and I also don't want to build one because I am not a modeling expert. Making 3D models of the existing object is very time consuming, especially the step of texture mapping (Textures of my sofa should be used instead of general textures to show the exact status of my sofa). Thus, I decide to recover it directly from images. Two photos (shown as figure 5.1) were taken using a Sony Cyber-shot DSC p92. No other parameter information about the camera is known.

To make the stereo matching processing easier and more accurate, color tags were attached to the key positions of the sofa, as shown in figure 5.1. These tags helped to decrease the ambiguity of key features in different viewpoints, which might greatly impact a lot to the recovered result. Where to put those color tags greatly depends on how person decomposes the object into different faces as a model. In the following steps, these tagged points are picked and matched in two images to calculate and reconstruct the model, which only takes several minutes.

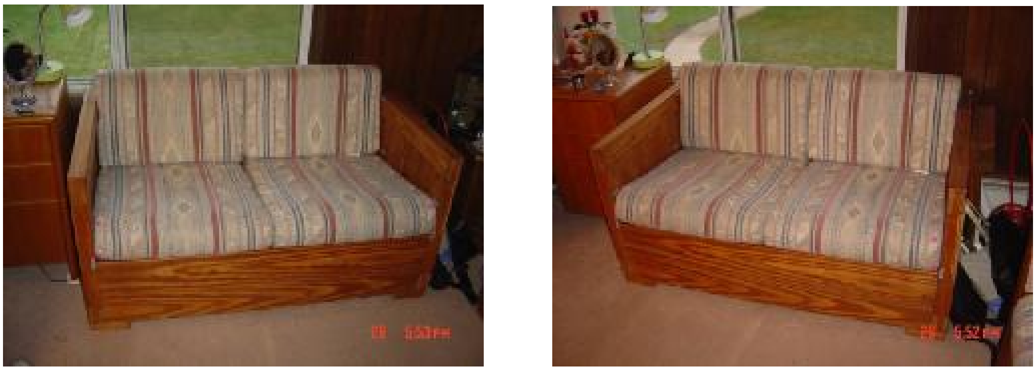


Figure 5.1 2D images of the sofa

Figure 5.2 shows the correspondence of key features of the sofa based on the color tags. Figure 5.3 shows the triangles used to reconstruct the model. With the feature point’s correspondences, triangles, and estimated camera intrinsic parameter, the 3D position of the features is calculated and the 3D model is recovered.

