Osaka University Knowledge Archive

Title	Design and Evaluation of AR-based Guiding Techniques for Learning and Training Applications
Author(s)	Techasarntikul, Nattaon
Citation	大阪大学, 2020, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/76646
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

Design and Evaluation of AR-based Guiding Techniques for Learning and Training Applications

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2020

Nattaon TECHASARNTIKUL

Dissertation

Title

Design and Evaluation of AR-based Guiding Techniques for Learning and Training Applications

Supervising Professor

Professor Haruo Takemura

Author

Nattaon Techasarntikul

January 14, 2020

Department of Information Systems Engineering
Graduate School of Information Science and Technology
Osaka University

List of Publications

Journals

1. Techasarntikul, N., Ratsamee, P., Orlosky, J., Mashita, T., Uranishi, Y., Kiyokawa, K., and Takemura, H. A Guidance and Visualization of Optimized Packing Solutions. In Journal of Information Processing, Vol.28, No.1, 2020. (accepted)
2. Techasarntikul, N., Ratsamee, P., Orlosky, J., Mashita, T., Uranishi, Y., Kiyokawa, K., and Takemura, H. The Effect of the Presence of an Embodied Agent in an AR Guiding System. In Transactions of the Virtual Reality Society of Japan, Vol.25, No.1, 2020. (accepted)

International Conferences

1. Techasarntikul, N., Ratsamee, P., Orlosky, J., Mashita, T., Uranishi, Y., Kiyokawa, K., and Takemura, H. Evaluation of Embodied Agent Positioning and Moving Interfaces for AR Virtual Guide. In Proceedings of 29th International Conference on Artificial Reality and Telexistence 24th Eurographics Symposium on Virtual Environments (ICAT-EGVE), Sep., 2019.
2. Techasarntikul, N., Ratsamee, P., Orlosky, J., Mashita, T., Uranishi, Y., Kiyokawa, K., and Takemura, H. Evaluation of Pointing Interfaces with an AR Agent for Multi-section Information Guidance. In Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE-VR), pp. 1185–1186, Mar., 2019. https://doi.org/10.1109/vr.2019.8798061

Domestic Workshop

1. Techasarathul, N., Mashita, T., Ratsamee, P., Uranishi, Y., Kiyokawa, K. and Takemura, H.: A Projection-based Packing Support System (in Japanese). In IEICE technical report, Vol. 116, No. 245, pp. 63–67, Oct., 2016. https://ci.nii.ac.jp/naid/40020992271/en/

Abstract

This dissertation presents the design and evaluation of several user interfaces used for guiding a human through a task with Augmented Reality (AR) techniques. I explore AR interfaces for guiding information used in scenarios of learning and working, i.e. in a museum and packing task, respectively. These interfaces were analyzed in both objective and subjective terms, including task completion time, task performance, memory performance, and user preference.

In the first part of this dissertation, I explore pointing interfaces used in AR information guidance of a large piece of artwork in a museum. An Embodied Conversational Agent (ECA) is included in the system implementation as a virtual guide, which is expected to positively impact user preference for the system. I also evaluated three pointing interfaces for the virtual guide. The results revealed that a line laser interface can help users find the position of related information the fastest while its obstructiveness was not much of a concern. There is also a trend that the fewer attempts taken by the user for seeking out an area of the content (the interface exactly points out a position), the less information that is recognized by the user.

In the second part of this dissertation, I explore positioning and movement interfaces for the virtual guide.

For a large-scale artwork, the distance from one position in the artwork to another is large, which means the virtual guide needs longer to reach its destination. Therefore, I propose the addition of a teleportation interface, which would decrease the time necessary for the guide to move and consequently the waiting time of the user when compared to the normal flying interface. I also take the virtual guide's position into consideration because when the virtual guide is outside of the artwork area, it can be out of the sight of the user, which leads to the question of how the user prefers the virtual guide to be in the environment. Results from a user study indicated that the flying interface has a positive rating relative to the teleportation interface because it is easy for the eyes to track. Although the teleportation interface helps reduce guiding time, its disappearing and reappearing characteristic makes the user feel uncomfortable with the guide. Regarding the virtual guide positioning, both

inside and outside positioning interfaces were preferred roughly equally by the users.

In the third part of this dissertation, I investigate the effect of the presence of an ECA virtual guide. The study was designed to test the effect of an ECA as a virtual guide in combination with five different guiding interfaces: audio, arrow, circle, arrow-move, and circle-move. Results of the evaluation suggested that having the virtual guide with any guiding interface can significantly increase either attractiveness, stimulation, or novelty for the guiding system.

In the forth part of this dissertation, I propose AR interfaces for a workplace scenario. For this purpose, I built an automatic solution suggestion system and solution indication interfaces to assist with object packing tasks. Evaluation results showed that a system with both rotation and movement indications can significantly reduce time and hand movements in packing tasks compared to packing without such support. However, there is a trade-off between these two interfaces in that the rotation type helps users complete a task faster, while the movement type is easier to understand and follow.

This dissertation concludes with a summary of findings and directions for future improvement of AR visualization, which contributes to related design guidelines in the field of human computer interaction.

Acknowledgements

I would like to thank my supervisor, Prof. Haruo Takemura and my second supervisor Prof. Kiyoshi Kiyokawa for their helpful comments and supports during my studies. Also to Prof. Tatsuhiro Tsuchiya for reviewing and commenting this dissertation.

Also I would like to thank Assoc. Prof. Tomohiro Mashita, Assoc. Prof. Yuki Uranishi, Assoc. Prof. Jason Orlosky, and specially to Assist. Prof. Photchara Ratsamee who always mentor me.

I also thank all members of Takemura laboratory including OB and OG for talking, helping, commenting, and friendship. Also, thank you to Secretary Sakaki-san and Takahashi-san for helping me with various Japanese documents. After five and a half years in my lab, I have met so many people inside and outside the lab through conferences. I have learned many things and enjoyed my life, and I would like to continue to do so after my graduation.

Lastly, thank you for my family who always supports, encourages, and are with me for all this time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Augmented Reality (AR) is a technique for overlaying virtual information to the real world. A typical AR system consists of the following components: a camera used to observe an environment, an AR display, information (image or model) of the object to be tracked, a virtual content to be displayed, and an object tracking module.

Cameras are not the only sensors used for tracking in AR. Other sensors such as mechanical [1], magnetic [2], and ultrasonic sensors [3] are also used to track the pose of an object. However, due to availability of high quality cameras today, they are more popular than other sensors. There are two major techniques for tracking an object using an RGB camera: marker based [4] and natural feature tracking [5]. While marker based tracking looks for an artificial marker that is easily seen by both the camera and human, natural feature tracking looks for features such as dominant corners or edges in the image, which requires more processing power.

AR displays are available in many types as shown in Figure 1.1. They can be a head-up display that is normally built-in a vehicle's front window glass, a head-

Figure 1.1: AR displays: head-up display, head-mounted display, handheld display, stationery display, and projected display (images taken from [10–14]).

mounted display (HMD) that allows a user to walk around, a handheld display such as a smartphone or tablet, which has a compact form factor and affordable price, a stationery display such as a monitor scene, or a projection display that can cover a large area. Additionally, AR displays are not limited to the visual displays; other outputs such as Audio [6], Haptic [7], Olfactory [8], and Gustatory displays [9] are also able to create AR experiences by augmenting the human senses of hearing, touch, smell, and taste.

Since vision is the most important sense for humans [15], vision-based AR system is a mainstream in AR research that aims to provide supplementary information to help guiding a user in many scenarios. Most of the real world tasks in navigation, learning, training and maintenance scenarios usually have a pre-planned procedure. However, some tasks have more than one solution. In such cases, the procedures such as visiting paths, assembly, picking and packing orders may be altered.

AR that automatically suggests a solution and overlays a procedure of a task could help improve work efficiency by reducing the time required for the task since the worker does not need to recall, review, or think about the next step of the procedure.

However, in order to show the procedure of the task in AR effectively, a design for visualizing that information and its evaluation are important. Although nearly 1,500 articles of AR studies were published in 2014, only 10% of them conducted user studies [16]. Also, many of them only make a comparison of the system between AR and traditional (without AR e.g. a paper manual), or only a comparison of the interfaces without simulating real scenario usage.

1.2 Research Objectives and Challenges

In this research, I am interested in designing AR visualizations for guiding a user to perform or navigate prevalent tasks that can be found in learning and training scenarios.

In learning scenarios, I chose to design an AR system to help guide in a museum because such a place provides many exhibits to be explored by the visitors. However, detailed information of each exhibit typically provided in audio, printed text, or AR-based text descriptions is not able to attract the interest of a visitor for a long period like a human guide can. Therefore, I aim to develop the AR guide system by including a virtual guide to increase attractiveness on both the exhibit and use of the system. Also in training scenarios, many tasks have a pre-planned procedure in which once the worker is able to remember that procedure, there is no need for that system. Whereas, a packing task that is necessary in goods distribution and/or in e-commerce companies does not have a single solution. The packing solution is depended on the items and container which always varies for each packing task. Therefore, the AR packing assistance is considered useful in long term use.

Not only do I show the AR visualization and claim that AR visualization is better than a traditional way to perform a task, I introduce many designs for the interfaces and evaluate them to understand how the user perceived each one and their performance regarding each interface.

Along with this dissertation, several challenges (1–3) in the learning scenario and one challenge (4) in the training scenario were tackled as follows:

1. For a large exhibit, which interface is proper to use for pointing to information located at a specific position? [Chapter 3]

2. How should a virtual guide move and where should it be located when presenting information? [Chapter 4]

3. How does the presence of a virtual guide effect memory performance and user experience? [Chapter 5]

4. In a packing task, how can an automatic solution visualization help improve task performance, and which type of indication interfaces performs best? [Chapter 6]

1.3 Philosophy

This section presents my proposed concept of user-oriented AR guiding visualizations. While AR can improve the efficiency of work, a user's perspective should be concerned and evaluated such as AR instructions that are easy to understand and/or including a human-like interaction that helps the user feel comfort when using the system. Thus, I suggest the following implementations to improve visualizations for AR:

- *A clue regarding object position:* When the system suggests to the user to look at or pick something up, it should tell the user not only the item characteristic, but also the location of the item the real world.
- *A clue regarding 3D object manipulation:* When the system suggests a user to move something, it should not only display a target position and orientation, but also a sequence of how the object should move and rotate step by step.

By doing this, anybody can easily understand the instruction and is able to follow the AR guidance.

- *Human-like interaction:* Humans always learn and communicate with other humans. Having a human-like interface such as an Embodied Conversational Agent (ECA) that includes verbal and non-verbal communication would better engage a user to perform a task.
- *Estimation of virtual content position:* Virtual content that is presented to the user should be displayed at a proper position to make the user feel comfort and not distract him/her from the main focus.

1.4 Contributions

According to the overview, this dissertation yields the following contributions to the field of human-computer interaction:

1. Pointing interfaces used by a virtual guide in AR-based information presentation. [Chapter 3]
2. Positioning and movement interfaces for a virtual guide [Chapter 4]
3. The merit of a virtual guide in AR-based information presentation. [Chapter 5]
4. An auto-generated packing solution system framework and solution indication interfaces [Chapter 6]

1.5 Dissertation Organization

Figure 1.2 presents overview of this dissertation which consists of seven chapters as follows:

Chapter 1 Introduction: This chapter illustrates background, challenges, philosophy, and contributions of this research.

Chapter 2 Literature Review: This chapter presents related works in information visualization for training and learning scenarios using AR and usage of the ECA.

Chapter 3 An AR Embodied Conversational Agent for Information Guidance: This chapter proposes a framework using for information guiding tasks, and evaluation of pointing interfaces used by a virtual guide.

Chapter 4 AR Embodied Conversational Agent Positioning and Movement Interfaces for Information Guidance: This chapter shows a study of positioning and movement functions to be implemented for the virtual agent and evaluation results.

Chapter 5 AR Embodied Conversational Agent Effectiveness for Information Guidance: This chapter investigates several user interfaces using with the virtual guide. Evaluation of each interface with and without the virtual guide conditions are described.

Chapter 6 Guidance and Visualization of Optimized Packing Solutions: This chapter presents a novel system for packing solution suggestions and user interfaces for indicating solutions. A user study and results are included to this chapter.

Chapter 7 Conclusion: This chapter summarizes the presented frameworks and the finding of the studies. Also, the chapter also suggests recommendations for future research.

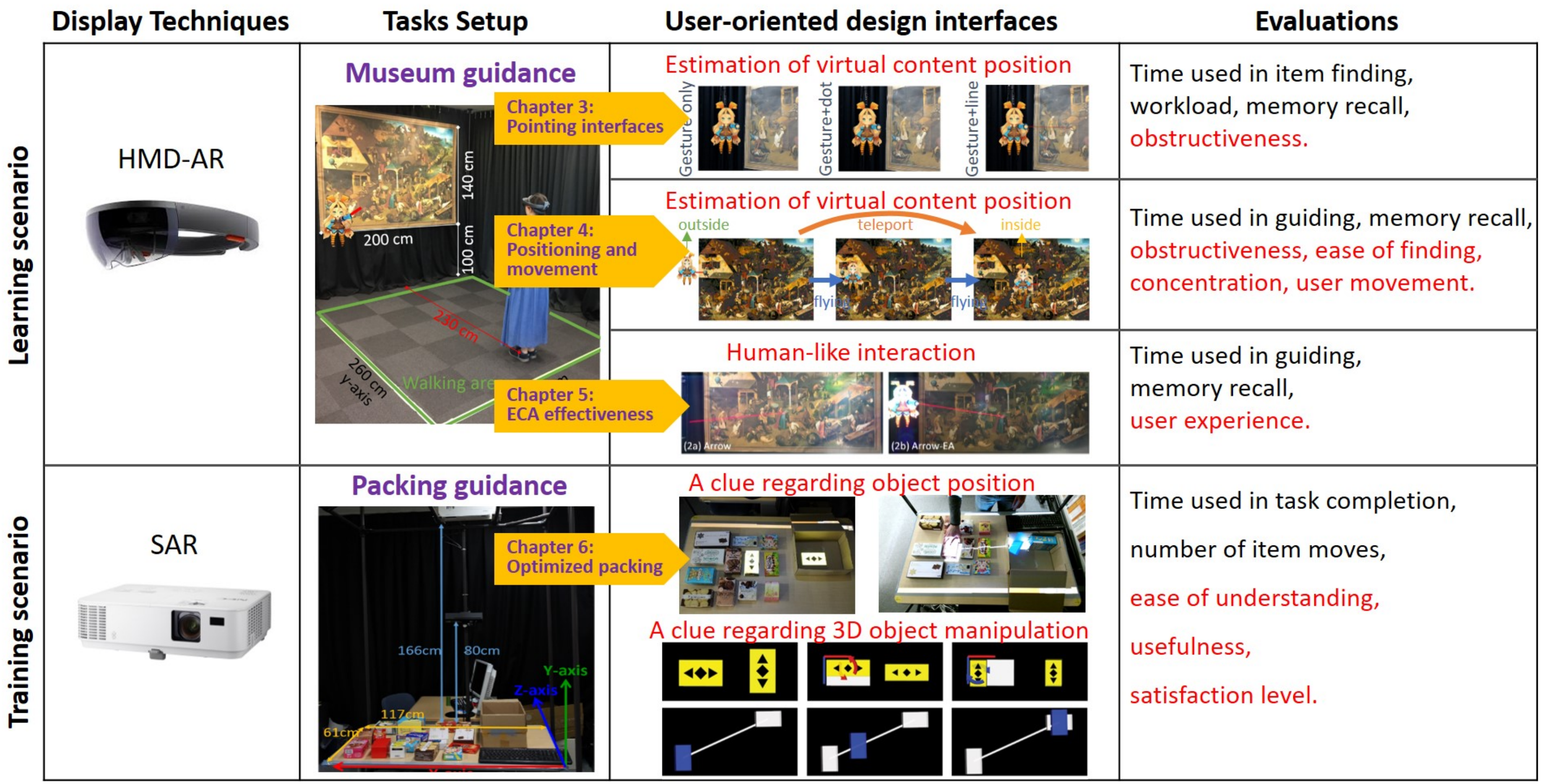


Figure 1.2: The overview of this dissertation. Designs for user-oriented visualization and user perspective related evaluation are highlighted in red.

Chapter 2

Literature Review

In this chapter, previous work related to information guidance in AR on training and learning scenarios is introduced. Also, previous research about Embodied Conversational Agents (ECAs) is discussed.

2.1 AR Systems Used in Training Scenarios

One main application field that utilizes AR is the support or training of human workers to efficiently perform tasks such as assembly, repair, and inspection. Since most of this work involves object manipulation, head mounted displays (HMD) and spatial displays (projector) that allow hands-free interaction are often more suitable to use in training scenarios. One of the first AR systems for assisting work was developed in 1992 at Boeing [17]. The system superimposed wire harnesses for guiding instructions: simple wire frames, icons, and text onto the workspace via an HMD. Later, many applications for other situations have been developed. Other systems have used HMDs to show task instructions including engine oil level checking that used a markerless tracking method to locate the position of the oil tank in an aircraft [18]. The system showed the oil check procedure with an animated 3D model and progress bar overlaid on the real oil tank. There are systems that used AR to give instructions for motherboard assembly [19], and Lego assembly [20], which put a marker at the working area for tracking the object. On the other hand, some systems use a projector to project instructions directly onto the workspace. These include applications such as a spot welding inspection system that showed a position and order of an item to be checked [21] and an object manipulation system that showed a user how to move a particular item from one to another position [22, 23].

Using a projector allows the worker to become both hands-free and does not need head equipment. However, if the viewpoint changes, the perceived image can be distorted. Therefore, the projection system is suited to the task that has no requirement for movement, while an HMD can be a burden but allows the user to move around.

2.2 AR Systems Used in Learning Scenarios

Learning is a lifelong activity for a wide range of people. It can be done both inside and outside the classroom, in which AR can help learning become more entertaining and provide a broader visualization in 3D. In the classroom, researchers have developed an interactive augmented book that helps students with reading comprehension [24], learn about electromagnetism [25], and study earth-sun relationships [26]. One of the early studies using AR in a museum was conducted in 2008 by Miyashita et al [27]. Based on an interview, they found some issues with implementation showing that the size of AR text-description refers to the real word coordinate, so that sometimes the AR text becomes too small when the user is standing far away from the exhibit. Some participants also commented that switching their attention between an AR monitor and the real artwork while holding the monitor with both hands was difficult.

Besides the hand-held device, Mason conducted a user study of AR museums using Google Glass [28]. Many participants reported that seeing the exhibit and video description at the same time was convenient when using the AR glasses. However the users watched the video for only around 10–20 seconds despite it being a total of 70 seconds, and then switched their attention to the real object in front of them. This indicates that the users tend to easily switch their focus if the current object of focus is not interesting enough. The participants also reported that reading long texts in AR glasses is not comfortable when compared to a smartphone. Regarding a survey of user requirements of using a smart-glass for a museum visiting [29], Vainstein et al. found that most participants preferred to control the system using an accompanying mobile device, not by voice, due to the quiet environment of the museum. They also preferred to be able to adjust text size, change audio volume, and fast forward or pause media. They also suggested that the system should automatically stop the presentation when the they move away. Moreover, many of them commented that text display was not easy to read.

Thus, head-worn type AR display devices benefit hands-free interaction when compared to a hand-held display. However, current limitations in technology may prevent reading long text descriptions comfortably.

2.3 Embodied Conversational Agent

An Embodied Conversational Agent (ECA), also referred to as an agent in this dissertation, is a computer interface that has a humanoid or body-like appearance and communicative behavior including natural language, facial expressions and/or gestures [30]. Due to the communicative behavior that is required in the ECA, most ECAs are able to move parts of their bodies to create natural interaction.

ECAs were developed to interact with a user as an assistive or human companion via computer screen, virtual reality, or augmented reality. Many roles of the ECA have been developed such as an interior advisor [31], hospital companion [32], news reporter [33], weather reporter [34], museum guide [35], and job trainer [36].

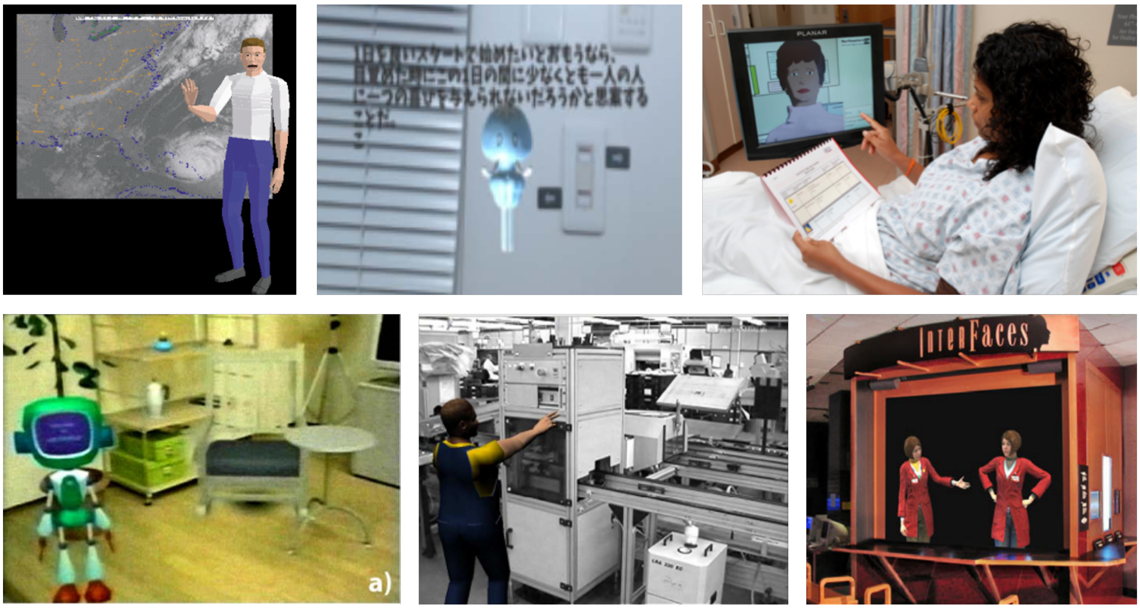


Figure 2.1: Example of Embodied Conversational Agents.

Regarding other aspects from information conveying, these ECAs were also rated positively on both subjective and objective ratings. Van Mulken et al. have studied the effect the presence of an agent has on learning performance [37]. They conducted a user study on presentations of technical and non-technical information with and without agent conditions. They found that the presence of the agent has a significantly positive effect on only the technical information at the subjective measure. Participants who listened to the presentation with the agent presence condition rated that the presentation was easier to understand, that they felt more entertained, and that the tests were less difficult than those who were presented without the agent, but there was no significant difference found on the test scores. Beun et al. have studied on effects on memory recall of the presence of an ECA [38]. They conducted a user study with realistic agent, cartoon agent, and non-agent image conditions. They found that participants who listened to the realistic agent had a significantly higher memory recall score compared to others who listened to non-agent image. However, there was no significant difference found between realistic agent and cartoon agent.

Bickmore et al. studied user preferences between a human and ECA on a complex document explanation task [39]. They found that users were more satisfied with the explanation from the ECA than a professional human with the reason that the agent has infinite patience, could repeat words, and did not criticize them if they didn't understand something. Also, Hasegawa et al. have investigated the effect of ECA posture in an e-learning system [40]. Their results indicated that in the ECA posture changing condition, students felt engaged and wanted to learn more about the topic, but there was no significant difference in the number of correct answers.