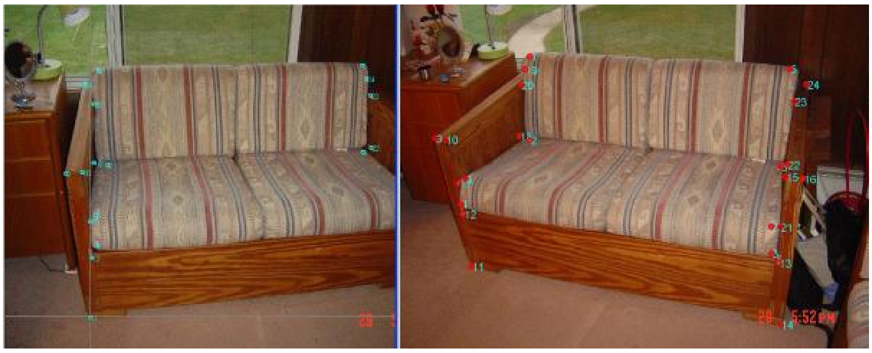


Figure 5.2 Correspondences of features of the sofa

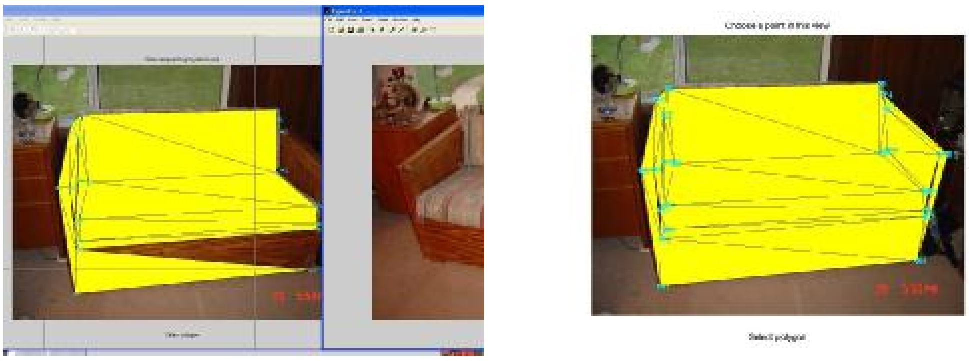


Figure 5.3 Selecting triangles for the reconstructed sofa

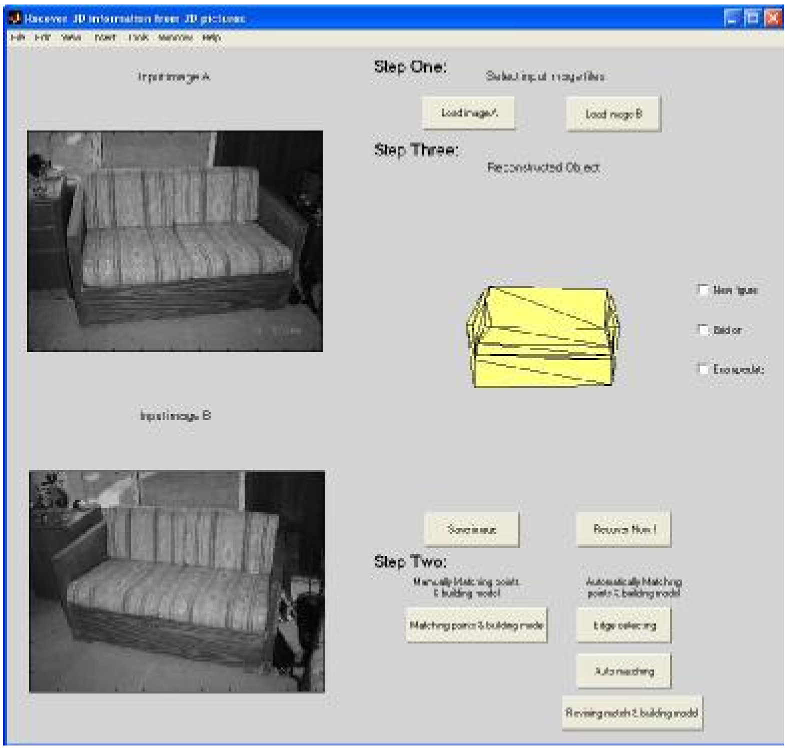
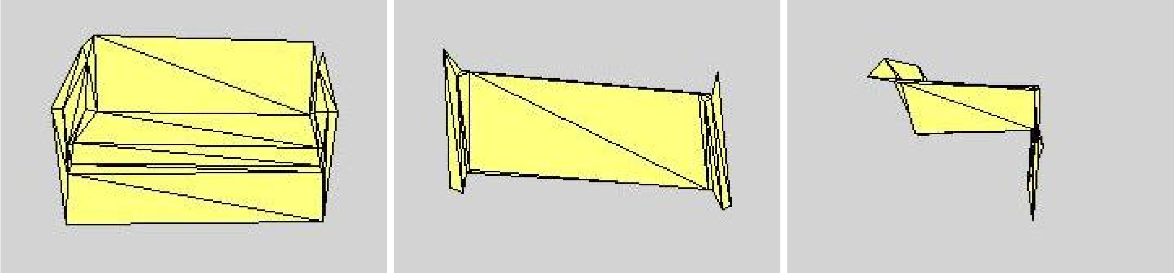

Recover 3D information from 2D pictures
File Edit View Insert Tools Window Help
Input image A
Step One:
Select input image files
Load image A
Load image B
Step Three:
Reconstructed Object
□ New figure
□ Grid on
□ Exteriors
Input image B
Save image
Recover Now !
Step Two:
Manually Matching points & building model
Automatically Matching points & building model
Matching points & building model
Edge selecting
Auto matching
Revising match & building model

**Figure 5.4 Recovery interface**

Figure 5.4 shows the recovery interface and procedures. Considering that the work required to revise mismatches from the automatic matching process is similar to that of picking correspondences manually, I just complete the matching manually in this case. Figure 5.5 shows the recovered model of the sofa from three different views, from which it is demonstrated that we can see the accuracy of the recovered model is acceptable.



**(a) Front view** **(b) Top view** **(c) Side view**

**Figure 5.5 Recovered sofa in three views**

The recovered 3D model is saved as a standard VRML file with texture mapping, which could be easily imported to almost any CAD software for editing and rendering. Figure 5.6 shows the recovered sofa with textures.

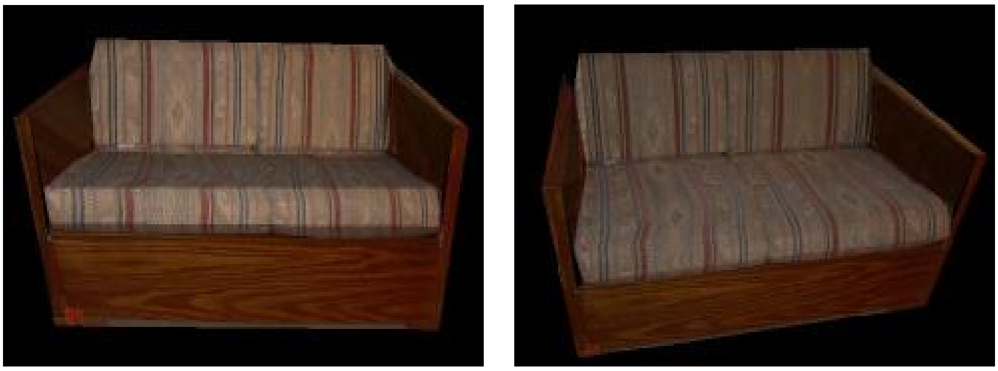


**Figure 5.6 Recovered sofa with textures**

Now the recovered 3D sofa is prepared to use in AR e-commerce. Before adding it into an AR e-commerce database, the step of normalization should be taken, so that what online shoppers see using AR e-commerce is the model in actual size and in the right direction. This process is required because the AR rendering greatly depends on the marker size and marker system used. Figure 5.7 shows the normalization process (zoom, translation, rotation, and adding light source) using 3DS Max.

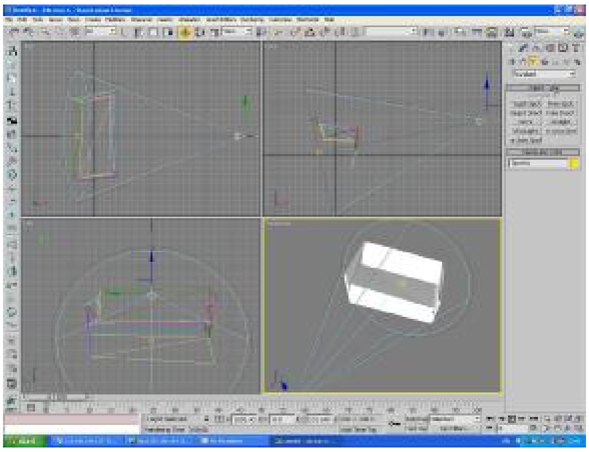
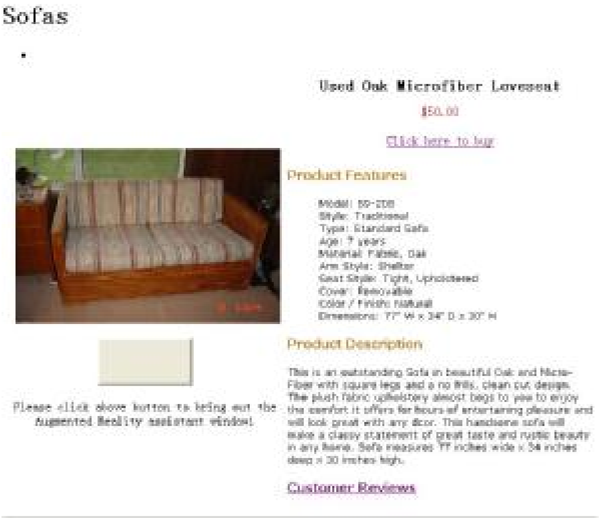