Moreover, Campbell et al. conducted a user study on an AR navigation task [41]. They found that by following an ECA who leads while walking, the time used in task completion was faster than following arrow-type indicators that were placed along the way.

Wagner et al. have developed an art history application [42]. They conducted a user study comparing several conditions from non-ECA to 2D ECA, 3D ECA, and AR ECA to explore its effectiveness. While there was no significant difference in time to perform the task and the average number of correct answers, the ECA was perceived helpful. Based on their interview results, although the ECA in AR condition looked natural and realistic, many participants commented that it increases required use time because the user has to wait for the walking animation. Also the ECA lacked eye contact with the user.

When considering AR guidance development, adding an ECA to the system could bring many merits. The ECA can act as a virtual guide, virtual companion, or virtual instructor who talks, smiles, pays attention to the user, and points out interesting parts of the artwork to the user. While the effect of ECA on memory retention of the information can still be ambiguous, the ECA-type interfaces are preferred by the user. Therefore the user would pay more attention to the task, which at the very least can positively increase the experience of the task.

Chapter 3

An AR Embodied Conversational Agent for Information Guidance

3.1 Introduction

Embodied Conversational Agents (ECAs) are a promising type of guidance interface. Due to its lifelike behaviour i.e. it has anthropomorphism and is able to communicate, a human user feels accompanied by this type of agent, despite knowing that it is a computer-generated character. Based on previous research, an ECA not only has positive effects on user preference but also on user performance such as memory retention [38] and navigation [41]. As an ECA is developed to interact with a human, applying its abilities to a real world situation could help support a user in some tasks in which human experts are unavailable or unable to perform some kind of support.

In this research, I focus on information guidance for a large-scale multi-section object. An example could be in a museum or exhibition as shown in Figure 3.1, where a piece of artwork or a panel is bigger than reaching distance and there are many pieces of information contained in that object. In such a scenario, an ECA type virtual guide that is able to fly to the top and point to an interesting part of the artwork

Figure 3.1: Example scenario of a crowd environment (image taken from [43]).

could perform a guidance in an interesting way that the human guide cannot.

To demonstrate and explore the possibility of this idea, I developed an information guiding system by using ECA as a virtual guide. I then tested the system on a printed version of a painting at actual size. I have also developed 3 pointing interfaces to help guide the user to information.

3.2 System Design

The AR guide system is designed to provide descriptions of all available regions of a large painting containing multiple details by pointing out each position of guidance and explaining the information about that piece (region of the image) to the user. I decided to develop the system using the Microsoft HoloLens, an optical see-through head-mounted display [44], which is a standalone device that allows a user to walk around. It also can perform real-time localization by mapping its own coordinate system to real world coordinates and visualize high quality mixed reality images.

While a projector-based system frees the user from the head equipment, it is necessary to install many cameras and projectors in the exhibition area and also calibrate the projector-camera viewpoint. This requires more work for setting up in a museum that contains multiple rooms, and it is also hard to change the viewpoint later.

In terms of the user experience, the HMD provides one-to-one interaction, which means that an appropriate guide adapted for each visitor is possible to implement.

This system uses images from the device's forward-facing camera to track the position of the painting itself, and also either receives button input from a Microsoft Xbox controller, clicker device, or a hand gesture to trigger the system to explain subsequent points of information. The system was developed with the Unity game engine version 2018.2.1f1, with integration of HoloToolkit 2017.4.3.0 and Vuforia AR Engine 8.0.10. A system diagram is shown in Figure 3.2.

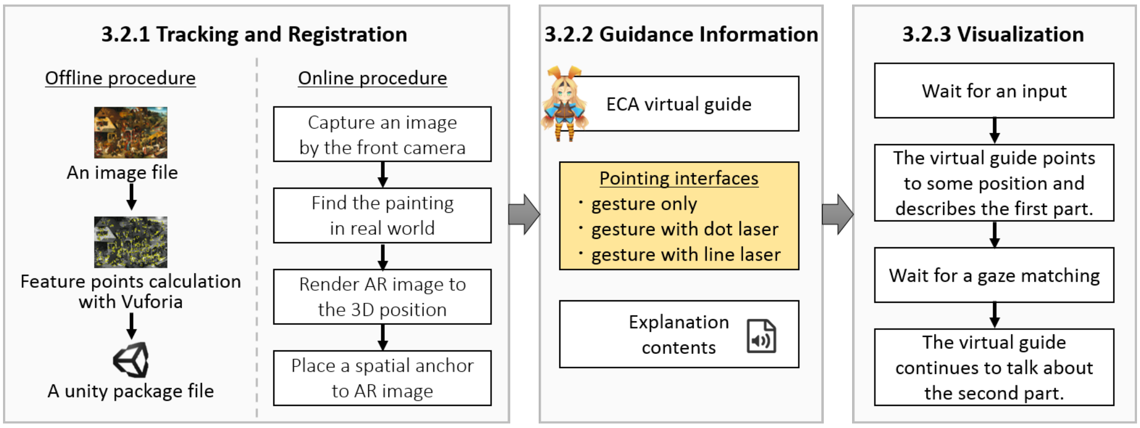


Figure 3.2: A diagram of the system implementation consists of image tracking, guidance information, and visualization.

3.2.1 Tracking and Registration

The painting image was tracked by the markerless image tracking function from the Vuforia AR Engine. The image tracking includes offline and online procedures.

In the offline procedure, an image of the painting was uploaded to the Vuforia web service in order to automatically calculate and generate its feature points which are encrypted as a Unity package. The dimension of the actual painting was measured to perform accurate tracking.

In the online procedure, the system captures an image by the front camera of the HoloLens in real-time and looks for the painting. Once it is found, the system maps the 2D image position of the painting to a 3D world position in the real environment and renders an AR guidance information overlay on top of the painting. Once the overlay is registered, predefined items for the painting are also registered automatically at their relative positions so that the system can accurately refer them in the painting.

However, when the user walks close to the painting, the front camera will not be able to see the whole painting, causing the system to lose the painting position, which results in a possible shift in the overlaid AR image position (due to user movement). To solve this problem, I placed a spatial anchor to the AR image immediately after the system is able to locate the AR image position in 3D and turn off the image tracking function. The spatial anchor is a 3D point representation in the HoloLens (real environment) coordinates. When it is attached to a virtual object (AR image), its position will be kept and tracked over time until the system shuts down. Therefore, the registration accuracy of the AR overlay can be maintained. I also saved the anchor data to the HoloLens application's roaming folder for subsequent guidance. When the system is started up for the second time, the system will look for the anchor data in the disk. By loading the anchor data, the system gets a position in 3D real world so that it can place the AR image to that position without performing the image tracking again.

3.2.2 Guidance Information

There are 3 components in the guidance information module including an ECA virtual guide, pointing interfaces, and explanation contents.

ECA Virtual Guide

Model and Size: Regarding selection of the model appearance, the choice of the character effects the user's impression. For example, the ECA with a human body is perceived as more professional than a geometric shape or other objects with a face [45]. While there is not much difference between the human-type and non-human-type ECA on usefulness and enjoyment, trust in the agent was rated very low for the non-human-type. Therefore, I decided to use a human-type ECA as a virtual guide, as it could benefit trust and acceptance from the user more than a non-human-type ECA.

Due to the limitations of the HoloLens, the virtual graphics that are seen by the user take up an area of 30-degrees horizontal and 17-degrees vertical in front of the head position, which does not cover all of the area seen by the user's eyes. Displaying the human-size of the ECA, the users usually see the embodied guide image cutoff which can affect the user experience. To solve this problem, I scaled the size of the virtual guide down to fit in the display height.

I decided to use a character that has a slightly larger head proportion so that the user can see the guide's facial expression more easily. The virtual guide was a young female character model called SD-Unitychan [46]. This model is provided by the Unity Technologies Japan and it is free to use. It is already rigged and has some basic facial animations provided so that programming the movement of the hands, mouth, and eyes can done easily. I set the height of the virtual guide to 50 cm, which fits the device display height when the user stands inside proxemics based social distance [47], around 160 cm away from the virtual guide.

Interaction: The virtual guide contains both verbal and non-verbal communication functions to perform a natural interaction to a human user. When the virtual guide is talking, its lips move randomly every 0.1 seconds among 5 vowel shapes: a, i, u, e, and o, and 3 closed-lip shapes. This speed was adjusted to match the Japanese speech rate of 6–11 syllables per second [48].

For non-verbal communication, eye contact is an important means for creating a communicative relationship [49]. Therefore, the virtual guide always makes eye contact with the user by facing its body towards the position of the user while describing the painting. This can make the user feel that the agent is talking to him/her. Each eye blink is at an interval of 2–4 seconds. Also, the virtual guide uses an open hand gesture to point toward the explanation and indicate the direction of the explanation. A single left or right hand is determined from the side closest to the position of the text.

Position: Regarding the setting of the virtual guide's position in the real world, it can either be standing on the floor or floating in the air. However, in a typical museum and exhibition, the exhibit is normally placed at the human's view position i.e. not at a floor. Therefore, I set the virtual guide's position to a region that was floating in the air near the exhibit.

In this primary implemented system, I decided that the virtual guide should not overlap the painting because it will occlude details of the painting. Also, I want to test the benefits and obstructions of each pointing interface in this study without worrying about obstruction of the painting area by the virtual guide. Therefore, I set the virtual guide to stand still on one side of the painting.

Pointing Interfaces

Three pointing interfaces: a gesture only interface (G-only), a gesture with a dot laser interface (G-dot), a gesture with line laser interface (G-line), were designed to use with the virtual guide as depicted in Figure 3.3.

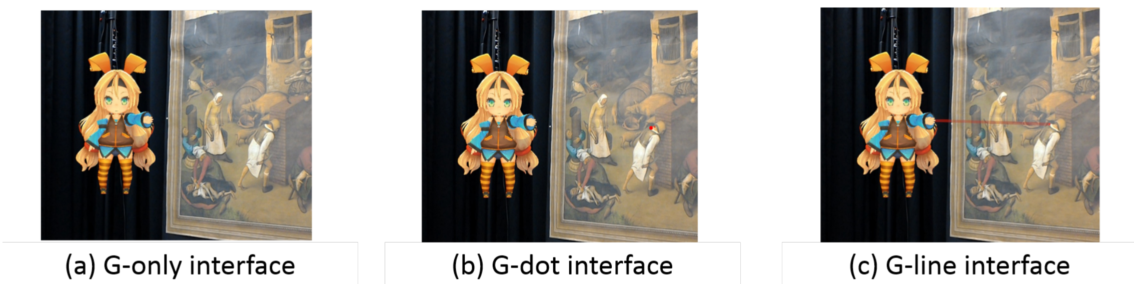


Figure 3.3: Images showing each of the three proposed pointing interfaces as used by the virtual guide.

G-only: This is a basic interface that will let the virtual guide make a gesture toward the direction of the position in concern, similar to human pointing. The virtual guide was designed to use an open hand toward a user direction instead of the index finger, which is a polite way to refer to the position of something. However, in case the artwork is too large, the virtual guide's hand would not be able to reach the pointing position, thus only direction information is delivered by the G-only interface.

G-dot: This interface indicates both a direction and an exact position. It is based from G-only, but adding a laser pointer that shows a small red dot (around 1 cm diameter) at the position to which it is pointing. The size of the dot was kept as small as possible, but still visible to the user in order to prevent occlusion of the artwork details. Although the G-dot interface displays both direction and position clues, it may still hard to find the position in some cases. For example, when the user has some observed direction shifted, and the pointing position is very far from the agent, this would prevent the user from looking at the correct area.

G-line: This interface draws a line that leads the user's eyesight to the position. It builds on top of the G-dot interface in addition with a red ray (1 cm thickness) drawn from the virtual guide's hand to the pointing position. This interface would be intuitive to follow and useful for finding the destination position. However, it may obstruct some painting areas along its drawing path.

Explanation Contents

Regarding a survey on user requirement of using smart-glass for a museum visiting, many users rated that they did not prefer the system to display text because it is hard to read [29]. Therefore, this system implements the explanation using a voice description and graphical visualization to give information about the exhibit. Twelve explanations around the painting are selected. Each text description is shown in Table 3.3. During guidance with the system, the voice explanation is divided into two parts. The first part mentions the section in the painting to look at and what the image in that section looks like. The second part tells the story or hidden meaning of that image. Since participants in the system evaluation are expected to be either Native Japanese speakers or Non-native Japanese speakers, Both English and Japanese female voice descriptions are provided, which are generated from the Amazon Polly Text-to-Speech service.

3.2.3 Visualization

In the visualization module, information guiding for each item starts after the user gives input to the system i.e. a hand gesture, click on a clicking device, and/or button press on an Xbox controller. Then the virtual guide raises its hand up to point to a particular position and says the first part of the description including section position to look at i.e. lower/upper and left/right/center, and it describes the image that it is talking about. After that, the system will wait for the user to look at that position, for which there is a 5*5 cm transparent square placed on that image area.

In the user view, he/she will see a white dot (around 0.4 cm diameter) that indicates the center of the HoloLens. This white dot moves along with the user's head direction. When the white dot is inside the area of the current talking piece of the image, the system will trigger a sound to notify the user that he/she found the target. Then the virtual guide continues talking about the second part of the description.

3.3 Evaluation

3.3.1 Experiment Setup

The painting used for testing was "The Netherlandish Proverbs," painted by Pieter Bruegel the Elder in 1559 [50]. The painting is in the public domain, meaning it is free to use. The image was printed at a size of 200 by 140 cm and hung on a wall 100 cm above the floor. Through informal testing, I observed that 140 cm from the picture was an ideal distance for the user to be able to see all the details in the image clearly.

Twelve different labels for subsections of the image were prepared, which were grouped into 3 groups for testing with the three pointing conditions. Figure 3.4 shows each group: 1, 2, 3 in pink, orange, and blue color circle, respectively. In the first group the explanation starting from 1A (the double pink circle) to 1B as indicated by the arrow, 1C, and 1D. The second group and the third group are shown in the same A, B, C, D orders.

In the experiment, twelve participants (4 female, 8 male) from lab members with an average age of 24, ranging from 18 to 28 years old were recruited to help evaluate the effectiveness of the system and pointing interfaces. The three pointing interfaces (G-only, G-dot, G-line) were balanced between 6 different orders to alleviate learning effects, meaning every two subsequent participants were tested on a different order of the interfaces than the previous two participants.

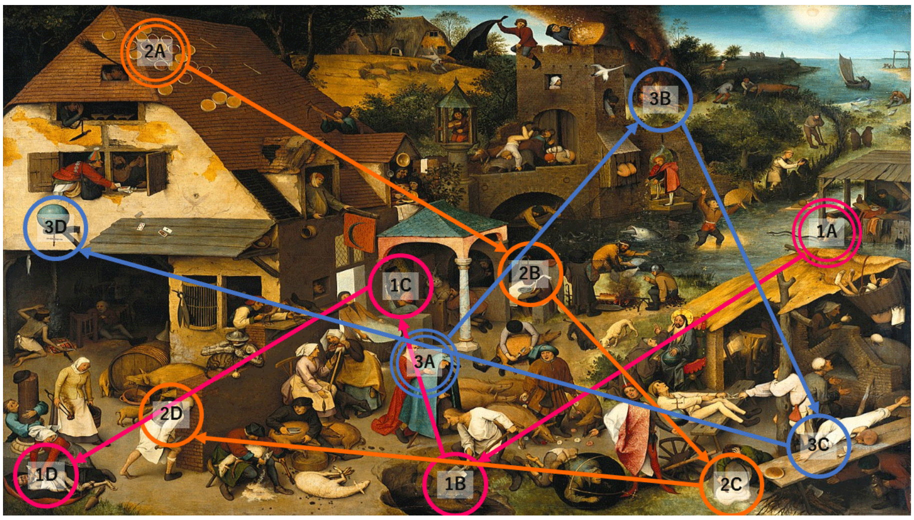


Figure 3.4: The painting and detailed positions for explanation that were used in evaluation of the system. Three groups of sections (different colors) were used for randomizing the sets of trials between the three different pointing interfaces.

3.3.2 Procedure

First, the experiment proctor asked the participant to fill in his or her demographic information. Then, the participant sat on a chair placed 140 cm away from the painting as shown in Figure 3.5. From the user's point of view, the painting visual angle is around 70 degrees horizontal and 45 degrees vertical. The experiment proctor let him/her wear the HoloLens and confirmed that he/she saw the white dot that represents his/her gaze and the virtual guide at the left corner of the painting. Then, the experiment proctor explained to the participant how to use a clicker device to trigger the virtual guide to give information about each item in the painting, and a sub-questionnaire was administered between testing with each pointing interface for each of the three interfaces.

When the participant finished the experiment and questionnaire, he/she was asked to complete a memory test without knowing in advance.

The test consisted of the painted image and 12 phrases that were referred to during

Figure 3.5: Image of AR virtual guide using her pointing interface for explaining large scale artwork. The blue rectangle represents the display's field of view for the AR guidance.

the explanation dialogue corresponding to each area as depicted in Figure 3.6. The participants were then asked to draw an arrow from each phrase to its corresponding position. After finishing the memory test, each person was interviewed regarding the system usage, questionnaire, and memory test to obtain insights and feedback.

Figure 3.6: Image of a memory recall test in Microsoft Excel. The user has to drag an arrow head from the text-boxes of image meaning to the image position.

3.3.3 Hypotheses

Since the G-only interface only provided direction to part of an image, the user had to listen to the virtual guide to refer to the corresponding position. On the other hand, the G-dot interface provided a mark for the destination so that time used for searching in the G-dot interface would be shorter than the G-only interface. Also, since the G-line interface provides a ray from the virtual guide hand the target region of the image, it is likely that the user could quickly find the position without listening to the virtual guide talk. Therefore, though the G-line interface allowed the user to find the position fastest, it may not help the user remember details regarding that piece, and it would also be the most obstructive when viewing the painting.

Regarding the virtual guide, we expected that with her smiley face expression, gaze towards the user when talking, and gesture, she would make the painting more interesting to explore. Accordingly, I formulated the following hypotheses.

- **H1**: The gesture with line laser interface (G-line) will result in the lowest search time when compared to other interfaces.
- **H2**: The gesture with dot laser interface (G-dot) will result in the highest memory score when compared to other interfaces.
- **H3**: With the virtual guide, the system will make the user feel more interested in the painting.

3.3.4 Experiment Results

Search Time

Time used to find a particular position was measured from the time that the virtual guide made a pointing gesture and started talking to the time that the user looked at the correct position. The average time used in the G-dot and G-line interfaces was 59.2% and 31.4% of the G-only interface, respectively.

I conducted a parametric repeated measures Analysis of Variance (ANOVA) test, followed with a post-hoc analysis using pairwise t-tests with Holm's adjustment to evaluate the time results. A significant difference between the 3 interfaces with $F = 14.64$, $p < 0.001$ was found. These results are summarized in Figure 3.7, where the asterisk ***, **, and * symbols between each set of bars indicates a significance level of $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

Workload

Afterward, I measured the workload after the user finished each guiding interface by using a 20 point NASA-TLX scale [51] that contained seven types of rating including Mental, Physical, Temporal, Performance, Effort and Frustration. A higher score corresponds to higher workload on each dimension except the performance dimension in which a higher score suggests the participant is less satisfied on his/her task performance. A non-parametric analysis Friedman rank sum test was conducted, following with a multiple comparison Wilcoxon signed rank test. The results reveal a significant difference on Mental scale with $\chi^2 = 15.368$ and $p < 0.001$, Performance scale with $\chi^2 = 14.6$ and $p < 0.001$, and Effort scale with $\chi^2 = 13.762$ and $p < 0.05$. Each scale rating results are shown in Figure 3.8.

Memory Recall

A number of pieces from the image to which the user could match the description correctly varied from person to person. Table 3.1 shows the number of correct answers for each pointing interface. One-third of the participants (four) were able to remember all 12 explanations, half of them (six) could remember more than half, and the remaining two could remember half of the total explanations.

Although the G-dot interface got the highest sum score of all participants, there was no significant difference when compared to other interfaces. Table 3.2 shows score and pointing interface name by experiment order. Also, while there was no significant difference found, there is a trend that just above half of the participants forgot the detailed information of their first trail.

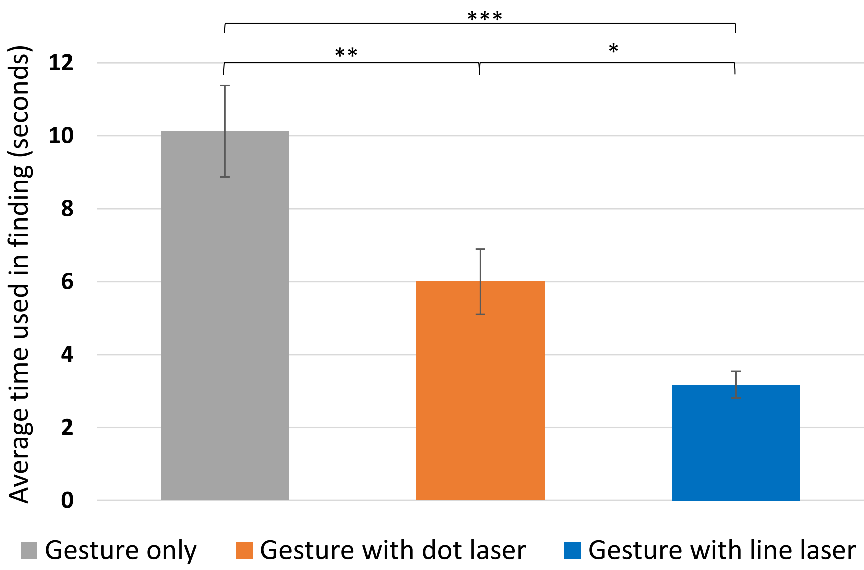


Figure 3.7: Average position searching time by each pointing interface (in seconds).

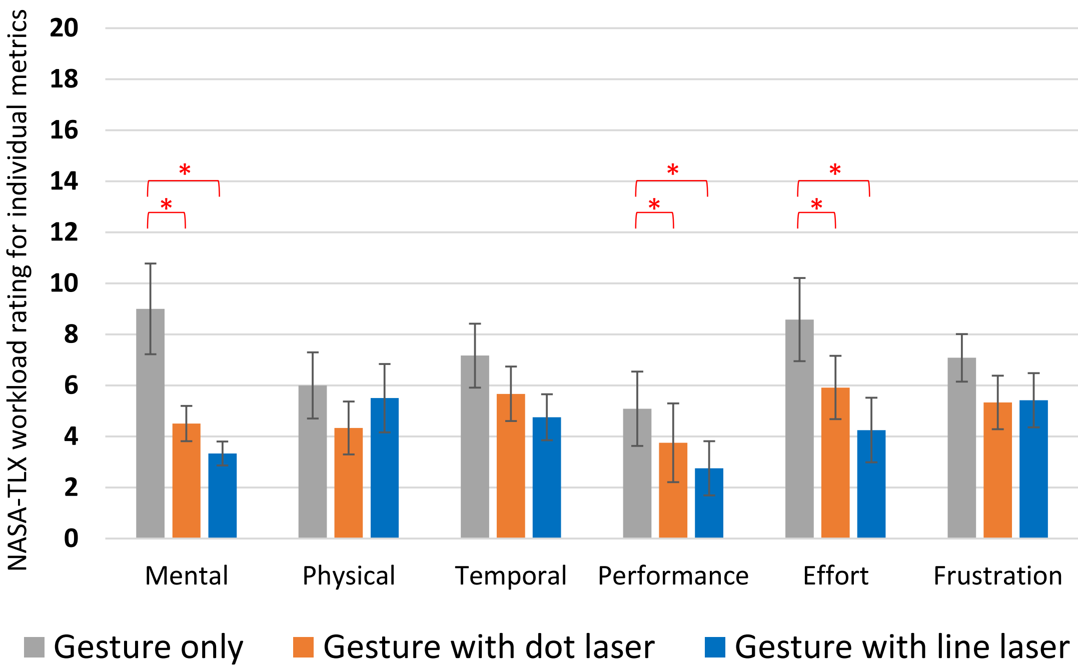


Figure 3.8: NASA-TLX workload rating for individual metrics.

Table 3.1: Raw score by pointing interface

	G-only	G-dot	G-line	Total(12)
	4	4	4	12
	4	4	4	12
	4	4	4	12
	4	4	4	12
	3	4	4	11
	4	3	2	9
	2	3	4	9
	4	2	2	8
	2	3	2	7
	3	1	3	7
	1	4	1	6
	2	2	2	6
Total (48)	37	38	36	

Table 3.2: Raw score with pointing interface name of participants memory recall by experiment order

	1st trail	2nd trail	3rd trail	Total (12)
	G-line (4)	G-only (4)	G-dot (4)	12
	G-only (4)	G-line (4)	G-dot (4)	12
	G-line (4)	G-dot (4)	G-only (4)	12
	G-dot (4)	G-only (4)	G-line (4)	12
	G-line (4)	G-dot (4)	G-only (3)	11
	G-line (2)	G-only (4)	G-dot (3)	9
	G-only (2)	G-dot (3)	G-line (4)	9
	G-dot (2)	G-line (2)	G-only (4)	8
	G-only (2)	G-line (2)	G-dot (3)	7
	G-dot (1)	G-line (3)	G-only (3)	7
	G-only (1)	G-dot (4)	G-line (1)	6
	G-dot (2)	G-only (2)	G-line (2)	6
Total (48)	32	40	39	

Obstructiveness

After testing each individual pointing interface, the experiment proctor asked the participant to rate the obstructiveness of the current interface to the real painting, rated from very obstruct, obstruct, neutral, not obstruct, to not obstruct at all. The results are shown in Figure 3.9. None of the participants rated any of the interfaces as very obstructive to the painting. In the G-only interface, eight participants (66.7%) answered that it did not obstruct at all. The one who rated agree the G-only interface is obstructing was concerned that the white dot representing gaze position was always shown in front of the user view. In the G-dot interface, five participants (41.7%) and six participants (50%) also answered that the interface is not obstructing at all and not obstructing, respectively. Also, two participants (16.7%) said that the G-line interface is not obstructing at all, and half of them answered that the interface is not obstructing.

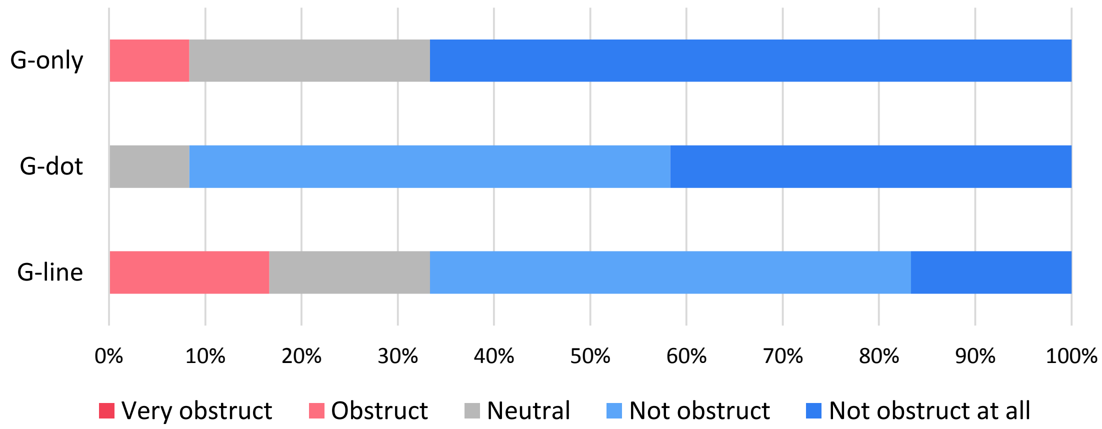


Figure 3.9: Subjective ratings for obstructiveness of each interface.

System Ratings

The proctor asked the participant to choose the interface that they preferred most. Six of them (50%) prefer the G-line interface, five of them (41.7%) prefer G-Dot interface, and only one (8.3%) prefer the G-only interface.

I have evaluated the system effectiveness by asking the participant regarding

- **Q1**: Before using the AR guide, I'm interested in this painting.
- **Q2**: With the virtual guide, it made the painting more interesting.
- **Q3**: I feel the quality of the painting image through HoloLens, is the same as if seeing through my own eyes.

Results of these aspects are shown in Figure 3.10. There are two participants (16.7%) and one (8.3%) who rated disagreed and strongly disagreed on **Q1**, which means that they were not interested in the painting used in the system evaluation. However, no one rated disagree on **Q2** which concludes that the system with an virtual guide made the painting more interesting. Also, five participants (41.67%) rated strongly agree, and two of them (16.7%) rated agree on **Q3** that indicate more than half of them do not recognize that quality of the painting seeing through HoloLens is little different.

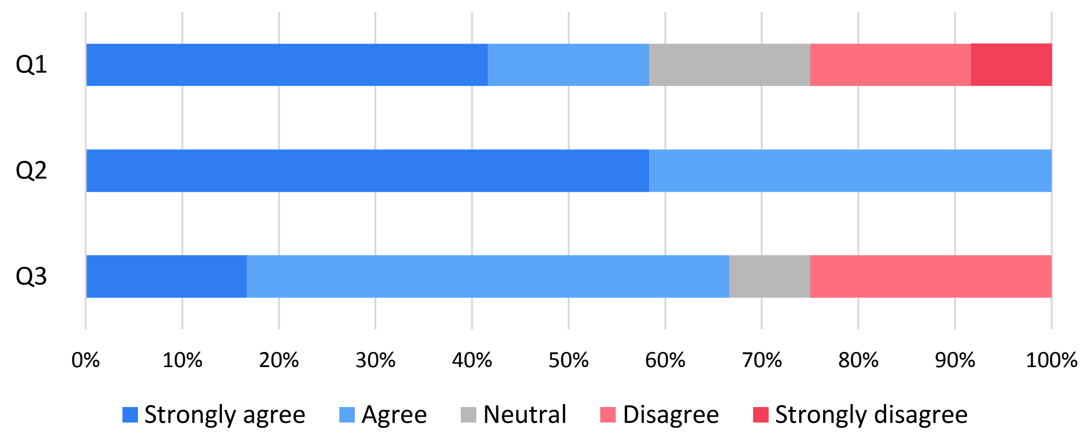


Figure 3.10: Subjective ratings from the questionnaire.
Q1: Before using the AR guide, I'm interested in this painting.
Q2: With the virtual guide, it made the painting more interesting.
Q3: I feel the quality of the painting image through HoloLens is the same as if seeing through my own eyes.

3.3.5 Discussion

In the evaluation, I found that the pointing interface of the G-line was the fastest interface in terms of helping the user search for the described position. This result supported hypothesis **H1**. On the other hand, although the G-dot interface got the highest memory score, these results were not statistically significant, so the hypothesis **H2** is only partially supported. Regarding the system rating, all of the participants rated strongly agree and agree on the question **Q2** that the system made the painting more interesting, which supported hypothesis **H3**.

Based on the interviewing of participants, some of them mentioned that the HoloLens, which weighs 579 grams, was too heavy, and it made them feel some neck pain when looking up for a long time. Participants who rated that the painting quality is different when looking through HoloLens mentioned that the color is a little bit changed, and the frame of the HoloLens obstructed he/she to see the painting.

Many participants suggested that the G-line interface helped them easily find the area of concern in the painting, however, they want to remove the line after they found the position. Some suggested to let the virtual guide move near the part of the image in concern and point it out with her hand.

With regards to the memory result, some participants who did not get a good score mentioned that he/she was not interested in the painting, he/she just enjoyed seeing and hearing the virtual guide guide but didn't think about remembering, and the description of some words in Japanese was not natural.

3.4 Conclusion

In this chapter, an AR guiding system for multi-section visual information with a virtual guide was proposed. Three types of pointing interface were designed to help the user find positions of interest inside large scale works of art more quickly. The system tracked the painting in real-time and projected the AR image onto the tracked image, which included positions of each sub-area to be guided and the virtual guide at the left side of the painting.

To explain information for a specific part of the image, the virtual guide points toward the position and describes where and at what to look and waits for the user to look at that area before talking about the details of that image.

To evaluate the effectiveness of the system, I have tested three types of pointing interfaces: gesture only, gesture with dot laser, and gesture with line laser. Experimental results showed that the line laser interface significantly helped the user find target regions by 31.4% less time than the Gesture only interface. The average memory recall of gesture only, gesture with dot laser, and gesture with line laser interfaces were 77.1%, 79.2%, and 75%, respectively, though these differences were not significant.

Table 3.3: List of descriptions

No.	Description	Duration
1.	At the bottom-left corner, you will see a woman who is trying to kill a devil with her hands. This part of the painting represents the phrase "to tie even the devil to a pillow", which means that stubbornness overcomes everything.	13 sec
2.	At the bottom of the painting. you will see a man who is filling a waterhole with the soil around. Inside the waterhole, there is a drowned baby cow. This part of the painting represents the phrase "to fill the well after the calf has already drowned". that means taking action only after a disaster. In other words, it's too late to take action.	20 sec

No.	Description	Duration
3.	At the right side of the painting, you will see a man who is catching an eel on its slippery tail. The phrase “to hold an eel by the tail” represented by this part of the painting, which means to undertake a difficult task.	12 sec
4.	At the left-bottom corner, you will see a man who is banging his head to a wall. This part of the painting represents the phrase “to bang one’s head against a brick wall”, which means to do things repeatedly, even though you have no chance of success.	14 sec
5.	At the center of the painting, you will see a man who is holding candles nearby a devil. The phrase “to hold a candle to the Devil”, represented by this part of the painting. This means to make friends or help anyone without thinking. This phase has an origin from the Roman Catholic practice of lighting candles for prayer to God.	18 sec
6.	At the Upper-right of the painting, you will see a man who is burning something and the fire spreads out over such a large area. Actually, he is making a campfire for warming himself. This part of the painting represents the phrase “to not care whose house is on fire, as long as one can warm oneself at the blaze”. Which means to take every opportunity regardless of the consequences to others.	24 sec
7.	At the upper-left side of the painting, you will see a lot of pies or tarts, that are being used as roofing materials. The phrase, “to have the roof tiled with tarts”, represented by this part of the painting, This means to be very wealthy and having an abundance of everything.	16 sec
8.	At the middle-bottom of the painting, you will see a woman in an eye-catching red dress and a blue cloak man. This part of the painting represents the sentence “she puts the blue cloak on her husband”. 500 years ago, when the piece was painted, red was a color of sin and blue often stood for cheating or folly. This implicitly means the woman is cheating on her husband.	23 sec

No.	Description	Duration
9.	At the lower-right of the painting, You will see a man who is stretching his arms toward 2 loaves of bread, that he can barely reach. The phrase, “to be barely able to reach from one loaf to another”, represented by this part of the painting, which is the same as the expression, “living from paycheck to paycheck”. That means to have difficulty living within a budget.	21 sec
10.	At the left side of the painting, you will see an upside-down globe. This part of the painting represents the phrase “the world is turned upside down. Which means everything is the opposite of what it should be. It also depicts the main theme of this painting, which is also called the Topsy Turvy World.	18 sec
11.	In the middle of the painting, you will see a fox and a crane, which are sitting together at the dinner table. This is an illustration from Aesop’s fables, the Fox and the Crane. One day the fox invited the crane to have dinner at his home, But he served soup in flat dishes. So that, the crane who had a long beak could not enjoy the food. After that, when the crane invited the fox to his home, he served soup in jugs with narrow necks. So that, the fox could not eat the soup. The story is telling that, If you trick someone, they might trick you back.	34 sec
12.	At the lower-right of the painting, You will see a man who is trying to put his food back to the pot. This part of the painting represents the sentence, “he who has spilt his porridge, cannot scrape it all up again”. which means once something is done it cannot be undone. In other words, there’s no use in worrying over the past which cannot be changed.	21 sec

Chapter 4

AR Embodied Conversational Agent Positioning and Movement Interfaces for Information Guidance

4.1 Introduction

In this study, I introduce a wearable AR guide system that makes use of an embodied conversational agent (ECA) as a virtual guide who points out and explains each part of a painting to a visitor. Although much research about ECAs exists [31, 41, 52], it has not been focused on using the ECA as a companion guide for real world use. In this sense, the virtual guide acts in the same manner as a human guide, for example by standing next to an exhibit while describing its contents and maintaining eye contact with the visitor. Moreover, in addition to implementing both verbal and non-verbal communication interfaces, a virtual representation of the virtual guide are also included. This virtual existence enables many functions that are not possible for a human guide, such as showing another virtual image, flying around to reach a high position, or using teleportation functionality to reduce travel time to other real world spaces. Since such virtual functions have not yet been studied in the context of real world situations, I conducted a user study to evaluate the following Research Questions (RQ):

- **RQ 1:** Do users mind if the virtual guide position inside the painting overlaps some of the art?
- **RQ 2:** Does teleportation help reduce total guiding time?
- **RQ 3:** Are there any differences in memory retention between each method of guidance?

The goal of this study is to explore which virtual guide movement and positioning interface users prefer and to explore whether different interfaces affect the user's tour experience and memory retention. I expect that the AR agent guidance system will help the user concentrate and remember information while shortening the time used for exploring information in the painting.

4.2 Design and Technical Implementation

The virtual guide contains three primary functions, including avatar communication, positioning, and movement.

4.2.1 Avatar Communication

The same verbal and non-verbal communication from the previous chapter i.e. eye contact, eye blinking, mouth movement, hand gesture are included in the virtual guide. In this study, A guide's narrative in Japanese was generated from commercial text-to-speech software [53] in which some voice intonations were adjusted manually to increase naturalness. Also based on the study on pointing interfaces in previous chapter, evidence suggested that a line laser interface helps the user more quickly find the part of the painting about which the virtual guide is talking, and most users do not consider this as an obstruction to the painting. As such, I decided to include a virtual line laser to help explicitly point out the position of each item in the painting.

4.2.2 Avatar Positioning

In this study, I adopted a positioning setting of the virtual guide from the previous chapter which keeps it floating in the air around the height of the painting so that the user can see the guide in their normal view of the painting. However, when the virtual guide is floating around at the same height of the painting, it can overlap the painting while giving detailed information.

To study whether the painting overlap has a negative effect on the guiding or not, two guiding position interfaces were created: position outside and position inside. The actual view of both interfaces are displayed in Figure 4.1 and 4.2.

Using the position outside interface, the virtual guide does not occlude the area of the painting by staying at one side of the painting at the same height of the explained segment. The virtual guide is located to the side: left or right that is the closest to the explaining position. In contrast, the virtual guide with the position inside interface occludes some area of the painting but stays next to the explained segment. I assumed that the users who pay attention to the virtual guide would prefer to have it stay inside their view while listening to it talk.

Figure 4.1: The actual view through HoloLens with outside type position interfaces.

Figure 4.2: The actual view through HoloLens with inside type position interfaces.

4.2.3 Avatar Movement

When the virtual guide starts talking about the new subsection of the painting, a new position will be applied. To move the virtual guide from the current position to the new position, I provided 2 movement interfaces: teleportation and flying. The teleportation interface moves the virtual guide to a new position immediately so that the user doesn't have to wait for the agent to reach the new position and start explaining a new item. In contrast, the flying interface gradually moves the agent as a straight-line path to the new position, which takes time if the agent needs to move from one side of the artwork to another. It takes 2-3 seconds for the virtual guide to fly from one side of the painting to the other.

By combining avatar positioning and movement, there are 4 guiding interfaces, which include: Outside-Flying (OF), Inside-Flying (IF), Outside-Teleport (OT), and Inside-Teleport (IT) to be used by the virtual guide. Sequential frames of each interface are shown in Figure 4.3

The virtual guide with the OF interface does not occlude the painting while describing the corresponding piece of information in the painting. However, when the agent flies to the other side of the painting, it occludes the painting and the user must wait for the movement to be finished before listening to the next narration.

The virtual guide with the OT interface does not occlude the painting while guiding a participant. When it needs to move to another location in the painting, it will disappear from its current position and appear at the new position. The user can turn his or her head and quickly find the new positions of the agent and explanation. The virtual guide with the IF or IT interface floats near the description while occluding other parts of the painting. With the IF interface, the agent will fly from its current position to the next position, which takes longer when compared to the IT type.

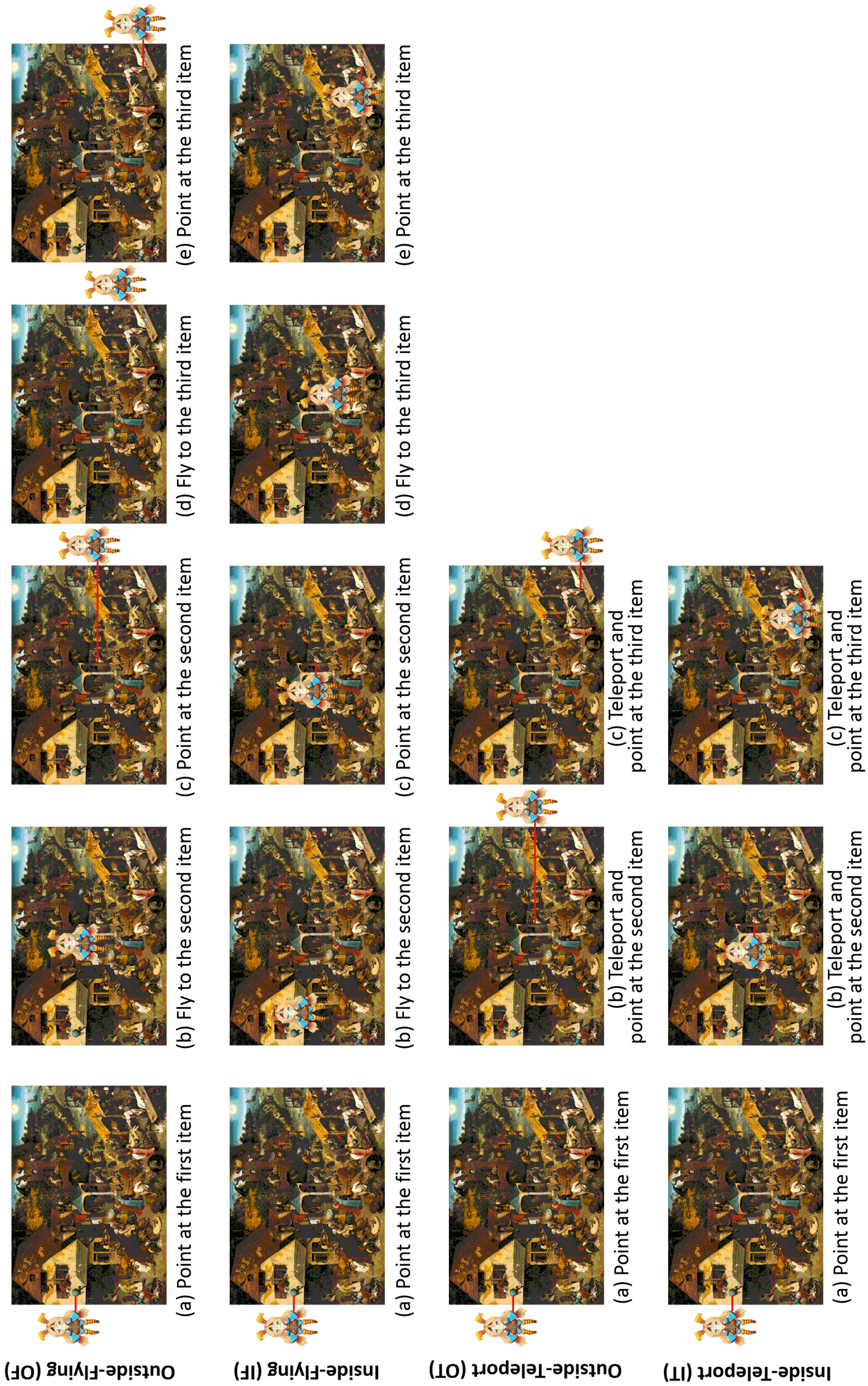


Figure 4.3: Sequential images of each interface.

4.3 Evaluation

4.3.1 Experiment Setup

The same painting and setting on the previous chapter was used for testing the proposed interfaces. However, in this study the user is allowed to walk around a 250 by 260 cm area in front of the painting as shown in Figure 4.4.

Sixteen different subsections of the image were prepared, with sets of four sections grouped together to test the four guiding interfaces (OF, OT, IF, IT), the orders of which were balanced to alleviate learning effects. The guiding started in the same order, from the first trial to the forth trial, and in the same order in each trial from the most left circle to the most right circle as depicted by the directional arrow in Figure 4.5. An explanation for each part saying a location of that part in the

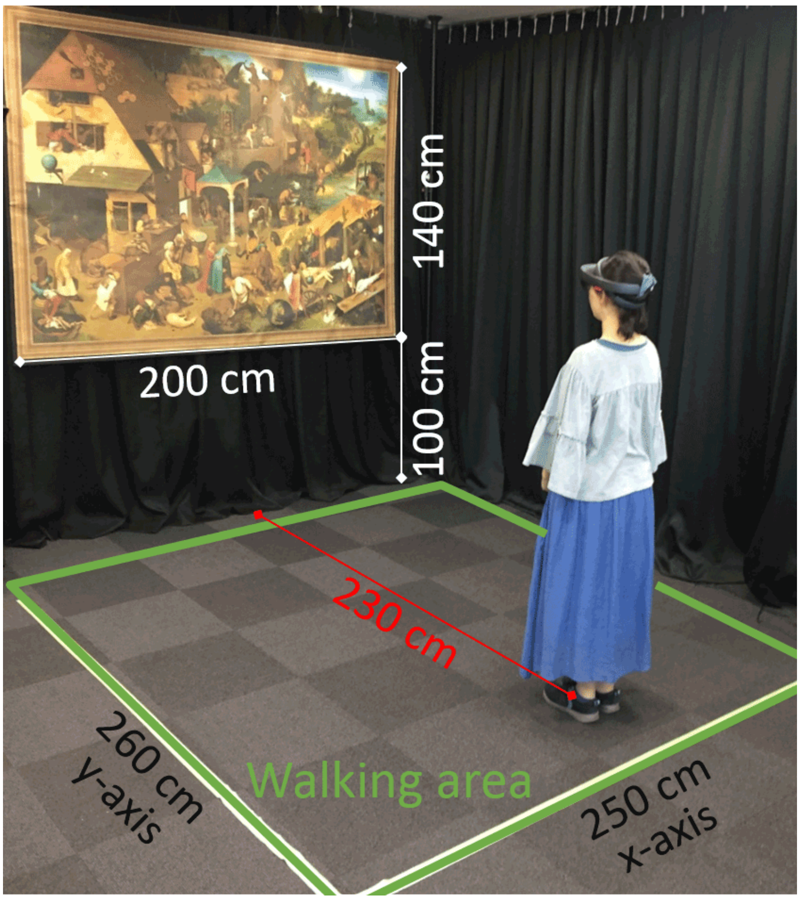


Figure 4.4: The area in which the experiment took place.

painting, the image characteristic to look at, and the meaning of that image. The explanation audio ranged from 11 to 17 seconds, with an average of 14.38 seconds. One example of an explanation is "At the right side of the painting, you will see a man who is catching an eel on its slippery tail. This is called to hold an eel by the tail, which means to undertake a difficult task." Before the explanation start, the virtual guide will move to the position near the describing part with the selected positioning and movement interface. After the avatar reaches the determined position, it will point to the direction of an image and start talking.