**Figure 5.7 Normalization of the recovered sofa**

After processing, the normalized 3D sofa is added to the database in the server together with other introductory information. The AR e-commerce web page for this second-hand sofa is prepared as shown in figure 5.8. With the web page and the database server running, online shoppers are able to use it and “visually” bring this sofa into their own home to see whether it is a good fit, all they need is a laptop, markers and a web camera. Figure 5.9 shows online shoppers "trying" the sofa in a sitting room.



Sofas

Used Oak Microfiber Loveseat

$50.00

Click here to buy

Product Features

Model: 59-208
Style: Traditional
Type: Standard Sofa
Age: 7 years
Material: Fabric, Oak
Arm Style: Shelter
Seat Style: Tight, Upholstered
Cover: Removable
Color / Finish: Natural
Dimensions: 77" W x 34" D x 30" H

Product Description

This is an outstanding Sofa in beautiful Oak and Micro-Fiber with square legs and a no frills, clean cut design. The plush fabric upholstery almost begs to you to enjoy the comfort it offers for hours of entertaining pleasure and will look great with any decor. This handsome sofa will make a classy statement of great taste and rustic beauty in any home. Sofa measures 77 inches wide x 34 inches deep x 30 inches high.

Customer Reviews

Please click above button to bring out the Augmented Reality assistant window!

**Figure 5.8 AR e-commerce web page for the used sofa**

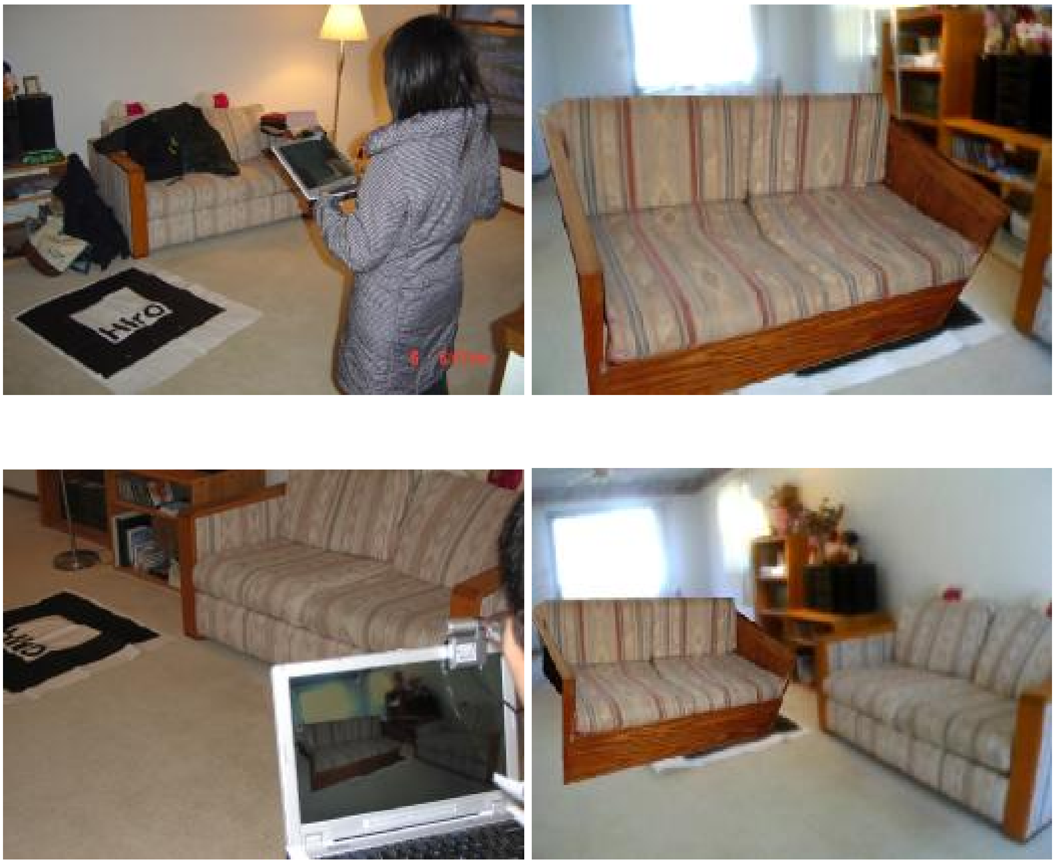


**Figure 5.9 "Trying" the sofa with AR e-commerce**

AR e-commerce is a very useful and interesting technology that provides the capability to see how products would fit in an actual space. Before purchasing, users can naturally see what happens in their own environment, so they can instantly tell whether or not any product would a good fit. AR e-commerce is a tool that can make online shoppers more confident.

However, for some remote users who are not in the actual environment, the presented AR e-commerce might not have great advantage. For example, if a household member who is not at home also wants to participate in the decision of buying sofas for the home, the shared AR scenario might work to some extent but would not be the best choice. In this situation, the immersive AR environment would provide better remote presence. With the immersive environment, users can have the immersive and stereo view of the remote scenario, which would give a much better sense of “being there”. The intelligent GPU-based occlusion method makes the virtual part of the AR scene even more real with the correct occlusion and depth perception.

Figure 5.10 shows an example of using immersive AR environment for remote presence, which helps for the decision-making about the sofa.

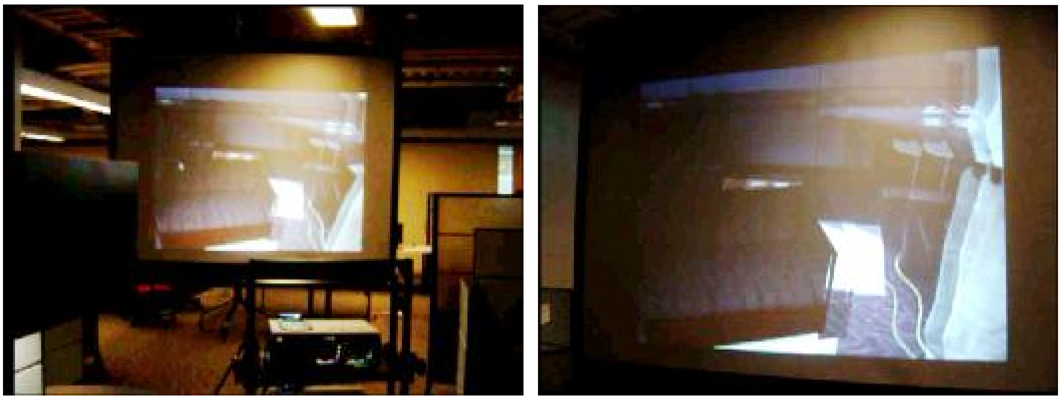


**Figure 5.10 Using an immersive AR e-commerce environment for remote presence**

# CHAPTER 6. CONCLUSIONS

Traditional e-commerce systems have reached a limitation that need to be overcome, because they do not provide enough intuitive information for online shoppers to help online shoppers with their decision making, especially for products like furniture, clothing, shoes, and jewelry. This research studied the feasibility of using AR to enhance e-commerce and provide more direct information to online shoppers, users' feedback about the AR e-commerce, and two other related technologies. One technology we presented in this dissertation, which can directly recover a 3D model from 2D pictures, can reduce the cost of model making and make AR e-commerce more easily adopted by online stores. The other technology, immersive AR with GPU-based occlusion, provides remote users with a better feeling of presence by offering immersive and stereo view.