In the experiment, 28 participants (13 female, 15 male, mean age of 21, SD of 1.61) were recruited to help evaluate the effectiveness of the guiding interfaces. Participants were mostly students, and came from various departments, studying in the Schools of Language, Letter, Human Sciences, Health Sciences, Biosciences, Law, Economics, Sciences, Engineering, and Engineering Sciences. The ratio of sciences and non-sciences students was 11:17. They were paid with a bookstore card valued at 500 yen (around 5 USD) for participating in the 20 minute experiment.

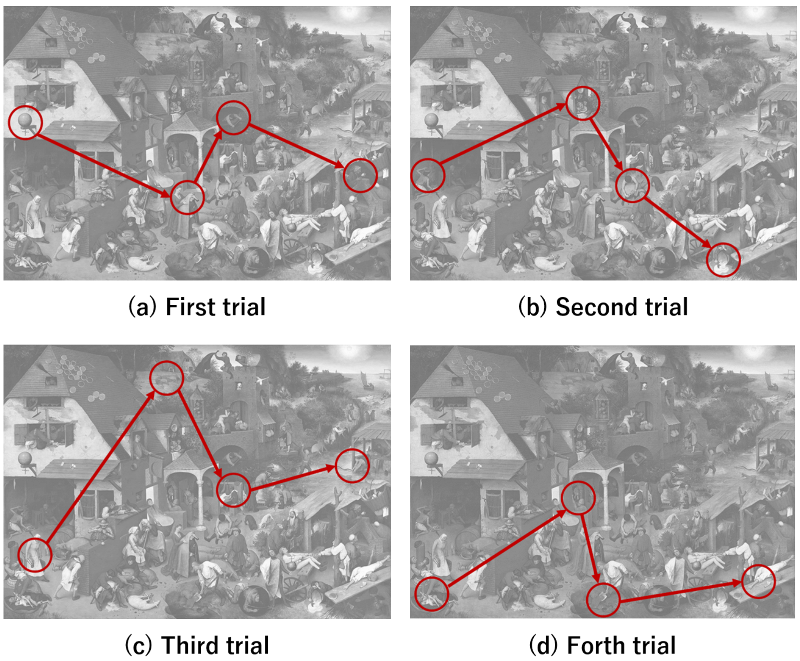


Figure 4.5: The painting and detailed positions for explanation that were used in the evaluation of the system.

4.3.2 Procedure

First, the experiment proctor explained the experiment procedure, including a description of how to use the AR guide system to each participant. The experiment proctor also informed that there would be a test after the experiment and asked the participant sign a consent form. Then the application was started and the participant was asked to stand 230 cm away from the center of the painting. At this position, the participant could see the overlaid graphics in a corresponding physical area of 130 * 70 cm.

Then the experiment proctor let him/her wear the HoloLens and confirmed that he/she sees the virtual guide at the left corner of the painting, which was a standby position. Next, the experiment proctor informed the participant that the experiment will start when he/she presses a button and he/she is free to move around the designated viewing area. The times when the user triggered explanations and standing positions in the environment were recorded during the experiment. The experiment started with demographic questions. Then, the virtual agent guided the 4 segments of the image. The participant had to press a button on a wireless controller to trigger the virtual guide to proceed to the next location. After participants finished listening to all four explanations present in the painting, they had to complete a subjective score related to each of the guiding interfaces. Items were scored on a 5-point Likert scale for the following statements:

1. The virtual guide obstructed the area of explanation.
2. the laser pointer obstructed the area of explanation.
3. A position of the virtual guide was proper.
4. It was easy to find the area of explanation.
5. This interface looked natural.
6. This interface helped me concentrate on the description.

7. This interface helped me concentrate on the area of the explanation.

The guidance and subjective questionnaire were repeated four times for the four interfaces, selected randomly. Finally, the participant has to select one of his/her preferred guiding interfaces (from the four, OF OT IF and IT). Note that both the demographic questions and subjective questionnaire were embedded inside the HoloLens application so that the participant did not need to take off the device and put on again in the middle of the experiment.

When the participant finished the experiment and questionnaire, the experiment proctor then asked him/her to complete a memory test on a computer. The test consisted of an image of the painting and 16 text boxes with a different phrase placed around the image. Each phrase was referred to during the explanation dialogue corresponding to the meaning for that part each area. The participant was then asked to draw an arrow from each phrase to its corresponding position. For an example, he/she had to draw an arrow from the phrase "To undertake a difficult task" to the position that related to the phrase which is an image of a man who is catching an eel on the right side of the painting. Participants were allowed to take as much time as they needed and could give up anytime. After that, each participant was interviewed regarding their impression of the guide, why he/she preferred the selected positioning and movement interfaces, and how well he/she could perform the test.

4.3.3 Experiment Results

The effects of each interface were evaluated on both objective and subjective measures. The objective measures included the time that participants spent on experiencing the AR guide system with each interface and a score from a memory recall test conducted after the participant used all 4 interfaces. A repeated measure Analysis of Variance (ANOVA) test [54] was used, following with a post-hoc analysis using pairwise t-tests [55] with Holm's adjustment [56] to analyze the objective measures.

Subjective measures were collected by asking the participant to rate each interface regarding obstructiveness, appropriate position, ease of finding the corresponding explanation, naturalness, and assistance to concentration using 5-point Likert scale [57]. A non-parametric repeated measures Friedman test [58] was used, following with a post-hoc analysis using Wilcoxon signed-rank test [59] with Holm's adjustment method. A significance level of $p < 0.001$, $p < 0.01$, and $p < 0.05$ were indicted using asterisks ***, **, and *, respectively, between each set of bars in the results plots and figures.

Time Used

The time for each guiding interface that each participant used to navigate the 4 parts (from the left to right side of the painting) is shown in Figure 4.6. Outside-Flying (OF) took the longest, with an average of 66.04 seconds, followed by Inside-Flying (IF) at 64.29 seconds, Outside-Teleport (OT) at 62.58 seconds, and Inside-Teleport (IT) at 62.53 seconds. An ANOVA was conducted first, which revealed a significant difference between interfaces, ($F = 5.99$, $p < 0.001$). Then the post-hoc analysis was conducted as a follow-up and found that mean time spent was significantly shorter with OT than with OF ($p < 0.0025$), and IT also required less time than OF ($p < 0.0025$).

Memory Recall

The number of correct matches between painting locations and descriptions varied from person to person. The average score for each group of descriptions and the difference of each interface are shown in Figure 4.7. Regarding guidance, participants answered the questions correctly at 58% with OT, 54% with OF, 53% with IF, and 50% with IT, respectively for each interface. Results from the ANOVA test indicated no significant difference was found between interfaces, ($F = 0.297$, $p = 0.827$).

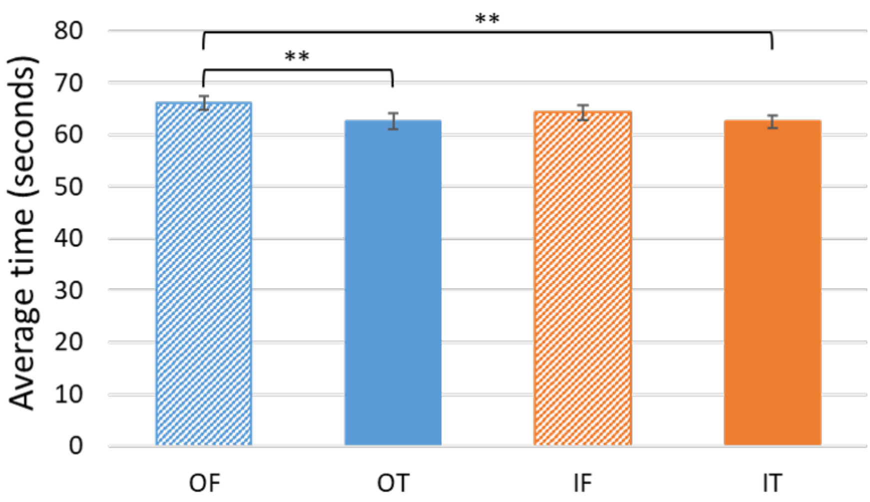


Figure 4.6: Average time that the participant used for with each guiding interface.

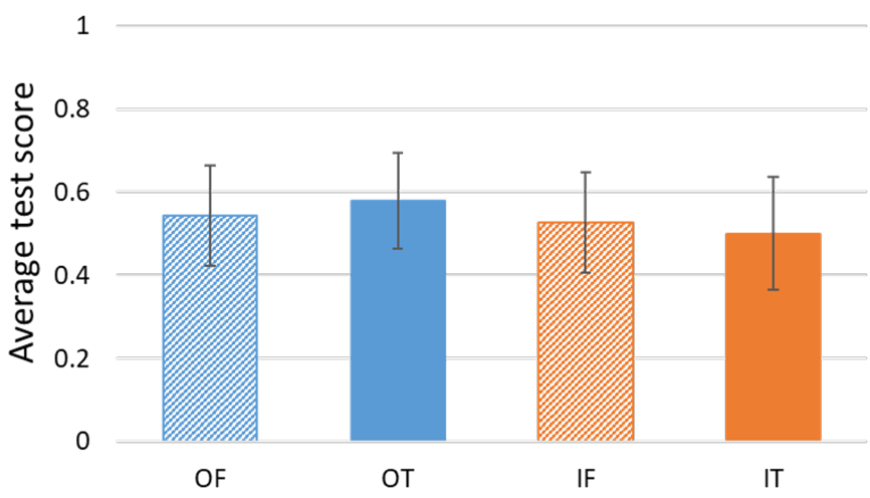


Figure 4.7: Average number of correct participant answers when they experienced each guiding interface.

Obstructiveness

In addition to objective measurements, the results of obstructiveness to the painting for the virtual guide and laser pointer for each interface are shown in Figure 4.8. With respect to obstruction, results showed that 28.57% of participants rated agree and 71.43% rated disagree on IT interface, whereas 14.29% rated agree and 78.57% rated disagree for the IF interface. None of them rated agree for the statement that the virtual guide with OF and OT interfaces were obstructive.

Regarding whether the laser pointer obstructed the area of explanation, 17.86% of participants rated agree and 82.14% rated disagree on OT interface. Both OF and IT interfaces 10.71% of participants rated agree and 85.71% rated disagree. For the IF interface, 3.57% of them rated agree and 85.71% rated disagree, while 10.71% rated neutral. A Friedman test was used and found a significant difference between methods for the virtual guide obstructiveness, $\chi^2(3) = 27.046$, $p < 0.001$. The post-hoc analysis result reveals the same level of significance ($p < 0.5$) between IF and OF, IF and OT, IT and OT, and IT and OF. However, no significant differences were found for the laser pointer obstructiveness, $\chi^2(3) = 6.252$, $p = 0.09$.

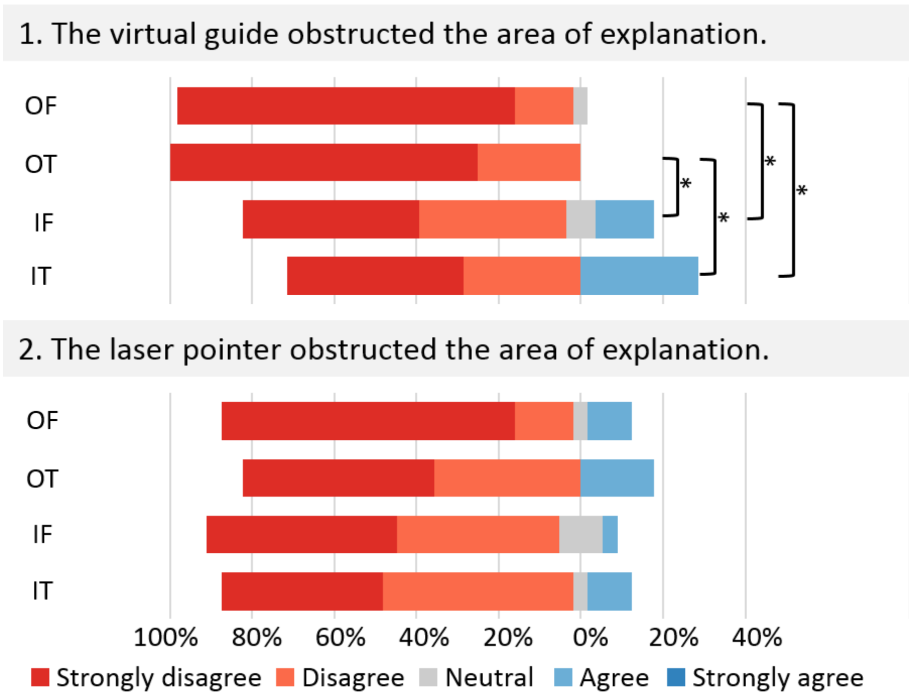


Figure 4.8: Subjective ratings for each interface regarding obstructiveness.

Interface Position, Ease of Finding, and Naturalness

Subjective results regarding appropriate position, ease of finding the explanation position, and naturalness of each interface are shown in Figure 4.9. Regarding a position of the virtual guide was proper or not, 60.71% of participants rate agree or above and 21.43% disagree or below on IF interface. With OF interface 57.14% of them rated agree and 28.57% disagree. With the IT interface 50% of them rated agree and 39.28% disagree. While the same 39.28% of participants rated both agree and not agree for the OT interface.

Regarding whether the area of explanation was easy to find, 89.28% of participants rated agree, while 7.14% of them rated not agree with the IF interface. For the OF interface, 60.71% rated agree and 25% rated disagree. For the IT interface, the number of the participants that rated agree and disagree are similar at 42.85% agreed and 39.28% disagreed. While 25% rated agree and 64.28% rated disagree to the OT interface.

Regarding whether the interface looked natural or not, 75% of participants rated agree and 7.14% rated disagree to the IF interface. With OF interface 67.86% of them rated agree and 10.71% rated disagree. While with the OT interface, both agree and disagree were rated equally 42.86% from the participants. For the IT interface, 42.86% agreed and 32.14% disagreed.

Results from a Friedman test found significant differences on ease of finding and naturalness of each guiding interface with $\chi^2(3) = 28.913$, $p < 0.001$, and $\chi^2(3) = 14.325$, $p < 0.01$, respectively. How ever no significant difference was found for the virtual guide position ($\chi^2(3) = 3.824$, $p = 0.281$). The post-hoc analysis revealed significant differences in the rating of ease of finding, and OT had a lower score compared to OF ($p < 0.01$) and IF ($p < 0.001$). Also, the IF interface has a higher score compared to IT ($p < 0.01$). Together with the naturalness rating result, the same level $p < 0.05$ of significant differences between OT and OF, OT and IF, and between IF and IT interface were found.

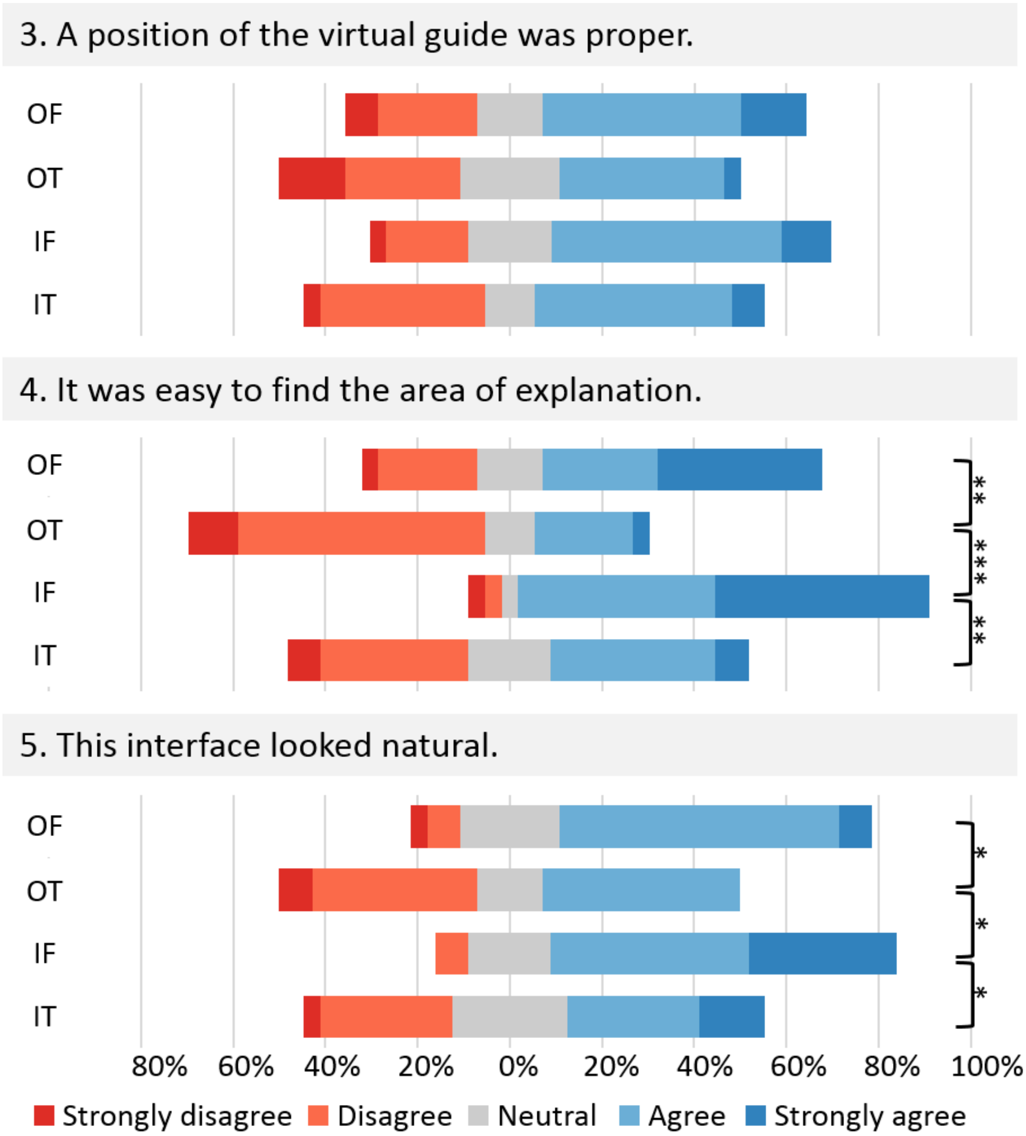


Figure 4.9: Subjective ratings for each interface regarding interface position, discovered position and naturalness.

Concentration

Results of concentration on the content description and area of explanation effected by each interface are shown in Figure 4.10. Regarding concentration on the content description, 89.28% of participants agreed and 3.57% disagreed with the OF interface. For the IF interface 82.14% of them agreed and 7.14% of them disagreed. A total of 75% of participants rated agree for the OT and IT interfaces. While 14.28% and 7.14% rated disagree on OT and IT, respectively.

Regarding concentration on the area of explanation, 89.28% of participants agreed with IF interface, but none of them disagreed. For the OF interface, 78.57% agreed and 10.71% disagreed. For the OT interface, 67.85% agreed and 17.86% disagreed. For the IT interface, 64.28% agreed and 10.71% disagreed.

Results from a Friedman test found a significant difference on the area of explanation concentration ($\chi^2(3) = 11.751$, $p < 0.01$), but no significant effect on the content description concentration ($\chi^2(3) = 5.2241$, $p = 0.156$). The post-hoc analysis revealed significant effects on aiding concentration of the explanation area between IF and IT ($p < 0.01$) and between IF and OT ($p < 0.05$).

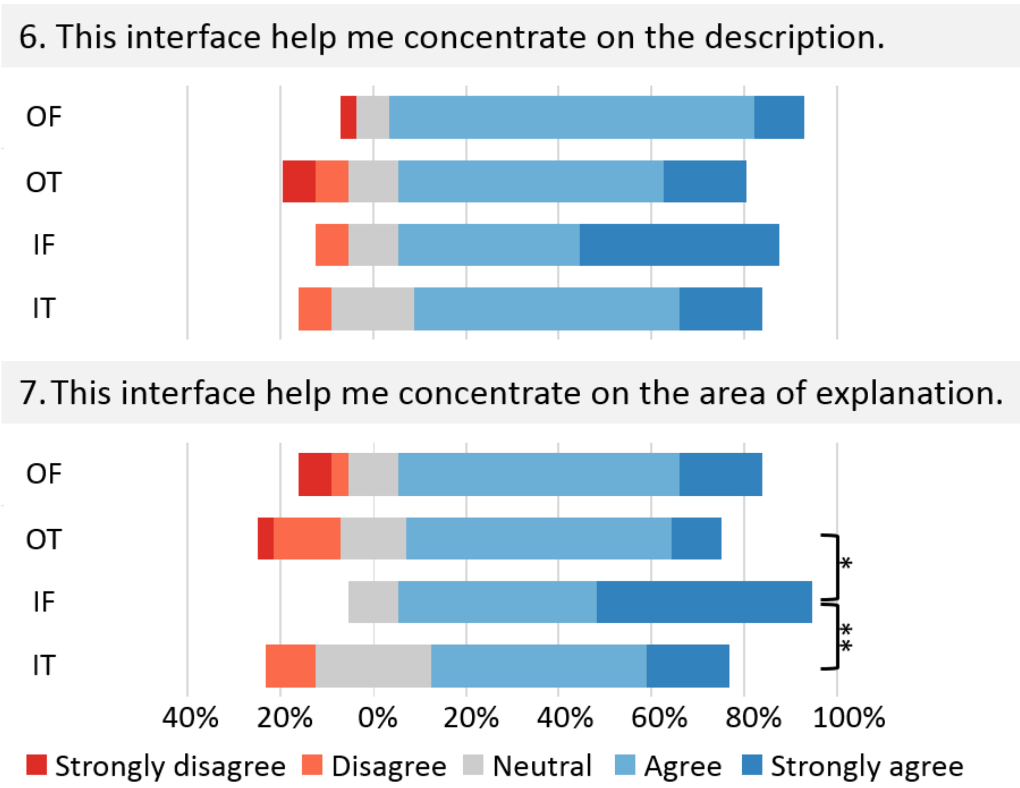


Figure 4.10: Subjective ratings for each interface regarding concentrate on the explanations and area in concern.

Interface Ratings

The participants were asked to choose the interface that they preferred most at the end of all questionnaires. Twelve participants (42.86%) preferred the IF interface, eleven (39.29%) preferred the OF interface, three (10.71%) preferred the IT interface and only two (7.14%) preferred the OT interface. Results shown in Figure 4.11 correspond to the previous subjective rating results that were voted for the most. As the IF and OF interfaces had many positive significant differences in several areas including easy to find, looked natural, and help in concentration.