In Chapter 2, We presented the design and development of an AR e-commerce prototype, which helps online shoppers "visually" bring the product into their own environment in order to make them more confident in their online shopping. Several key issues for AR e-commerce, were analyzed and discussed, including the creation of realistic product models, normalization, real time rendering, and seamless model merging into real-world scenes. A usability experiment was also designed and conducted, to compare AR e-commerce with traditional e-commerce and VR e-commerce, and to study the effectiveness of AR for enhancing e-commerce. Usability experiments results verified that the developed AR e-commerce system could be used to provide more direct product information to online shoppers and thereby help them make better purchasing decisions. But as mentioned by participants, this strength only works for certain products (for example decoration related products) instead of all products. It is also evident that some limitations still exist in the proposed approach. According to the study participants, the major limitation of using the AR e-commerce system is that it is currently not as easy to use as the traditional or VR e-commerce systems, because more computer devices and higher computer experience level are required to use this new type of e-commerce system. The other possible reason participant thought AR e-commerce is not as easy to use as the other two types of e-commerce is that user are still not familiar with Augmented Reality and its interaction method, which makes the easiness to use even worse. To improve the easiness to use, new Augmented Reality interaction methods still need to be studied and improved, to make it more convenient and intuitively friendly for users without too much computer experience. Also the interface could be designed as convenient as possible for users. For example, online shopper could have multiple choices to use this system like upload static picture with markers, or videos pre-made so that users do not have to take the computer around for each product every time. The application could also be specifically implemented on PDA and cell phone, which is available for most consumers and is also light to carry, because some participants complained that the laptop is too heavy to carry around all the time. The rendering methods used also need to be improved to help integrate virtual models into real scenes more seamlessly. For example, large amount of texture mapping should be used to improve the realness of the virtual product. Real time occlusion could also be implemented to help with consumers' depth perception about the virtual product. Especially, the stableness of computer vision algorithm used in the AR system needs to be improved, to make the marker tracking more stable, even in a poor lighting condition, so that the virtual object will not disappear from time to time which obviously interrupts the presentation of augmented scene and immersion. New smarter algorithm should also be studied and developed for partial marker tracking so that user do not need to worry about that the virtual product will disappear because the marker is partially occluded. And even though wireless internet access is currently still not fast enough to transfer high-resolution product models in real-time, it might not be a problem in the near future. However, in a word, users of this study preferred the AR e-commerce system more than traditional e-commerce and VR e-commerce systems by considering its strength and weakness, which means the great potential of AR e-commerce.

Chapter 3 described the comprehensive tool to recover 3D objects from 2D photos with wide baselines. This tool integrates automatic feature extraction, automatic feature matching, manual revision, feature recovery, and model reconstruction technologies to conveniently recover 3D models directly from 2D photos. The method and tool were demonstrated using practical case studies, and each was shown to be an effective and convenient means. Specifically a hierarchical matching algorithm, a universal camera intrinsic matrix estimation technique, and a new automatic texture-mapping algorithm were presented as main noteworthy and novel contributions of this part of research. However since the proposed method was only designed for recovering 3D scenes from two 2D photos, non-visible areas in either photo cannot be recovered, which means that the reconstructed model is not a complete object. So fusing technology should also be used to merge different reconstructed parts into a complete object. In this case, four photos could be used to recover a complete 3D object (two photos for up front part, and the other two photos for down back part). Currently, since the automatic feature extracting method is based on simple segment edges (like buildings), features of objects with complicated curve edges (like trees and grass) will be treated as noise and filtered out because of the length. So more human interaction are needed to be involved in such situation. But color tags could be used in these complicated objects to reduce the human interaction. Also even though the hierarchical matching method performs better than cross correlation method, it takes more time for calculation. In this part of research, the use of estimated method of camera intrinsic parameter bring the convenience to users, so that they do not have to conduct the calibration to know the camera parameter in each situation. However the precision of the recovered 3D model using estimated intrinsic camera parameter matrix still need improving. A good solution is to use optimization technology to search for the real focal length with the aim of minimizing the error, which could be measured by re-projecting the recovered 3D point back to the photos. Even though there is weakness to be improved, the presented recovery method is a very potential and convenient modeling approach. The strength of this method is even more obvious for specific modeling for certain object.