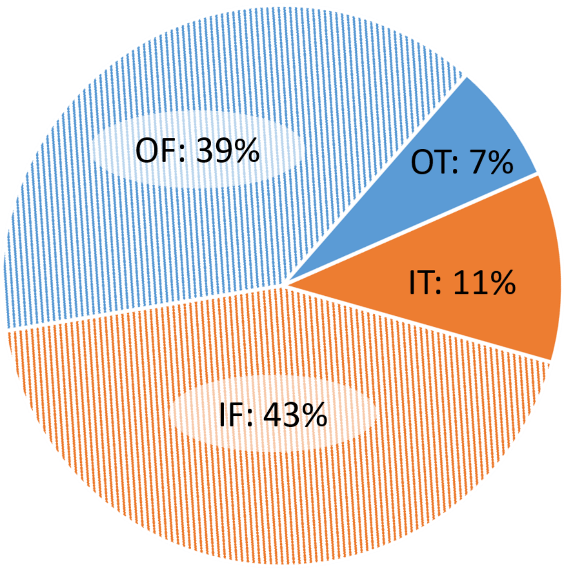


Figure 4.11: Plot showing the number of participants that ranked a particular method as the most preferred.

4.3.4 Discussion

In the evaluation, when the virtual guide is positioned inside the painting (IF and IT interfaces), a few participants agreed that the virtual guide obstructed the area of explanation, while most other participants disagreed. This evidence partially supports **RQ 1**, with the hypothesis that users do not mind if the virtual guide's position overlaps the painting.

Regarding the average time used by each interface, both teleportation interfaces (IT and OT) had a lower average time used. Therefore, **RQ 2** was supported, in which the teleportation type movement interface helped reduce guidance time. However, these types of interfaces received many negative subjective ratings. Based on the interview, some participants mentioned that when the guide disappears, they didn't know where it would reappear. Also, they had to find the virtual guide position again most of the time when using the teleportation type interfaces because there are no hints as to where the new position of the virtual guide would appear. The results from the questionnaires also confirmed that the teleportation interfaces are neither proper in position nor do they help the user find the area in context, and they do not look natural. I suggest that although the virtual guide is able to teleportation in the real world, this is not a preferred interface for most of the participants. Therefore avoiding its use or creating a movement path or destination information for the agent may be appropriate so that the user will not lose the guide's position as the guidance progresses.

Although the IF interface was rated that it help the participant concentrate on both content description and area of explanation, but scores from the memory recall test indicated that it does not outperformed any other interfaces on memory retention. Therefore, **RQ3**, which states that the different type of position and movement interface of the virtual guide does not affect memory recall of the image-related content, is rejected. Based on the interview, many participants told us that there were too many items introduced so they could not remember them all. Many times they only remembered either the image or the phrase. Also, some participants said that they just enjoyed the guide system but were not interested in the painting, so they did not try to remember the content. Some participants mentioned that they did not like the outside position interfaces because they had to trace the virtual guide pointing to the image, while the inside position interfaces made it easier to find the part of the image because the virtual guide and the target position were in the same view.

By observing participants movement during the experiment as shown in Figure 4.12, I found that participants who preferred IF and IT interfaces mostly stood around their starting position. There was only one participant who preferred the IF interface that walked around the designated area. Six of eleven of participants who preferred the OF interface and one from two of the OT interfaces walked around the area, while half of the remaining participants stood around the start position.

This indicates that if the virtual guide moves a small distance (e.g. inside the painting area) the user will not move much from his/her standing position. However, if the virtual agent moves for a certain distance (e.g. from the left to right side), the user tends to follow that movement.

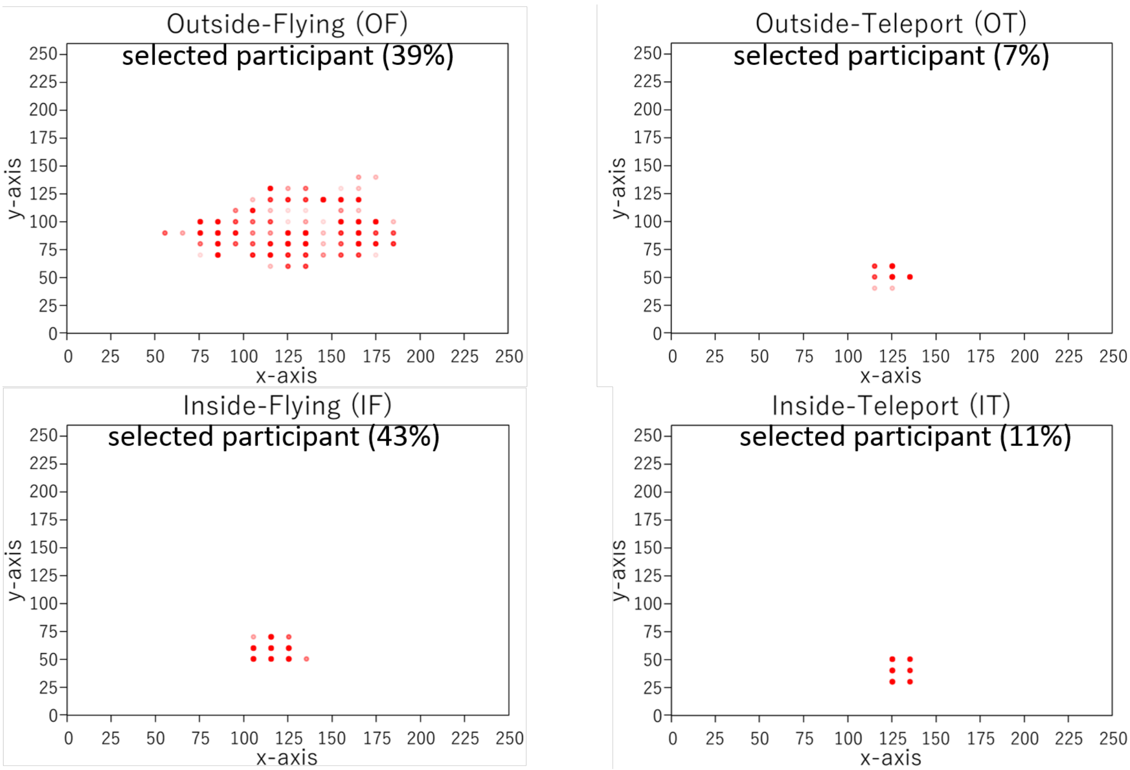


Figure 4.12: Plot showing movement of a sample participant.

4.4 Conclusion

In this chapter, I proposed an AR guidance system for multi-section visual information using an virtual guide. I also implemented two positional interfaces and two movement interfaces for the virtual guide. I developed the system that tracks the location of a large piece of art and locates the position of regions of interest, after which I provided guidance to the corresponding location. The system included an virtual guide that moves and points toward the position, describes where and at what to look, and explains information for a specific part of the image.

Base on results from the user study, most of the participants preferred the Inside-Flying type, with the Outside-Flying pulling a close second since the flying interface was easier to track the position of a virtual guide than the teleportation interface. Many participants were not concerned that the virtual guide obstructed the art when it stays in front of the painting as long as it does not occlude the part at which the user is looking. I found that the virtual guide that guides by moving inside the area of the painting helped participants find the position in the description immediately without having to trace the laser from the virtual guide's hand to the area in the description. This inside positioning interface also helped the user concentrate on a particular area of the painting. Regarding outside positioning interface, movement to another far position of the virtual guide affected the user's movement.

In future studies, I would like to explore guiding interfaces for areas with obstacles such as other visitors to provide a dynamic guide system that automatically avoids or adapts its position based on its location and the order of descriptions.

Chapter 5

AR Embodied Conversational Agent Effectiveness for Information Guidance

5.1 Introduction

Many studies on Embodied Conversational Agent (ECA) have claim that the ECA has merit in either improves user performance or increases positive attitude to the task. Such as the users felt the explanation of technical information is less difficult and more entertaining with the presence of the ECA image, however no effect on comprehension [37]; The users can remember more words significantly when the information is talking by the ECA compared to the sound only [38]; The users are more satisfied and not felt pressured by the ECA over an expert human in a healthcare related document explanation [39]; and The ECA increases learning engagement, concentration, content understanding, and learning repetition of the student in e-learning system, while there is no significant difference in test scores [40].

However, there is no study comparing effectiveness of using ECA in museum guide situation. Also, in museum guiding different types of guiding interface may affect the need of presence of ECA. Therefore in this study, I have developed several guiding interfaces for the AR information guiding system to help assist the explanation by pointing out the specific location of information throughout the painting. Then, I conducted a user study to evaluate each guiding interface with and without the presence of the ECA-type virtual guide. I assumed that any guiding interface with the presence of the virtual guide would increase the user experience rating of the tour guide and memory retention about the exhibit contents.

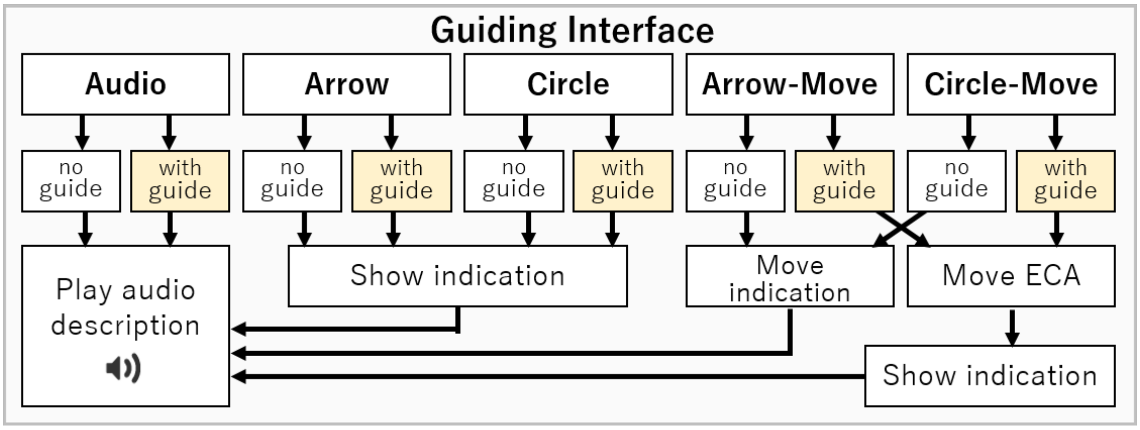


Figure 5.1: Flow of the guiding interfaces.

5.2 Guiding Interfaces

In this section, I have developed several guiding interfaces to help indicate the position or region of content descriptions.

The three guiding interfaces include the Audio interface, which is a baseline comparison, Arrow interface, which shows the direction from a reference position to the target position, and Circle interface, which specifies the sub-region, are implemented as static guiding interfaces. I then add a moving condition to give a clue to a user when the system points to the next item with the Arrow and Circle interfaces. Thus, there are five guiding interfaces in total including Audio, Arrow, Circle, Arrow-Move, and Circle-Move.

Each guiding interface is also combined with or without an ECA presence condition as shown in Figure 5.1.

The interface with the EA condition will show the agent either at the reference point, an initial position of the EA located at the left side of the painting, or a point near the item to be described. Also, the ECA function including make eye contact, make a gesture toward the direction of the content position. The image of each interface and the ECA presence condition are shown in Figure 5.2.

Figure 5.2: Actual view through HoloLens of each guiding interface.

5.2.1 Audio interface

The Audio interface is created to simulate a traditional audio guide that provides in many museums. Although it does not provide any graphical indication to a sub-region inside the painting, the audio description tells the user where the region is in the painting, and what to look at. One of the description audio used in the guiding is: “At the left-bottom corner, you will see a man who is banging his head to a wall. This part of the painting represents the phrase, to bang one’s head against a brick wall, which means to do things repeatedly even though you have no chance of success.” The same audio descriptions are used throughout other interfaces.

5.2.2 Arrow interface

The Arrow interface was developed to specify the item position located in the painting by drawing a red line with an arrowhead. The Arrow interface draws a line from a reference point toward the position of the region in concern while playing the audio description. The Arrow interface with EA condition shows the EA at the reference point and draws a line form the virtual guide fingertip toward the item position.

5.2.3 Circle interface

In some situation that a piece to describe is located on another side of the painting, the Arrow interface would draw a line across the painting that hinders the user view. Therefore the Circle interface was created to compromise this problem. This interface plays the audio description and draws a circle with the size of 10 cm radius around the region in concern at the same time, this way it will not occlude the main pointing position and the user can focus on that region.

Both Arrow and Circle interfaces have their own strength and shortcoming, the

Arrow interface provides a reference point to look back and trace a new pointing position. However, looking back and forth every time for a new description is considered redundancy in time and makes the user tired. The Circle interface (without EA) helps the user focus in the content area but lacks a reference point. To trace the next position of the circle indication, the user has to look for it in every time that the next information was introduced. Therefore, a movement reference type interface was developed as an extended version to the Arrow and Circle interfaces to help the user navigate to the next item easier.

5.2.4 Arrow-Move interface

This interface draws 10 cm length of a horizontal align line with an arrowhead. The arrow indication gradually moves from the current position to the new pointing position with the speed 10 cm per frame, i.e. around 40 cm per second. The direction of the arrowhead is determined either left or right side based on the direction of the next pointing position. After the arrow indication reaches the position in concern, the audio will be played.

5.2.5 Circle-Move interface

The Circle-Move interface is similar to the Arrow-Move interface. It will move the circle from the current region to the next region with the same speed to the Arrow-Move interface and then play the audio description.

In case of Arrow-Move and Circle-Move interfaces with EA condition, firstly, the virtual guide will move close to the description region with the same speed of the indication's movement speed, normally took time 1–3 seconds depends on a distant. After reaching the position near the region to be explained, the system activates the virtual guide function and shows either arrow or circle indication, and then let it say an explanation.

5.3 Evaluation

To investigate the effects of the ECA regarding several types of guiding interface, a laboratory study comparing with and without the ECA presence condition on an individual interface was conducted.

5.3.1 Hypotheses

The first hypothesis is that different guiding interfaces will require different guiding times. While Audio, Arrow, and Circle type interfaces play the image description immediately, Arrow-Move and Circle-Move require slightly more time to show the movement interface. Regarding memory retention possibility, since graphical interfaces provide an overlaid indication to the area of description, I assumed that the user would easier found the exact position and hence remember the correct position in the content. Also when guiding with the ECA, it would able to attract the user attention. Therefore overall user experience rating for the system would be rated higher than non-agent condition, especially on attractiveness and novelty scales. Accordingly, I formulated the following hypotheses:

- **H1**: The time used for movement-type interfaces will be longer than the Non-movement-type interfaces.
- **H2**: Guiding with any graphical indication interfaces will result in higher memory recall than guiding with the Audio-type interface.
- **H3**: Guiding with the ECA will result in a higher user experience rating compared to the non-agent presence condition.

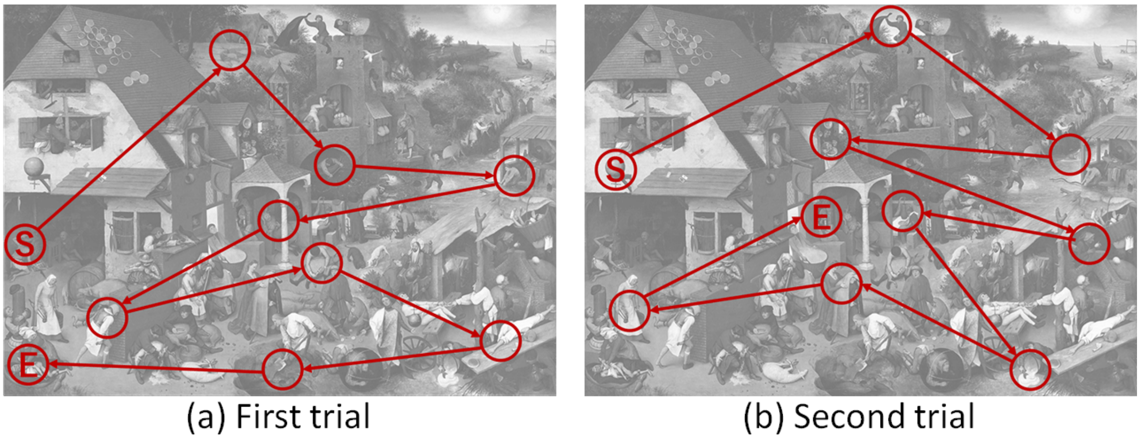


Figure 5.3: The positions for explanation that were used in the system evaluation.

5.3.2 Experiment Setup

This AR guide system was tested with the painting named "The Netherlandish Proverbs" which contains many proverbs hidden in sub-details. It is suit for evaluating the system in terms of memory recall effect and user experience rating. This image was printed at 200 by 140 cm and hung it on a wall 100 cm above the floor. A 250 by 260 cm free walk area in front of the painting was provide to facilitate viewing, as same as in the Chapter 4.

Twenty different subsections of the images were prepared, with sets of two sections group. The guiding is divided into 2 sections starting from the first trial to the second trial, in the same order sub-sections to every participant. The trial order and position of each sub-section used in this evaluation are shown in Figure 5.3. The circle with the letter 'S' inside indicates the start sub-region position to be described, the next guiding region is depicted by the arrow that draws from the current one to the next one, the end position of the guiding area for each trial is depicted by the circle with the letter 'E' inside.

An oral explanation for each part describes the location of that part in the painting, the image characteristic to look at, and the meaning of that image. The explanation audio ranged from 11 to 16 seconds, with an average of 13.55, SD of 1.7 seconds.

In this experiment, 40 participants were recruited (20 female, 20 male, mean age of 25.87, SD of 10.36) to help evaluate the effectiveness of the guiding interfaces. Participants were both university students and people who visited Osaka university Toyonaka campus during an open house day. They are paid with a bookstore card valued at 500 yen (around 5 USD) for participating in the 20 minute experiment. Participants were assigned into 5 groups for testing individual interface regarding the presence of the ECA.

I have adopted a User Experience Questionnaire [60], a seven-stage scale of 26 items using for measuring attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty of the system as shown in Table 5.1. Each measuring item is a set of positive and negative terms in a random side of a 7-point Likert scale. For example “annoying/enjoyable”, “friendly/unfriendly” and so on are part of items using for measure the attractiveness scale. I included this questionnaire inside the AR guiding system as a popup window between each guiding trial.

I also prepared a matching test of the image region and meaning for testing memory retention of the users as shown in Figure 5.4. The application contains 20 questions in a different order to the guiding system, shows one by one and 40 sub-regions to be selected (20 regions are introduced by the system and others are dummy answers).

5.3.3 Procedure

First of all, the experiment conductor let a participant sign a consent form, answer demographic questions and bring them to the experiment space. The participant will stand 230 cm away from the center of the painting. At this position, he/she could see the overlaid graphics in a corresponding physical area of 130 * 70 cm. Then the participant was let to wear the HoloLens and confirm visibility of the virtual guide at the left side of the painting which is a standby position.

Next, the experimental procedure was explained to the participant that the experi-

Figure 5.4: Image of a memory recall test built by Unity game engine. The user has to select a correct image that related to the description above.

Table 5.1: Questions in each perspective of the user experience questionnaire.

Attractiveness	**Perspicuity**
annoying/enjoyable	not understandable/understandable
bad/good	difficult to learn/easy to learn
unlikable/pleasing	complicated/easy
unpleasant/pleasant	confusing/clear
unattractive/attractive	
unfriendly/friendly	
Efficiency	**Dependability**
slow/fast	unpredictable/predictable
inefficient/efficient	obstructive/supportive
impractical/practical	not secure/secure
cluttered/organized	does not meet expectations/meets expectations
Stimulation	**Novelty**
inferior/valuable	dull/creative
boring/exciting	conventional/inventive
not interesting/interesting	usual/leading edge
motivating/demotivating	conservative/innovative

ment will divide into 2 parts which guiding with the virtual guide and without the virtual guide in a randomized order. The system is controlled by pressing a button on a wireless controller to listen to the next item explanation. After 10 items were introduced, a build-in-app questionnaire popup will appear so that he/she has to rate their perspective of the current guiding interface. Then the system will guide another 10 segments of the image with the other condition. The experiment conductor also informed the participants that they had to do a matching test about this painting after they finished listening to both parts.

5.3.4 Experiment Results

Time Used

The average time that participants used to navigate 10 items around the painting for each guiding interface is shown in Figure 5.5. The line on the top of each interface bar indicates a standard error of mean. The Audio interface took the shortest time at 155.73 seconds, following with the Audio interface with the ECA presence condition at 156.62 seconds, and the Circle-Move interface took the longest time at 168.19 seconds, following with the Arrow-Move interface at 168.18 seconds. A two-way ANOVA analysis

Memory Recall

The average number of correct matches between painting locations and meanings is shown in Figure 5.6. Participants who are guided with the Audio interface with the ECA presence condition got the lowest correct answer at 38.75%, following with the Audio interface 42.50%. The highest correct answer score is from participants who guided with the Circle interface with the ECA presence at 67.75%, following with the Circle interface and Circle-Move interface at 60% of the correct answer. However, a result from two-way ANOVA analysis shows no significant difference neither interface type ($F = 2.002$, $p = 0.103$) nor the agent presence condition ($F = 0.019$, $p = 0.892$)

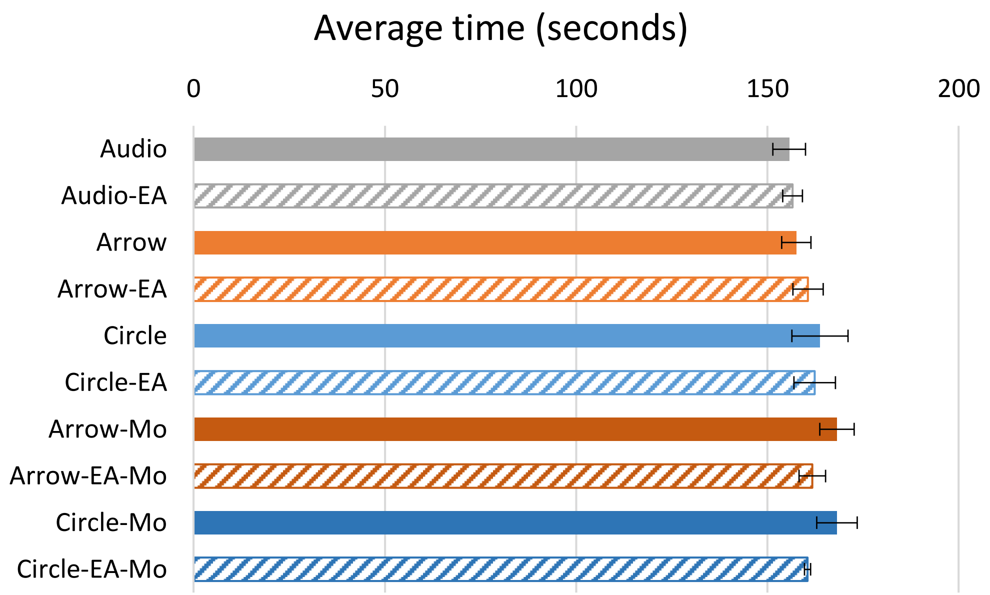


Figure 5.5: Average time that participants used during each guiding interface.