In Chapter 4, We proposed an immersive AR prototype system which makes use of the characteristics of CAVE-based virtual reality to greatly improve visualization and usability of an AR application. The system integrates virtual objects into a real scene captured by a set of stereo remote cameras. As the first step, we implemented the system on a one-wall portable Cave at this moment. More cameras with 90-degree viewpoint or a 360-degree camera mounted on a controllable robot are needed to implement this system on a six-wall surrounded CAVE. However this type of Augmented Reality is some kind of changing the way of augmentation, and user will have more feeling of “immersive” instead of “see-through”. And unlike Immersive Virtual Reality, users with this kind of facility act more like to an “observer” instead of a “participant”. But users still have the interaction to virtual objects and the robot mounted with cameras to move and operate. In this part of research, a GPU-based method for computing occlusion was also developed between real and virtual objects, in real time. This method uses information from captured stereo images to determine depths of objects in the real scene. Our method introduces a new fast matching approach using layers. Each of real objects is segmented as a layer. Each layer is assumed to be at the same depth, either in front of or behind the virtual object, which is also treated as a layer with same depth. So the accuracy of occlusion results is also slightly affected by viewpoints. For future studies, more experiment could be conducted to determine the error distance. Also the approach only computes occlusion for full layers. It cannot handle mutual occlusions between each certain layer and virtual object, for example part of the hand is behind the virtual object while others is in front of it, which is the limitation of this approach. However this method is not limited to use in a CAVE-based system only. It can be used in any AR system that uses stereo video cameras including mobile applications using PDAs or cell phones. Since the fast matching approach depends upon stereo image information, so depth information cannot be determined for objects that only appear in a single camera. The occlusion method also works for dynamic backgrounds, in which virtual objects and real objects move independently, since depth information is dynamically calculated for every frame. Results and performance comparisons show that the GPU-based method is much faster than prior CPU-based methods. The proposed immersive AR system could also be used in applications involving tele-presence (e.g., tele-operation, tele-training, tele-maintenance, and tele-tours) so that users could experience higher degrees of “immersion” and greater sense of presence, feelings of “being there”. A wide area of applications can benefit from the developed method with respect to both immersiveness and occlusion performance. Besides, as mention in the chapter 4, the frame rate drops a lot from 73.5 to 8.5 when video input and AR function is add, because of the data acquisition and marker tracking. So it will be interesting to implement image processing and marker detection algorithm of ARToolkit in GPU, which might improve the performance of AR.

In Chapter 5, A case study is used to show how all three of presented technologies work together. The case study depicts a scenario how a person sells his furniture on ebay.com using AR e-commerce. The 3D model of the specific sofa is recovered with textures using the tool developed in chapter 3 and processed in totally several minutes. With the plug-in built in the web pages, other buyers could connect to this page to try and see how this specific sofa fit their home, which helps a lot to make their decision. Even though shoppers might not be more willing to buy products because of this new type of e-commerce, more shoppers will come to this page because it brings more shopping confidence, which definitely improves the chance of products to be sold.  The case study helps to understand the effectiveness and potential of AR e-commerce and related technologies.