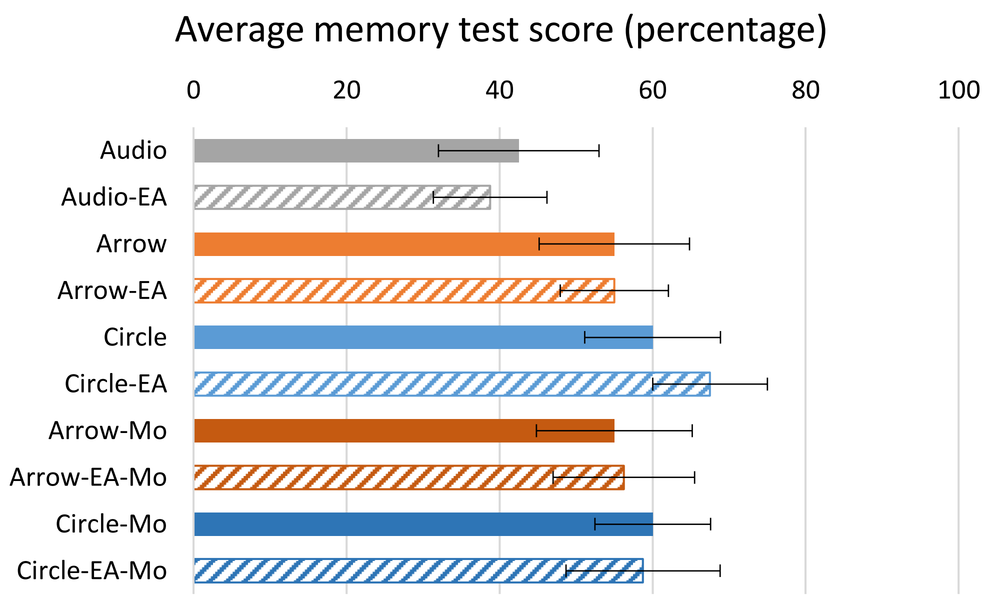


Figure 5.6: Average percentage that participants answer correct when they experienced each guiding interface.

User Experience Rating

An analysis of variance of Aligned Rank Transformed Data was conducted to find a significant difference for the 2-way non-parametric analysis [61]. The results revealed that there are significant differences between different interface types for attractiveness ($F = 3.278$, $p < 0.05$) and stimulation ($F = 3.298$, $p < 0.05$), and between agent presence condition between attractiveness ($F = 10.006$, $p < 0.01$), perspicuity ($F = 5.369$, $p < 0.05$), stimulation ($F = 13.616$, $p < 0.001$), and novelty ($F = 16.432$, $p < 0.05$).

Non-parametric pairwise comparisons using a Wilcoxon rank sum test [62] with Holm's adjustment [56] were conducted for testing the differences between subjects for the five guiding interfaces. There was a significant difference found between the Audio and Circle interfaces on the attractiveness ($p = 0.041$) and a near significant difference on the stimulation ($p = 0.056$).

Within each individual guiding interface, a non-parametric pairwise comparison using a Wilcoxon signed rank test [59] was conducted to test the with and without ECA presence conditions. Although the ECA presence condition has rated higher in the most scale for every interface except Circle interface. Significant differences were found on some scale including: Audio interface regarding novelty ($p = 0.024$), Arrow interface regarding attractiveness ($p = 0.020$), Arrow-Move interface regarding attractiveness ($p = 0.036$) and stimulation ($p = 0.022$), and Circle-Move interface regarding attractiveness ($p = 0.039$), stimulation ($p = 0.013$), and novelty ($p = 0.034$).

Subjective rating results regarding attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty scales are shown in Figure 5.7. The line on the top of each interface bar indicates a standard error and an asterisk * between bar indicates a significant difference ($p < 0.05$).

5.3.5 Follow-up Interview

In addition to testing users who expect to be a museum visitor, I also wanted to get feedback from a museum curator who conducts museum guide tours (a female of age 36 with 3 years of guiding experience). The interview was conducted face-to-face after the guide tour finished. I explained to her that the target of the work is to compare the effectiveness of the virtual guide through several guiding interfaces and videos of each interface. Then she was asked the following questions:

- Is the system useful in the museum guidance and how?
- Do you think the virtual guide should presence in the guiding system?
- What is the defect of the current system and how it can be improve?
- What kind of issue you found in your working and how the system can help?

The museum curator commented on the usefulness of the system, stating that it can be used anytime when guidance is limited due to a museum curator's time. She also mentioned a difference between the system and a human guide, which was that the human can conduct a conversation or answer some questions to the visitor in the mean time. It would be great if the guiding system could do that. Regarding the virtual guide presence, she preferred to have to virtual guide stay next to the item because it can attract interests of the visitor to that position. The arrow moving interface was considered to cause confusion since it does not exactly highlight a particular area, while the circle would be better. Also the guiding interface can disappear when the visitor looks at the position or when time passes. Regarding issues during work, she mentioned that when guiding she doesn't know the background knowledge of the visitors with respect to the exhibit. Also when there are different ages of people joining the same group such as kids and elders, guiding with the same style would lead some visitors to cease following the guide's description. In order for the guiding system to perform many guiding styles, adaptation to different people is needed.

5.3.6 Discussion

I found that although movement-type interfaces required more time to see the animation, time used for these interfaces was not significantly difference from the other interfaces. Therefore I reject the hypothesis **H1**. I also found that the ECA presence condition decreased the time used in the movement-type interfaces. This is because the agent size is bigger than the interface and its hand point to the interface again, hence it has a shorter movement distance and movement time required.

While there was no significant difference found between each interface in the both time used and test score, the average memory recalls score of Arrow and Circle type interfaces were all higher than the Audio-type interfaces. Therefore this result is only partially supported the hypothesis **H2**. I suggest that testing on more subject would help clarify the result.

Regarding user experience rating results, interfaces with the ECA presence condition had a significant difference in attractiveness, stimulation, and novelty scales. However there was no effect on perspicuity, efficiency, and dependability scales. Therefore the hypothesis **H3** is only partially supported.

I also particularly looked into the question items including "annoying/enjoyable", "unpleasant/pleasant" and "obstructive/supportive", which revealed some negative effects since the ECA may have disturbed comfort during the appreciation of the painting. I found that only Circle-type interface has a lower rating on the ECA presence condition (5.50/5.37 for enjoyable, 5.62/5.50 for pleasant, and 5.50/5.25 for supportive, each for absence/presence conditions of the agent, respectively).

Based on the interview, 75% of participants prefer the ECA over the non-ECA condition with various reasons, such as "It was enjoyable", "It is more natural to have ECA, so that I know where the voice is come from", "It helps me find to subregion easier since itself is easier to find", and "I feel like there is someone explain the content for me, so I must listen".

Most of the negative opinions for the guide with the ECA are from participants who try non-movement interfaces including Audio, Arrow, and Circle interfaces that the agent is standing at one side of the painting. Their comments are “I always look back and forth between an avatar and region of content that made me hard to concentrate”, “I’m interested in this painting. Having an avatar nearby distract me from the painting.”, and “I didn’t look to the avatar at the corner, so there is no need of it”.

Based on the finding, the need for the ECA depends on either the agent visibility of the users, and the personal interests of the exhibit. If they prefer to explore the arts thoughtfully by themselves, having the virtual guide may distract them. In case of the non-movement interfaces that the virtual guide is outside the user field of views, 37.50% of participants who used these interface said that they do not prefer the ECA presence. In contrast, regarding the Arrow-Move and Circle-Move interfaces that the ECA condition locate near the description region, 93.75% of participants who used these interfaces prefer to have to the ECA guiding them. Also, all scales in the user experience result of these movement-type interfaces confirm effective of having the ECA.

5.4 Conclusion

In this chapter, I proposed an AR guidance system for multi-section visual information using an ECA-type virtual guide and designed five types of guiding interface: Audio, Arrow, Circle, Arrow-Move, and Circle-Move. All interfaces were designed to help the user find positions of interest inside large scale works of art more easily. Each interface is can be combined with the ECA presence condition to increase the AR tour guide experience. For the experiment, the positions of the regions to be explained within a tracked image was located, and the system guided users to one of these regions. To explain information for a specific part of the image, the selected guiding interface indicates the position inside the painting and describes where and at what to look, and then gives an explanation about that piece. I evaluated the effectiveness of the each guiding interfaces and the presence of the virtual guide by comparing with and without the presence of the ECA. The results showed that guidance with the ECA either increased the attractiveness, stimulation, or novelty of the system. Based on the interview, over 75% of participants preferred to have the ECA present and thought it was attractive, natural, or make him/her feel able to concentrate on the content. However, no significant differences in memory retention appeared, which is likely more dependent on the interest in the painting or content itself.

One limitation of the present study is that I tested only with a particular painting. In the future, I would like to test the system with different artworks to evaluate the generalizability of the system. I would also like to extend the system to support navigation guidance and test it at a real museum.

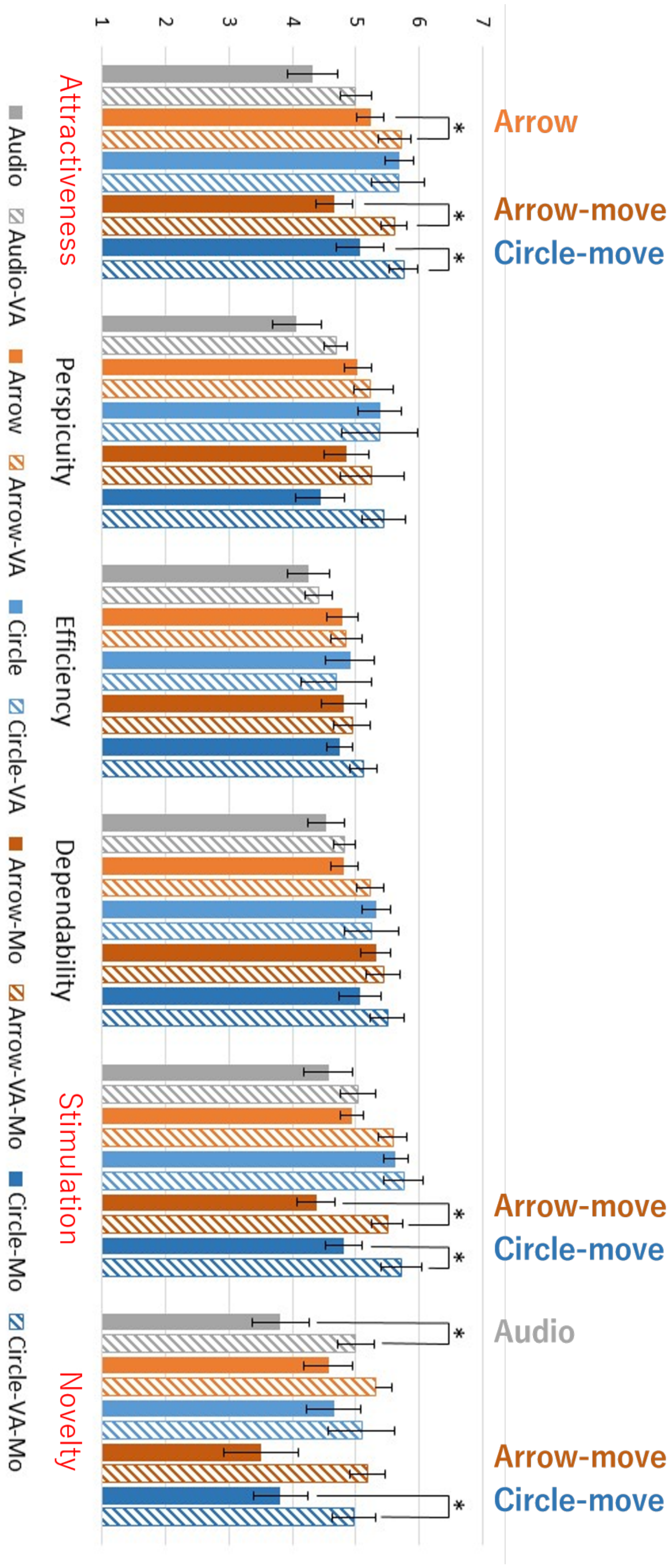


Figure 5.7: Subjective ratings for each guiding interface.

Chapter 6

Guidance and Visualization of Optimized Packing Solutions

6.1 Introduction

Packing, defined simply for the purposes of this dissertation as the process of arranging smaller items into a larger container, is both a complex and prevalent issue in logistics management due to the fact that it is an NP-hard problem [63]. Ensuring proper packing solutions allows workers, and hence the company, to optimize shipping and transportation costs [64]. In mathematics, many heuristic algorithms [65–67] have been developed to solve for the optimal solution of packing problems. These algorithms have sometimes been applied for large-scale packing operations, for example when loading thousands of containers onto ships or airplanes. For smaller scale packing problems such as those faced by consumer-oriented delivery or local shipping carriers [68], fewer methods are currently available for logistics and optimization. Moreover, real-time presentation or instructions have not been applied or tested.

Overlaying a task solution visualization onto the workspace helps reduce both completion time and the number of errors [69, 70]. Such a real-time support or worker

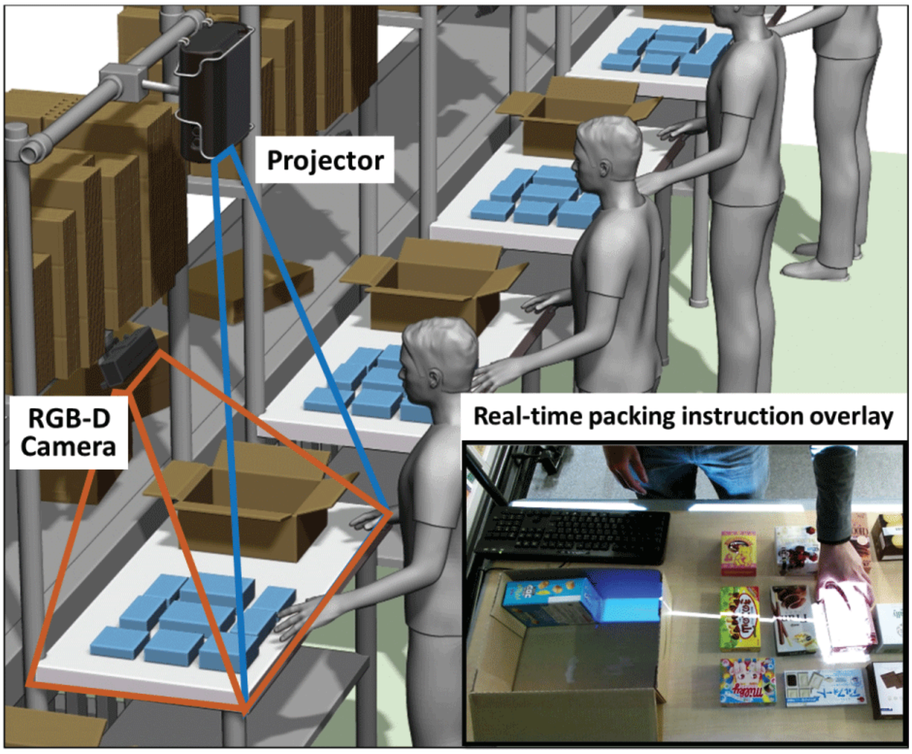


Figure 6.1: Model of the use case for the packing support system in a typical fulfillment center (top). Projection onto a sample workspace with virtual overlays provides real time instructions to guide users to the optimal packing solution (lower right).

training system would be useful to individual workers and scalable across the entire supply chain. As a potential solution, I have developed a real-time packing guidance system using Spatial Augmented Reality (SAR), a branch of Augmented Reality (AR) that involves overlaying computer-generated information onto the real world using projection techniques [71].

This projector-based visualization is considered the least burden to the worker because the user does not need to hold any AR display devices, and thus the interface becomes hands-free. Also, compared to the HMD device, a smaller HMD is considered less of a burden to wear. However, there is a trade-off between size and field of view. Considering a packing task in which the user does not move too much but needs to look around the workspace to find and pick the item, a visualization from a projector that is able to cover a large area of the workspace would be more suitable.

The image in Figure 6.1 provides an example of the system's functionality and operation for assisting a warehouse shipping task. The hardware for the system primarily consists of two parts, the first being an RGB-D camera that can ascertain and segment packing items. This camera is coupled with a projector that projects packing guidance onto the user's workspace. The software includes several primary methods to facilitate support. First, color and depth information of the items to be packed and a target container are acquired using the RGB-D camera. Next, object segmentation and dimension estimation are computed using the segmented point cloud, and finally the optimal packing solution is found using the bin-packing algorithm proposed by Baltacioglu et al. [72]. I have also designed two packing overlays (virtual instructions) called Rotation Instruction (RI) and Movement Instruction (MI). These overlay techniques are then combined with the packing optimization algorithm into a single real-time system. When the user begins packing, step-by-step visualizations of package placements and orientations are projected onto the workspace.

To test how well these visualizations could assist users, I conducted experiments to evaluate the performance of the system by comparing two overlay methods and the unsupported packing process. Results showed that overlays reduced packing time and the number of moves required per object versus a baseline, but that no single type of instruction significantly outperformed any other. Contributions of this study include:

- the hardware setup and implementation of the packing support system as a whole,
- algorithms that obtain and segment objects and output an optimized packing solution in real time, and
- the design and evaluation of two types of virtual overlays that guide the user during the packing process.

6.2 Related Work

6.2.1 The Packing Problem

In mathematics and computing, the packing problem is typically described as the general goal of packing smaller items into a larger container, and the problem can often be multi-dimensional. For example, a simple one-dimensional packing problem is disk partitioning on traditional computing systems [73]. Two-dimensional packing problems would include tasks such as efficient cutting of wood or metal plates [74]. Three-dimensional packing problems are most typically found in logistics, and the goal is almost always to reduce the cost of packaging materials, vehicles, and fuel, hence the overall cost of transportation. This class of problems is typically categorized by input minimization and output maximization [75] based on the type of logistics problem being solved. Input minimization considers the packing all of the items using the smallest number of containers possible (as e.g., bin-packing), whereas output maximization seeks to pack the largest possible subset of items into a single container (e.g., the knapsack problem).

Packing algorithms are usually executed in two steps, including container selection and packing position determination. A sorting step is sometimes included since the ordering can potentially increase packing efficiency, as presented in [65]. In the case where more than one container is used, a method such as branch and bound [76] or heuristics such as first-fit-decreasing or best-fit-decreasing [65] are often used for selecting a container. In order to decide an item's position inside a container, placement heuristics such as Wall-building [67], Layer-building [72], Corner points [76], and Extreme points [65] have been developed in the past.

Though wall-building and layer-building approaches pack items using a guillotine partitioning method, this does not necessarily utilize the entire volume of the container. Both the corner point and extreme point methods are generally able to utilize a container space more effectively, but have a non-rotation constraint. Therefore, for cases with the same input, the wall building and layer building methods with a rotation constraint would provide a more optimal packing solution.

Most of packing solutions are typically provided in a format such as a text file [65], [76], which normally contains a percentage of bin utilization and details of each packing item that are listed by their dimensions, position, orientation, and container number. By providing these data, it is easy to compare efficiency between different packing methods by listing them in a table [75]. However, in practical use, it is hard for humans to comprehend what a particular packing solution looks like or how it was ordered or packed. Erick et al. introduced a method using OpenGL to draw three-dimensional computer graphics for simulating a packing solution [66]. Observing a graphical simulation not only allows a user to better identify free space in a container or how to pack the item, but also helps developers and researchers check their packing algorithm's precision and efficiency. Moreover, with the visual and interactive capabilities of a simulated model, the user is able to understand and learn a more disciplined approach to solve 3D packing problems.

Nowadays, 3D simulation of packing solution services are available in both non-commercial [77] and commercial [78], [79]. However, existing systems display their solutions on a flat monitor. Thus, the user needs to switch his or her attention between the screen and the workspace, which can be time-consuming and prevent spatial learning. Alternatively, displaying the solution onto the workspace can potentially be more efficient for both time, effort, and learning of the packing task.

6.2.2 Spatial Augmented Reality in object manipulation

SAR has a great potential to increase efficiency by allowing people to view the entire work space efficiently without wearing any device. One usage of SAR is to guide a local user to move an object around a working space. For example, Tsimeris [80] developed a number of SAR visual cues that convey translation and rotation information to instruct object arrangement. Adcock et al. [22] used semantic information of object physical properties, by drawing a shape of the manipulating object at the work area. Uranishi et al. [23] used a grid-pattern to help both instructor and user and to indicate and identify the position of the object. However, these systems generally use translational or 2D rotation cues, and require manual operation on the part of an instructor or professional user. This motivated us not only to propose 3D rotation cues for manipulating objects using SAR, but also to automate the process of optimizing a packing solution.

6.3 System Design

The projection-based packing support system is designed to help users pack items into a container as densely and quickly as possible without the assistance of a remote collaborator. A flowchart outlining the entire process flow is depicted in Figure 6.2. The system first starts by using the overhead RGB-D camera to acquire workspace and object information as inputs. I then use an object segmentation algorithm to estimate each object's dimensions and calculate a packing solution. Finally, the packing solution is displayed to the user in a step-by-step fashion by projecting the instructions (visualizations) directly onto the work area and updating after every completed step.

When testing a prototype system, I selected items and a container that were rectangular solids, which are often found in a number of real packing scenarios because they can be used to minimize open space in a container in comparison with other shapes. During experimentation, all items were initially placed orthogonally to the table and with enough space between them to facilitate segmentation so that the scanning process only has to happen once, as opposed to requiring the user to scan each item individually.

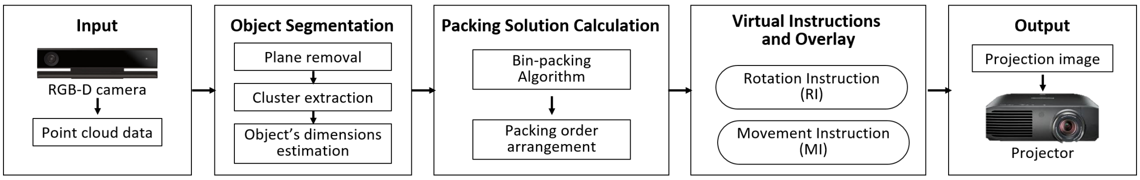


Figure 6.2: Flowchart of each step of the system from raw input to final output.

6.3.1 Object Segmentation

In this step of the process, information about each packing item from an input point cloud image is extracted and divided into individual point cloud segments, and calculated dimensions through the Point Cloud Library [81].

First, the table surface was estimated from the input point cloud, as shown in Figure 6.3(a), by using the Random Sample Consensus (RANSAC) algorithm [82]. RANSAC is an iterative method that estimates parameters of a mathematical model from the data, which in this case is used to fit all points from the point cloud into possible plane models. After numbers of iterations are done, it returns the plane model of best fit as well as the points fitted to that model, referred to as inliers. A sample result of the planar model from the RANSAC algorithm is shown in Figure 6.3(b), red color points are the group of inliers included in the plane model, and all other points are outliers. The estimated plane parameters a, b, c, and d of this inliers group in $ax + by + cz + d = 0$ form are -0.035, -0.377, 0.925, and -0.8263, respectively.

The plane inliers (table) from the point cloud were removed so that only objects that align on table remain. This allows us to use the Kd tree search method [83] to find clusters of point cloud data that represent individual objects. The results of a sample cluster extraction are shown in Figure 6.3(c), where different groups of clusters are shown in different colors. In each cluster, the biggest point cloud segment is selected as the container. Other remaining clusters are labeled as packable items, each item and container cluster were stored with a reference (x,y,z) position relative to the workspace. Finally, the smallest bounding box possible was assigned to each cluster and set each object's dimensions equal to its bounding box size, as shown in Figure 6.3(d). In order to correctly orient and display visualizations, Eigen values and vectors were also extracted for each object, and use them to rotate the object to align the axis in the target container. Note that in the experiments, 1-2 mm of error was found in the estimation of each bounding box dimensions from the real object, therefore a manually adjusted of the dimensions was conducted create more accurately represent the items being tested.