In conclusion, this dissertation focuses on the study and development of AR e-commerce and related techniques. It links AR e-commerce to model recovery tools with cost consideration from the sellers' perspective, and also to immersive AR environments with the consideration of remote users. Taken together, these studies demonstrate an innovative and powerful approach that can enhance e-commerce and online shopping and could also benefit other related industries of modeling, photo geometry, and Augmented Reality, etc. For example, IKEA, one of largest furniture stores, might find AR e-commerce and new modeling method presented very interesting and helpful for their online business. And Leica PhotoGeometry Suite product, which is one of most powerful tools to reconstruct 3D terrains from 2D remote sensing images, might find the GPU method presented helpful to improve their performance of image processing and finding correspondences from image pairs.

# APPENDIX A. E-COMMERCE APPLICATIONS COMPARISON

Please take a few minutes to complete this survey by checking the box before the suitable selection and answering the questions. Thank you for your cooperation.

## Demographic Questions:

1. Gender ______

□ Female □ Male

2. What is your age? ______

□ 17—21 □ 22-26 □ 27-31 □ 32-36
□ More than 36

3. What is your education level? ______

□ High School □ Bachelor □ Master □ PH.D
□ Others

4. How much computer experience do you have? ______

□ Not At All □ A little □ Average □ A lot
□ Professional level experience

5. Do you shop online? ______

□ Not At All □ A little □ Average □ Very Often
□ More than often

## Information Questions:

6. Do you want to buy any product for the current scene?____________________
Please list you reasons? ______________________________________
______________________________________________________

## Comparison Questions:

7. Please rate your overall satisfaction for the three e-Commerce web pages on a scale from lowest (1) to highest (5)?

Traditional ____/5 VR enhanced ____/5 AR enhanced ____/5

8. Please evaluate the product information (e.g., color, size, fit your office or not, etc) provided by the three e-Commerce web pages on a scale from lowest (1) to highest (5)?

Traditional ____/5 VR enhanced ____/5 AR enhanced ____/5

9. Please evaluate the easiness to use of the three e-Commerce web pages on a scale from lowest (1) to highest (5)?

Traditional ____/5 VR enhanced ____/5 AR enhanced ____/5

10. Please rate your confidence level for your decision (e.g., buy or not), based on the information given by each of the three e-Commerce web pages, on a scale from lowest (1) to highest (5)?

Traditional ____/5 VR enhanced ____/5 AR enhanced ____/5

11. Which one do you like the most as an e-Commerce web page? ______

□ Traditional □ VR enhanced □ AR enhanced

12. What do you think might be the problem that prevents AR enhanced e-Commerce from being widely used?

____________________________________________________________

____________________________________________________________

____________________________________________________

13. What advantage do you think that the AR enhanced e-Commerce has, compared to the other two e-Commerce?

____________________________________________________________

____________________________________________________________

____________________________________________________

14. What disadvantage do you think that the AR enhanced e-Commerce has, compared to the other two e-Commerce?

____________________________________________________________

____________________________________________________________

___________________________________________________

15. Do you think that AR enhanced e-Commerce has the potential to replace onsite shopping in the future? Why?

____________________________________________________________

____________________________________________________________

____________________________________________________

## Suggestion:

16. Do you have any other suggestions about our AR enhanced e-Commerce web page? ( such as design, interaction, easy to use, devices)

17. Any final observations? Please write down some of your thought and feeling.

# ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this dissertation. First, I would like to thank my major professor, Professor Shana Smith, for the opportunity she gave me to work on these topics and for her enthusiasm and guidance throughout these topics.

I also wish to thank my committee members for their efforts and contributions to this work, and for their countless advices and helps in the process: Professor Frederick O. Lorenz, Professor Derrick J. Parkhurst, Professor Viren R. Amin, and Professor Julie A. Dickerson.

I am especially grateful for my spouse Chengyan Yue for her constant support, encouraging me to complete this program.