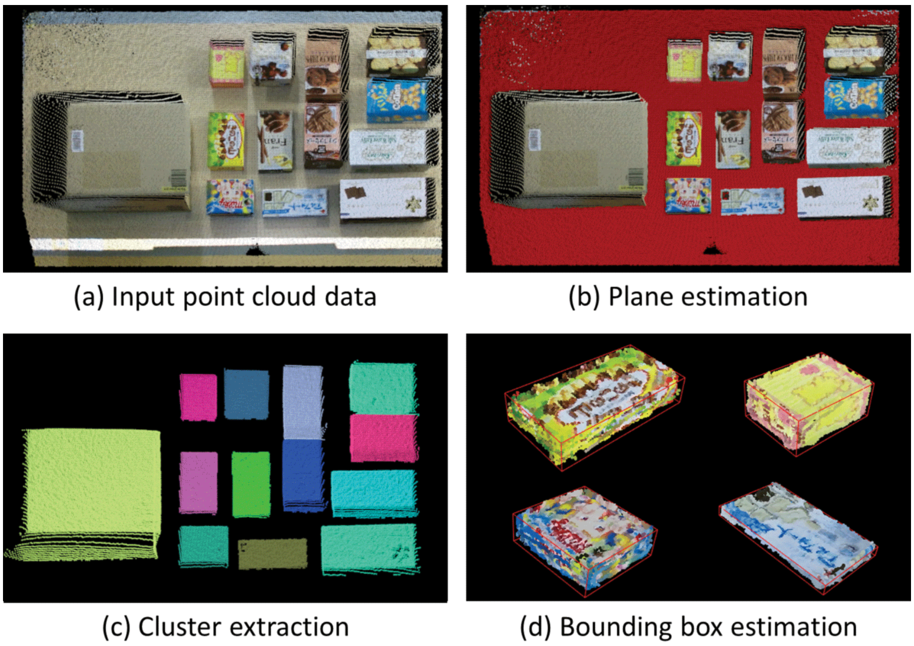


Figure 6.3: The object segmentation process. Showing the capture of point cloud data, estimation of the workspace plane, segmentation of the container and packing items, and bounding box estimation with overlaid textures.

Figure 6.4: Simulated visualizations of a packing solution showing segmented point cloud data fit into the container.

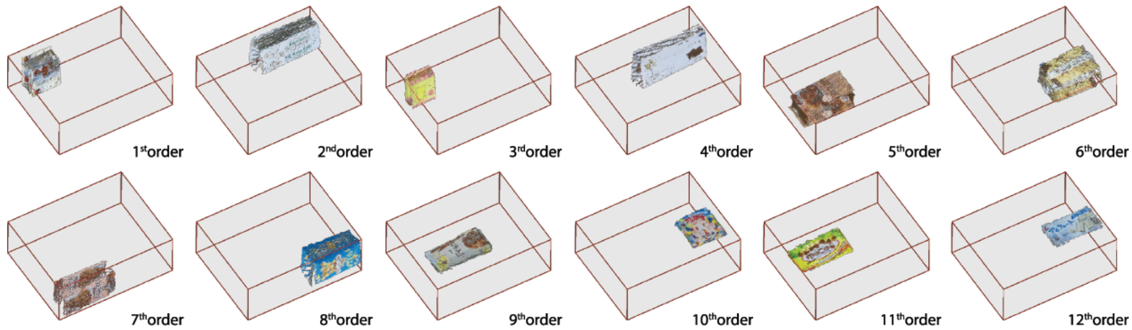


Figure 6.5: The sequential packing solution at each step with the red bounding box signifying the outer regions of the container.

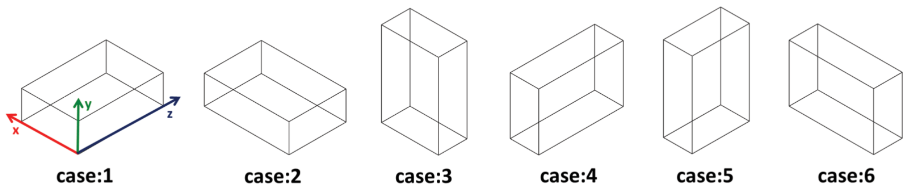


Figure 6.6: The six possible rotations for any packing item, which determine how the algorithm handles the instruction at a particular step.

6.3.2 Packing Solution Calculation

To find the densest packing solution for a particular container, the bin-packing algorithm by Erhan et al. [72] was applied, which is a heuristic that employs a layer packing approach and packs rectangular boxes in any orthogonal orientation. The x, y, and z dimensions of the container, total number of items, and lists of the item's dimensions are used as input to the algorithm. Outputs from the algorithm include the x, y, and z positions and dimensions of each item to be packed in the container.

A visual representation of one of these packing solutions along with each packing step is shown in Figure 6.5.

The process behind the packing algorithm is described as follows. First, I defined n as the total number of items and $\{d_{i1},\ d_{i2},\ d_{i3}\}$ as the three dimensions of a particular item i. I then defined *DIM* as the set of unique dimensions for all items as calculated by the following formula:

$$DIM = \cup_{i=1}^{n} \cup_{j=1}^{3} d_{ij} \tag{6.1}$$

The variable m represents the total number of members in *DIM*, and k is the index of members ranging from 1 to m. I then defined *VAL* as a set of the average free space values for each DIM_k. The variable c_i is set to the dimension from item i that is closest to the *DIM* value, and the value of c_i is from $\{d_{i1},\ d_{i2},\ d_{i3}\}$. Each member of *VAL* is calculated by using the following formula:

$$VAL_k = \sum_{i=1}^{n} |DIM_k - c_i| \tag{6.2}$$

I then constructed, *LAYER()*, an array of candidate layers height using the following formula:

$$LAYER() = \cup_{k=1}^{m}(DIM_k, VAL_k) \tag{6.3}$$

Afterwards, *LAYER()* members are arranged by increasing value of *VAL*. A small *VAL* value represents lower remaining free space between packing item and layer height *DIM*. I used the first *DIM* of candidate layers that has the smallest *VAL* as the first layer height. The algorithm then repeatedly tries packing solutions by selecting items that can fit into the layer along X dimension and repeats placement operations into the current layer along Z dimension until the layer surface is full. When no space remains to pack items into the current layer, a new upper layer along Y dimension is created, and the process is continued for the current layer. This bin-packing algorithm also tries to pack items with six different orthogonal rotations of the container as depicted in Figure 6.6, then compares each rotated container utilization and selects the container rotation that has the best packing utilization as the optimal solution.

After obtaining the optimal packing solution, I arranged the order of packing items as determined by the algorithm. Simply put, I followed the rules stating that lower item positions should be packed before upper item positions and that packing order should start from one side to opposite side of the container.

Now that the system could automatically obtain an optimal packing solution, an intuitive way to present this to the user was needed. In order to effectively convey the item's destination packing position and orientation, packing instructions should be both easy to understand and at the same time indicate sequential progress for the item to pick up. I then went on to design two different types of virtual overlays to convey information, as described in the following section.

6.4 Virtual Instructions and Overlay

As a user progresses with the packing task, steps toward reaching the optimal solution need to be displayed item by item, in real time, from start to finish.

In the initial design, I used point cloud images that were captured during object segmentation process to visualize a packing solution as depicted in Figure 6.7. However, this type of interface has many demerits. Due to its pixel-based image, the projected image is hard to see, especially when it is overlaid on a textured item. Moreover, if the packing solution shows the rotated point cloud image, the user will see a sparse point cloud because it did not be captured from the single view of the RGB-D camera.

Therefore, I proposed another two user interfaces, called Rotation Instruction (RI) and Movement Instruction (MI). Representative images of each of these techniques

Figure 6.7: An Initial design of the packing instruction shows position of item to pick up, and final position and orientation of the item to be packed in a container.

are shown in Figure 6.8 and 6.9, respectively. These methods are interactive that the user can advance or return via keyboard.

Prior to rendering, item and container positions and dimensions were obtained to locate the physical positions on the user's workspace. Container positions and dimensions in the real world are then used as reference positions for the final packing solution obtained from packing solution calculation described previously. The coordinate system depicted in Figure 6.10 was used to project item positions and orientations, where the positive X-axis is a horizontal line aligned with the workspace from right to left. The positive Y-axis is a vertical line passing through the center of the workspace upwards into the air. The positive Z-axis is a horizontal line through the center of the workspace pointing away from the user.

6.4.1 Rotation Instruction

In some situations, packing items may have similar faces and dimensions, so it can be hard to distinguish differences in orientation simply by observing images from the virtual overlays. For example, consider a plain, square box of a single color in which one side is only slightly longer than the other. By only looking at the item image itself, users may feel uncertain as to which orientation is actually being indicated. To solve this potential ambiguity, the Rotation instruction was designed, which uses graphical images to help identify the entire transformation of the packing item. For any given packing task, there are six possible orthogonal rotations in three dimensions of any object, which are shown in Figure 6.6. Similar to algorithmic solutions to Rubik's cubes which are divided into specific rotations, these six possible orthogonal rotations into two-dimension rotations between the X-, Y-, and Z-axes were classify as follows:

- case 1: no rotation
- case 2: rotate 90 degrees at Y-axis
- case 3: rotate 90 degrees at X-axis
- case 4: rotate 90 degrees at Z-axis
- case 5: rotate 90 degrees at X-axis and Y-axis
- case 6: rotate 90 degrees at Z-axis and Y-axis

In other words, a combination of rotations on any two axes can cover six different possible target orientations for any packing object.

The visualization of RI technique consists of 2 steps of rotation. The first step uses a red or blue arrow to indicate a rotation along X-axis (to flip an item up/down) or Z-axis(to flip an item left/right). The second step uses 2 comparable images to indicates a rotation along Y-axis. If there is a rotation only at Y-axis, the instruction will show only the second step image as shown in Figure 6.11 (a).

The first rotation step is images that show on the left side. A white rectangle indicates an initial orientation of an item. A yellow rectangle (left side) which overlays on the white rectangle position that indicates a final orientation of the item, a red and blue line at the corner of the rectangles show X and Z-axis direction, and a curved arrow indicates which direction to rotate the item. A left side of Figure 6.11 (b) and (d) shows a red curved arrow that guides a user to rotate the item 90 degrees around the red line or X-axis, from the white rectangle position to the yellow rectangle position. A left side of Figure 6.11 (c) and (e) which have a blue curved arrow, prompt the user to rotate the item 90 degrees rotation around the blue line or Z-axis.

The second step indicates a rotation along Y-axis, which shows two yellow rectangles on the left and right side of the image together with the first step images. Both yellow rectangles have black arrows in the middle to signify an orientation. The left side of yellow rectangle shows an orientation before rotating the second step and the yellow rectangle on the right side shows the final orientation of the item. Figure 6.11 (a), (d) and (e) which have 90 degrees different orientation of the yellow rectangles between the left and right side, prompt the user to turn the item clockwise or counter-clockwise (90 degrees rotation around Y-axis).

Figure 6.8: Rotation Instruction when projected to the workspace.

Figure 6.9: Movement Instruction when projected to the workspace.

6.4.2 Movement Instruction

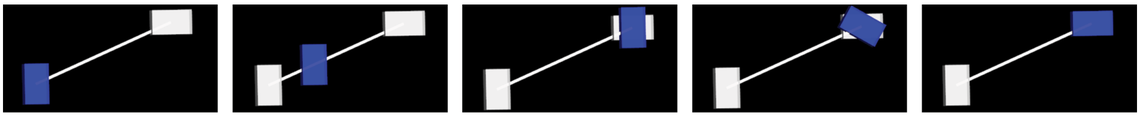
In contrast to the RI technique in which the user needs to infer intermediary steps between the start and finish, they lacked of motion cues. Therefore, I developed one final technique called Movement technique that shows smooth movement of the item from its current position to the target position. Designing of this overlay has a hypothesis that users would be able to follow the instructions without additional cognitive effort to interpret translations or rotations. Frames from the sequence of one of these moving objects is shown in Figure 6.12.

White rectangles show the current object's initial and final packing positions, and blue rectangle show the object's movement and rotation. This type of instruction first shows the object moving along a white leader line to the packing position. After the virtual object reaches the target position, it then gradually rotates 15 degrees along each orthogonal axis all the way to the correct orientation. After observing

Figure 6.10: Calibrated coordinate plane.

(a) Rotation at Y-axis (b) Rotation at X-axis (c) Rotation at Z-axis (d) Rotation at X and Y-axis (e) Rotation at Z and Y-axis

Figure 6.11: Rotation Instructions: in each sub-image, the yellow box with blue/red target rotation markers on the left is projected over the physical object's original position, and the right image is projected in the destination container at the target orientation and position.



(a) Initial pose (b) Intermediate translation (c) Final translation (d) Intermediate rotation (e) Final pose

Figure 6.12: Object Movement Instructions: (a), (b), and (c) show item in an image is moving from a current position to a target position. (c), (d), and (e) show item in an image is rotating from a current orientation to a target orientation.

this movement, the user can replicate the same translation and orientation into the target location.

The blue rectangle will start moving from the initial position after 0.5 second of an input key pressed by the user. Every 0.1 second it will update its position which is 5cm away from its last position along the white leader line. When it reach the final packing position, it will start rotating 15 degrees at every 0.1 seconds from Y-axis following with X or Z-axis. One axis rotation takes 0.6 seconds and two axes take 1.2 seconds. On average, items are 50 cm away from the container, so that the movement overlay would take average time around 2 seconds to show the packing cues for one item.

Finally, a user evaluation to determine the, efficiency, trade-offs, and subjective ratings of each technique were conducted.

6.5 Evaluation

To evaluate the effectiveness of real time support and compare two visualizations, a series of experiments to test packing performance were conducted. The goal was to assess the overall performance and subjective preferences of the two types of proposed instructions and to compare the results with unassisted packing.

6.5.1 Experiment Setup

To select compact packing tasks, I have collected various size of 50 snack boxes and several parcel boxes. Size of these boxes was measured manually. I then input the size information of each parcel box and all snack boxes into the bin-packing algorithm. The most two compact packing parcels are selected as the experiment tasks (refer as Task:1 and Task:2).

Task:1 has a filling rate of 87.60% consisted of 12 individual items to be packed into a container with dimensions of 32.0 cm × 11.5 cm × 25.0 cm as shown in Figure 6.13. Task:2 has a filling rate of 84.02% consisted of 17 individual packing items to be packed into a container with dimensions of 32.5 cm × 21.0 cm × 23.0 cm as shown in Figure 6.14. A single optimal solution provided by the system foe each task is shown in Figure 6.4 and lists of packing items are shown in Table 6.1.

The experiment was conducted in two parts: the first was a pilot experiment to create a baseline data of no-assistance packing, and the second was a within subjects condition to evaluate the two user interfaces. In the first run, 12 participants (average age of 22.33, ranging from 21 to 24) were asked to complete packing for Task:1 and Task:2 without any assistance. Initial boxes positions were the same between participants, and the task was complete when all items were packed into the destination container. In the second run, 24 participants (average age of 23.67, ranging from 20 to 29) were included to conduct packing for Task:1 and Task:2 with the RI and MI overlays in random order per participant.

Figure 6.13: Image of the packing items at the start of Task:1

Figure 6.14: Image of the packing items at the start of Task:2

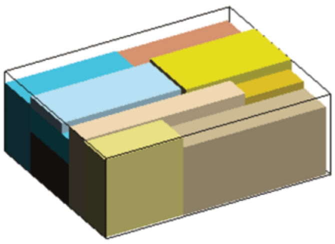


Figure 6.15: Simulated images of Task:1 optimized packing solution.

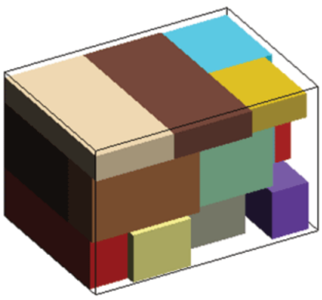


Figure 6.16: Simulated images of Task:2 optimized packing solution.

Table 6.1: List of packing items.

Task:1		Task:2	
Item	Dimensions (mm)	Item	Dimensions (mm)
1	158 × 98 × 63	1	214 × 70 × 56
2	159 × 96 × 63	2	152 × 70 × 80
3	160 × 60 × 95	3	140 × 65 × 127
4	157 × 60 × 95	4	220 × 68 × 53
5	230 × 114 × 37	5	75 × 69 × 85
6	84 × 106 × 35	6	38 × 65 × 100
7	203 × 97 × 53	7	160 × 95 × 60
8	111 × 111 × 52	8	160 × 95 × 60
9	163 × 25 × 90	9	158 × 95 × 60
10	116 × 35 × 95	10	150 × 95 × 121
11	150 × 30 × 89	11	160 × 90 × 24
12	170 × 15 × 77	12	60 × 95 × 157
		13	49 × 95 × 133
		14	230 × 37 × 114
		15	230 × 37 × 114
		16	116 × 35 × 95
		17	106 × 35 × 84

6.5.2 Procedure

Before starting the experiment, the experiment conductor explained each participant how to interpret the RI and MI instructions, and how to select the next or previous instruction by pressing the right or left arrow keys. Times used for an object segmentation which automatically determine items' longer and shorter side of tasks 1 and 2 were 67.10 seconds and 82.12 seconds, respectively. The packing solution calculation time was 0.48 seconds and 0.26 seconds for tasks 1 and 2, respectively. To make a more fair comparison, I have added these pre-processing times to the packing time for each task.

During the experiments, the experiment conductor measured time and move count for each item. After completing the two packing tasks, the participants to fill out a subjective questionnaire are asked.

6.5.3 Hypotheses

In general, I hypothesized that packing time would be reduced when assisted with the visualizations since trial and error would not be necessary to find a solution. On the other hand, I thought that the total number of moves required would be relatively similar despite taking more time. With respect to visualizations, I thought that the RI technique would be most efficient on shortening the packing time because it provides a 3D visualization of the item rotation, whereas users need to wait for the animation with the MI technique.

Regarding preference, I thought that the MI technique would be most preferred since it provides a packing example closest to the real world objects, but with the other technique the user has to take some effort to interpret the visualization before insertion. Accordingly, I formulated the following hypotheses:

- **H1**: The proposed packing support system and visualizations will result in reduced packing time and number of item moves compared to the no-guidance condition.
- **H2**: The Rotation Instruction (RI) will result in the lowest packing time when compared to other visualizations.
- **H3**: The Movement Instruction (MI) will receive the highest subjective ratings for ease of understanding, usefulness and satisfaction level when compared to other visualizations.

6.5.4 Experiment Results

This section presents a comparison of results for packing without support from the system (WO) and each of the two packing interface techniques (RI and MI).

I have analyzed packing time and the number of moves per item. I have also gathered subjective user data for each type of virtual instruction, which included ease of understanding, instruction usefulness, and satisfaction level.

Packing Time

I combined measured time to completion of each set of instructions with the time used in Object Segmentation and Packing Solution Calculation process. The total times are presented in Figure 6.17, which shows that the proposed techniques (RI and MI) reduced packing time when compared to the non-assisted (WO) condition. Reductions in time by using RI and MI were 57.89% and 55.63% in Task:1, respectively, and 30.32% and 34.26% in Task:2, respectively.

A parametric measures Analysis of Variance (ANOVA) test [54], followed with post-hoc analysis using pairwise t-tests with Holm's adjustment [56] were conducted to evaluate time to completion results. In Task:1, significant differences found between the WO, RI, and MI overlays with $F = 15.21$, $p < 0.001$, and in Task:2, significant differences between the WO, RI, and MI overlays also found with $F = 6.65$, $p = 0.003$. These results are summarized in Figure 6.17, where the asterisk ***, **, * symbols between each set of bars indicates a significance level of $p < 0.001$, $p < 0.01$, and $p < 0.05$, respectively.

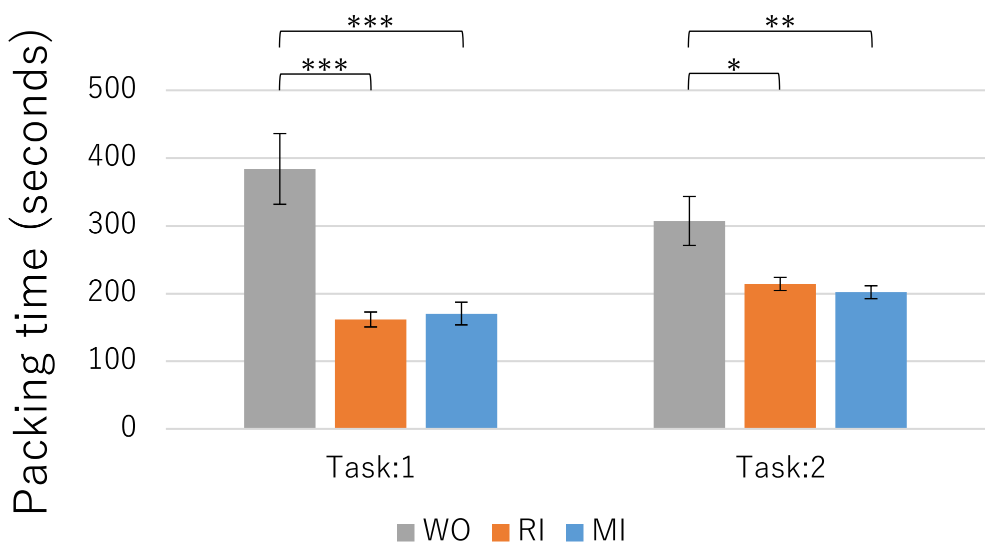


Figure 6.17: Average packing time by method (in seconds). WO indicates without instruction, RI indicates with rotation instruction, and MI indicates with movement instruction.

Number of Item Moves

Afterwards, the number of item moves was analyzed as presented in Figure 6.18. The chart shows that both RI and MI visualizations significantly reduced the number of total item moves, by 86.91% with RI, and 85.12% with MI in Task:1, and by 76.39% with RI, and 76.71% with MI in Task:2. The same analysis used for packing time was also conducted to evaluate item moves. Significant differences of $F = 39.19$, $p < 0.001$ for Task:1 and $F = 33.83$, $p < 0.001$ for Task:2 are found.

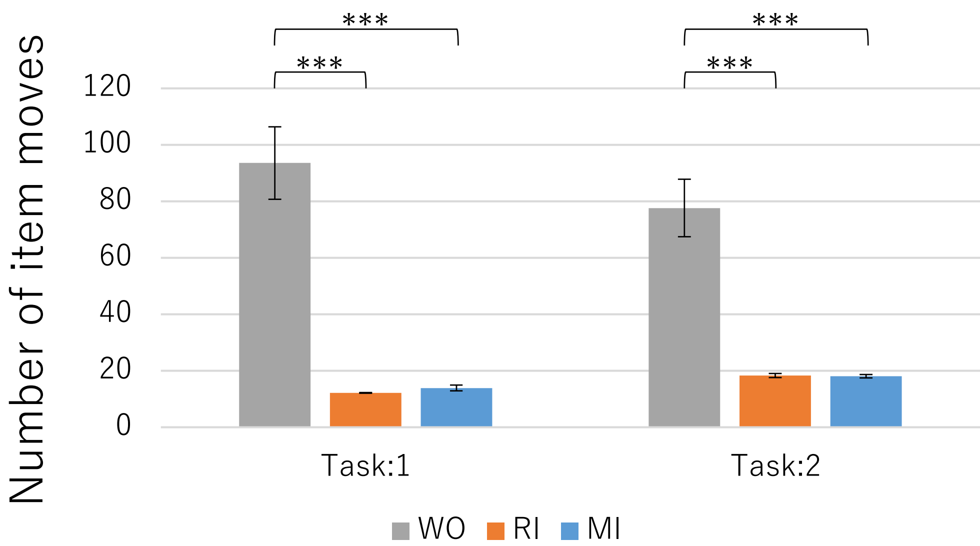


Figure 6.18: Average number of item moves by method.

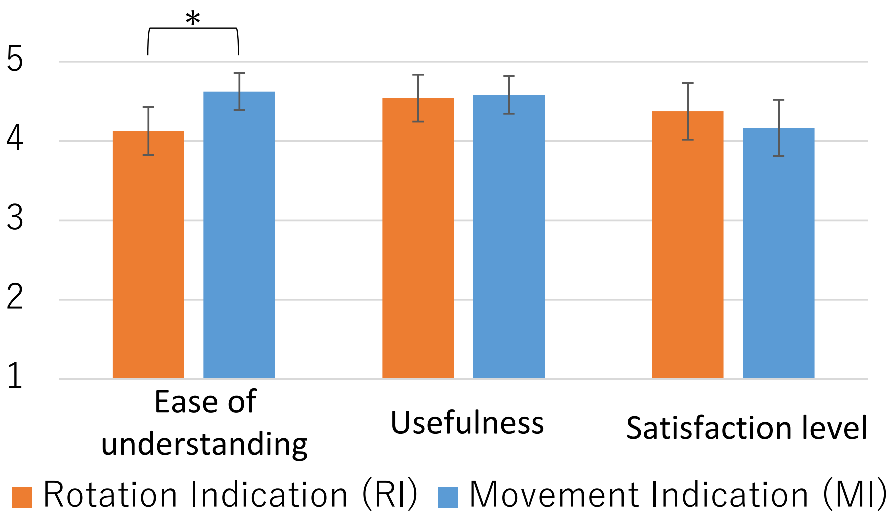


Figure 6.19: Subjective ratings from the questionnaire in the main evaluation.

Questionnaire

Finally, a questionnaire with a 5-point Likert scale was conduct to rate subjective opinions about the interface and overlays from the most negative response to the most positive response. Regarding each packing instruction, participants were asked the following 3 questions: "1. Is this packing indication easy to understand?", "2. Is this packing indication useful?", and "3. How much were you satisfied with this packing indication?". The summary of ratings are shown in Figure 6.19.

A Wilcoxon signed rank test [59] was used to evaluate questionnaire results. The results reveal a significant difference only on the ease of understanding ($p = 0.045$), which is indicated by the asterisk * symbol between each set of bars in the graph.

6.5.5 Follow-up Interview

In addition to testing typical users, I also wanted to get feedback from a person who was experienced in packing related problems. I interviewed a product designer (a 31 year old male with 3 years of design and problem solving experience) who was responsible for solving product damage and return issues that occur when shipping to a customer. The interview was conducted by face-to-face meeting. I explained that the goal of work was to reduce labor work and time for packing items into a parcel, and showed videos of each interface. Then he was asked the following questions:

- Is the system useful in a packing for your workplace and how?
- What is the defect of the current system and how it can be improve?
- What kind of issue you found in your working and how the system can help?

The product designer commented on the usefulness of the system, stating that it can help reduce time in thinking and trying packing, so that the worker just packs one time. However, the current system visualization projects 2D images that are not clear enough to see. Also the visual indication alone leads to confusion in finding the indication. Having sound indicate the characteristic of the item would improve the system in terms of reducing effort to find the projected visualization. Moreover, having packing process detection and automatically showing the next instruction would reduce time required to press the button to show the next step. Regarding issues found in his work, items to be packed have various sizes such as a small plastic box or a longer mop, so there is empty space in the parcel, which could lead to the item moving and breaking inside the box. Also each product has a different weight-restraint since putting the fragile one under the heavy one could lead to damage. Therefore having simulations of weight balance and object collision when the box is moving would help the worker safely pack products, which will reduce damage from delivery.

6.6 Discussion

In the evaluation, the results suggest that packing without support from the system took much longer on average than with the system. This evidence supports the hypothesis (**H1**), which suggests that the system is overall beneficial for packing support tasks, and can already be used practically.

Although Task:1 (12 items, filling rate 87.60%) has items to pack less than Task:2 (17 items, filling rate 84.02%), unassisted packing time for Task:1 was longer than Task:2 because of the higher filling rate. When a more compact of packing is needed, the packing is more difficult, thus the system would efficiently help reduce both packing time and number of item moves.

On the other hand, no significant results for the comparison of packing times between the two proposed visualization techniques were found. Therefore, the hypothesis (**H2**) is only partially supported, which suggests that the relative benefits of each individual method are quite similar.

Regarding user preference, even though rating on ease of understanding was significantly higher MI compare with RI, no other significant data was found on the other ratings. Accordingly, the hypothesis (**H3**) was only partially supported.

During the experiment, some participants took time more than others in the same condition due to mistakes and the time taken to think about how to rotate items. Such as, in the RI condition, when looking at the overlay, some users did not immediately know the direction to rotate, so they have to try rotating 2-3 times to make sure the direction was correct. Also in both two conditions, some participants did not carefully follow the instructions, so they incorrectly rotated the object. When continuing to pack later, the remaining items could not fit the container, so they had to go back several steps to correct the mistake.

Opinions from each participant about general use during the experiment were also collected. For the RI instruction took time to understand for a approximately half of the items, though with the latter half, participants had already started to become familiar with the instructions. In the MI condition, some participants said that seeing the image of the items moving was fun, while some said the movement was too slow, so they had to wait before they could pack the item. Also, some participants mentioned that the projections did not exactly fit the packing position, so they had to adjust the packing position by themselves. This feedback suggests that one primary improvement needs to focus on is the reduction of errors for the calibration of the projection surface. Sometimes the projected image was slightly distorted when the projection was mapped to a non-planar surface and color information was skewed due to projection on an existing colored/textured surface. This was especially true when objects of different heights were present in the workspace.

6.6.1 Future Work

To help automatically check for individual insertions and removals before proceeding to the next packing object, individual object tracking may provide a way for the system to enable backtracking and perhaps more intelligent instructions. Though manual interaction gives the user more control to some extent, more adaptive visualizations could help alert the user that their packing was not correct, facilitate training, develop better packing habits, and further reduce mistakes. Future work includes improving the calibration of the surface area, updating the RI instruction, and an object tracking approach rather than mapping point cloud data.

6.7 Conclusion

In this chapter, I have proposed a new projector based packing support system and three types of visualizations to help convey instructions. The system works by recognizing an external container and packing items in a flat workspace, and then performing object segmentation to extract both the items and container dimensions. Afterward, the system uses all input items to calculate a packing solution, which consists of packing positions and orientations, and the order and placement of each item are projected onto the workspace as a user completes the packing task.

To help evaluate spatial augmented reality (SAR) as an assistive tool, two types of visualizations were tested, including the Rotation and Movement based instructions, and compared these to non-assisted packing. Experimental results showed that the proposed packing techniques significantly reduced average packing time and movement requirements for packing tasks, with decreases average in task time up to 57.89% with Rotation instruction, and 55.63% with Movement instruction.

I hope that this research will bring packing support systems one step closer to practical use in the transportation industry and pave the way for further iterations of spatial augmented reality research.

Chapter 7

Conclusion

7.1 Summary of Findings

The goal of this work is to explore the best representation for guidance information of a guiding system using AR in terms of user performance and user preference. In this dissertation I have explored a variety of guidance information representations and found trade-offs between each of the design concepts.

In a learning scenario like the museum used in these studies, regarding the first study, having a virtual guide lead a user to interesting areas in the exhibit from 60% at the first time to 100% in the end of the experiment. Also over 60% of participants were satisfied with the slightly lower quality of the artwork seen through the HMD. Using a line laser as a pointing interface coupled with the virtual guide let the user find the position of the content the fastest. While over 60% of participants rated this interface as not obstructive, it did not help improve user performance in remembering content.

I have also explored position and movement type of the virtual guide while guiding. Results showed that although the teleportation interface reduced the time taken waiting for the agent to move, users did not prefer this because it lacked cues that allowed them to track and follow the guide. The flying type was rated positive due to its natural. Regarding position half of the participants prefer inside type and the other half prefer outside type. Based on analysis of user movement, I found that people who prefer the inside type tend to stand at a fixed position for most of the time, while the people who prefer outside type tend to walk around the exhibit area.

By comparing guidance with and without a virtual guide in several guiding interfaces, I found that most of the virtual guide presence conditions were rated positively for attractiveness, stimulation, and novelty of the system.

In a training scenario where task performance is more important than user experience as in the learning scenario, having a virtual guide may promote a user to continue performing a task, which would be useful in guiding a child or elderly individual to do something. However, in the case of the packing task that needed optimization of the packing time, the virtual guide is less important than audio and graphical visualizations that help the worker identify the item and packing solution. Since I have focused on improving a working performance, I didn't implement a virtual guide function in the study of using AR in the training scenario.

For the packing task, having an automated solution suggestion system helps decrease both time and hand movement for the task. I compared several solution indication interfaces, and the results suggested that the rotation indication helps the user pack faster, but can be hard to interpret. On the other hand, the movement indication needs the user to observe the box transition first and then mimic, which required more time than the rotation indication but was preferred by the most of the users.

7.2 Future Directions

This work has focused on presenting guidance possibilities and evaluating usage and performance of the systems. Installation of cameras in the environment to track the behaviour and progress of the user should be considered, and an individual user database system would provide more suitable user-oriented guidance.

By having a camera tracking the user's behaviour, various information can be observed such as exhibits of interest, boredom or loss of focus on explanations, or thinking or pondering behavior. Then a progress time and/or amount of information to be explain to match the user condition can be automatically adjusted. Also, having a database store information for the preference and experience of each user would help the system better suggest guidance information and style for the user the next time, or evaluate and match the type or preferences of a new user.

In a training scenario, tracking of user progress also has many merits. Suggestion of the next guidance should be automated or point out mistakes in real-time. Moreover, tracking the user's body condition, such as movement speed and facial expressions to evaluate working performance and task load can provide a better balance of work.

Bibliography

[1] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[2] Mark A Livingston and Andrei State. Magnetic tracker calibration for improved augmented reality registration. *Presence: Teleoperators & Virtual Environments*, 6(5):532–546, 1997.

[3] Joseph Newman, David Ingram, and Andy Hopper. Augmented reality in a wide area sentient environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 77–86. IEEE, 2001.

[4] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94. IEEE, 1999.

[5] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. Pose tracking from natural features on mobile phones. In *Proceedings of the 7th IEEE/ACM international symposium on mixed and augmented reality*, pages 125–134. IEEE Computer Society, 2008.

[6] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. Navigation system for the blind: Auditory display modes and guidance. *Presence*, 7(2):193–203, 1998.

[7] Olivier Bau and Ivan Poupyrev. Revel: tactile feedback technology for augmented reality. *ACM Transactions on Graphics (TOG)*, 31(4):89, 2012.

[8] Tomoya Yamada, Satoshi Yokoyama, Tomohiro Tanikawa, Koichi Hirota, and Michitaka Hirose. Wearable olfactory display: Using odor in outdoor environment. In *IEEE Virtual Reality Conference (VR 2006)*, pages 199–206. IEEE, 2006.

[9] Takuji Narumi, Shinya Nishizaka, Takashi Kajinami, Tomohiro Tanikawa, and Michitaka Hirose. Augmented reality flavors: gustatory display based on edible marker and cross-modal interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 93–102. ACM, 2011.

[10] Retrieved December 9, 2019 from `https://www.displaydaily.com/article/display-daily/ar-huds-and-martinis`.

[11] Retrieved December 9, 2019 from `https://www.theverge.com/2019/1/17/18186781/wacom-magic-leap-ar-augmented-reality-partnership-ces-2019`.

[12] Retrieved December 9, 2019 from `https://www.youtube.com/watch?v=vDNzTasuYEw`.

[13] Retrieved December 9, 2019 from `http://retail-innovation.com/topshop-in-moscow-had-a-virtual-fitting-room-on-trial`.

[14] Retrieved December 9, 2019 from `http://frag-games.com/2019/08/13/what-is-augmented-reality-ar/`.

[15] Paul Rookes and Jane Willson. *Perception: Theory, development and organisation*. Routledge, 2005.

[16] Arindam Dey, Mark Billinghurst, Robert W Lindeman, and J Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:37, 2018.

[17] Thomas P Caudell and David W Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the twenty-fifth Hawaii international conference on system sciences*, volume 2, pages 659–669. IEEE, 1992.

[18] Francesca De Crescenzio, Massimiliano Fantini, Franco Persiani, Luigi Di Stefano, Pietro Azzari, and Samuele Salti. Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, 31(1):96–101, 2010.

[19] Giles Westerfield, Antonija Mitrovic, and Mark Billinghurst. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, 25(1):157–172, 2015.

[20] Bui Minh Khuong, Kiyoshi Kiyokawa, Andrew Miller, Joseph J La Viola, Tomohiro Mashita, and Haruo Takemura. The effectiveness of an ar-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*, pages 57–62. IEEE, 2014.

[21] Jianlong Zhou, Ivan Lee, Bruce Thomas, Roland Menassa, Anthony Farrant, and Andrew Sansome. Applying spatial augmented reality to facilitate in-situ support for automotive spot welding inspection. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 195–200. ACM, 2011.

[22] Matt Adcock, Dulitha Ranatunga, Ross Smith, and Bruce H Thomas. Object-based touch manipulation for remote guidance of physical tasks. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pages 113–122. ACM, 2014.

[23] Yuki Uranishi, Goshiro Yamamoto, Zeeshan Asghar, Petri Pulli, Hirokazu Kato, and Osamu Oshiro. Work step indication with grid-pattern projection for demented senior people. In *Engineering in Medicine and Biology Society (EMBC),*

2013 35th Annual International Conference of the IEEE, pages 4698–4701. IEEE, 2013.

[24] Andreas Dünser and Eva Hornecker. Lessons from an ar book study. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 179–182. ACM, 2007.

[25] María Blanca Ibáñez, Ángela Di Serio, Diego Villarán, and Carlos Delgado Kloos. Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, 71:1–13, 2014.

[26] Brett E Shelton and Nicholas R Hedley. Using augmented reality for teaching earth-sun relationships to undergraduate geography students. In *The First IEEE International Workshop Augmented Reality Toolkit,*, pages 8–pp. IEEE, 2002.

[27] Tsutomu Miyashita, Peter Meier, Tomoya Tachikawa, Stephanie Orlic, Tobias Eble, Volker Scholz, Andreas Gapel, Oliver Gerl, Stanimir Arnaudov, and Sebastian Lieberknecht. An augmented reality museum guide. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 103–106. IEEE Computer Society, 2008.

[28] Marco Mason. The mit museum glassware prototype: Visitor experience exploration for designing smart glasses. *Journal on Computing and Cultural Heritage (JOCCH)*, 9(3):12, 2016.

[29] Natalia Vainstein, Tsvi Kuflik, and Joel Lanir. Towards using mobile, head-worn displays in cultural heritage: user requirements and a research agenda. In *Proceedings of the 21st international conference on intelligent user interfaces*, pages 327–331. ACM, 2016.

[30] Justine Cassell, Joseph Sullivan, Elizabeth Churchill, and Scott Prevost. *Embodied conversational agents*. MIT press, 2000.

[31] Mahoro Anabuki, Hiroyuki Kakuta, Hiroyuki Yamamoto, and Hideyuki Tamura. Welbo: An embodied conversational agent living in mixed reality space. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 10–11. ACM, 2000.

[32] Timothy Bickmore, Reza Asadi, Aida Ehyaei, Harriet Fell, Lori Henault, Stephen Intille, Lisa Quintiliani, Ameneh Shamekhi, Ha Trinh, Katherine Waite, et al. Context-awareness in a persistent hospital companion agent. In *International Conference on Intelligent Virtual Agents*, pages 332–342. Springer, 2015.

[33] Kota Gushima, Hina Akasaki, and Tatsuo Nakajima. Ambient bot: delivering daily casual information through eye contact with an intimate virtual creature. In *Proceedings of the 21st International Academic Mindtrek Conference*, pages 231–234. ACM, 2017.

[34] Tsukasa Noma, Liwei Zhao, and Norman I Badler. Design of a virtual human presenter. *IEEE Computer Graphics and Applications*, 20(4):79–85, 2000.

[35] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, et al. Ada and grace: Toward realistic and engaging virtual museum guides. In *International Conference on Intelligent Virtual Agents*, pages 286–300. Springer, 2010.

[36] Luca Vacchetti, Vincent Lepetit, George Papagiannakis, Michal Ponder, Pascal Fua, Daniel Thalmann, and Nadia Magnenat-Thalmann. Stable real-time interaction between virtual humans and real scenes. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 449–456. IEEE, 2003.

[37] Susanne Van Mulken, Elisabeth André, and Jochen Müller. The persona effect: How substantial is it? In *People and computers XIII*, pages 53–66. Springer, 1998.

[38] Robbert-Jan Beun, Eveliene De Vos, and Cilia Witteman. Embodied conversational agents: effects on memory performance and anthropomorphisation. In *International Workshop on Intelligent Virtual Agents*, pages 315–319. Springer, 2003.

[39] Timothy Bickmore, Laura Pfeifer, and Langxuan Yin. The role of gesture in document explanation by embodied conversational agents. *International Journal of Semantic Computing*, 2(01):47–70, 2008.

[40] Dai Hasegawa, Yücel Uğurlu, and Hiroshi Sakuta. A human-like embodied agent learning tour guide for e-learning systems. In *2014 IEEE Global Engineering Education Conference (EDUCON)*, pages 50–53. IEEE, 2014.

[41] Abraham G Campbell, John W Stafford, Thomas Holz, and Gregory MP O'Hare. Why, when and how to use augmented reality agents (auras). *Virtual Reality*, 18(2):139–159, 2014.

[42] Daniel Wagner, Mark Billinghurst, and Dieter Schmalstieg. How real should virtual characters be? In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, page 57. ACM, 2006.

[43] Retrieved December 9, 2019 from `https://www.louvre.fr/en/groups-schools/tours-official-guide`.

[44] Retrieved December 9, 2019 from `https://www.microsoft.com/en-us/hololens`.

[45] Jun Xiao, John Stasko, and Richard Catrambone. The role of choice and customization on users' interaction with embodied conversational agents: effects on perception and performance. pages 1293–1302, 04 2007.

[46] Retrieved December 9, 2019 from `http://www.unity-chan.com/`.

[47] Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannan, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. Proxemics [and comments and replies]. *Current anthropology*, 9(2/3):83–108, 1968.

[48] Satoshi Iwasaki, Mitsuyoshi Nagura, Shuji Ocho, and Tomoyuki Hoshino. Contribution of speech rate to speech perception in multichannel cochlear implant users. *Annals of Otology, Rhinology & Laryngology*, 111(8):718–721, 2002.

[49] Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, Takeshi Maeda, Takayuki Kanda, and Ryohei Nakatsu. Robovie: an interactive humanoid robot. *Industrial robot: An international journal*, 28(6):498–504, 2001.

[50] Wikipedia. Netherlandish proverbs, 2009. Retrieved August 27, 2019 from `https://en.wikipedia.org/wiki/Netherlandish\protect\discretionary{\char\hyphenchar\font}{}{}_Proverbs`.

[51] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50–9, pages 904–908. Sage Publications Sage CA: Los Angeles, CA, 2006.

[52] István Barakonyi and Dieter Schmalstieg. *Ar puppet: Animated agents in augmented reality*. na, 2004.

[53] Inc. AI. Aitalk®, 2009. Retrieved August 27, 2019 from `https://www.ai-j.jp/english`.

[54] Ellen R Girden. *ANOVA: Repeated measures*. Number 84. Sage, 1992.

[55] Bradley Efron. Student's t-test under symmetry conditions. *Journal of the American Statistical Association*, 64(328):1278–1302, 1969.

[56] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[57] I Elaine Allen and Christopher A Seaman. Likert scales and data analyses. *Quality progress*, 40(7):64–65, 2007.

[58] Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, 1(4):221–228, 1996.

[59] RF Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007.

[60] Maria Rauschenberger, Martin Schrepp, Manuel Pérez Cota, Siegfried Olschner, and Jörg Thomaschewski. Efficient measurement of the user experience of interactive products. how to use the user experience questionnaire (ueq). example: Spanish language version. *International Journal of Artificial Intelligence and Interactive Multimedia. 2013; 2 (1): 39-45*, 2013.

[61] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146. ACM, 2011.

[62] R Clifford Blair and James J Higgins. A comparison of the power of wilcoxon's rank-sum statistic to that of student's t statistic under various nonnormal distributions. *Journal of Educational Statistics*, 5(4):309–335, 1980.

[63] Julien Bramel and David Simchi-Levi. *The logic of logistics: theory, algorithms, and applications for logistics management.* Springer New York, 1997.

[64] AL Abramov and AA Kovtaniuk. Decision support systems in transport logistics. 2010.

[65] Teodor Gabriel Crainic, Guido Perboli, and Roberto Tadei. Extreme point-based heuristics for three-dimensional bin packing. *Informs Journal on computing*, 20(3):368–384, 2008.

[66] Erick Dube, Leon R Kanavathy, and Phoenix Woodview. Optimizing three-dimensional bin packing through simulation. 2006.

[67] John A George and David F Robinson. A heuristic for packing boxes into a container. *Computers & Operations Research*, 7(3):147–156, 1980.

[68] Daniel Hartmeier. A novel packing heuristic based on rigid body simulation, 2016. Retrieved September 25, 2019 from `http://www.benzedrine.ch/3D-ODRPP.html`.

[69] Bui Minh Khuong, Kiyoshi Kiyokawa, Andrew Miller, Joseph J LaViola Jr, Tomohiro Mashita, and Haruo Takemura. Context-related visualization modes of an ar-based context-aware assembly support system in object assembly. *Transactions of the Virtual Reality Society of Japan*, 19(2):195–205, 2014.

[70] Michael R Marner, Andrew Irlitti, and Bruce H Thomas. Improving procedural task performance with augmented reality annotations. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 39–48. IEEE, 2013.

[71] Oliver Bimber and Ramesh Raskar. *Spatial augmented reality: merging real and virtual worlds*. AK Peters/CRC Press, 2005.

[72] Erhan Baltacioglu, James T Moore, and Raymond R Hill Jr. The distributor's three-dimensional pallet-packing problem: a human intelligence-based heuristic approach. *International Journal of Operational Research*, 1(3):249–266, 2006.

[73] Narendra Karmarkar and Richard M Karp. An efficient approximation scheme for the one-dimensional bin-packing problem. In *Foundations of Computer Science, 1982. SFCS'08. 23rd Annual Symposium on*, pages 312–320. IEEE, 1982.

[74] Andrea Lodi, Silvano Martello, and Michele Monaci. Two-dimensional packing problems: A survey. *European journal of operational research*, 141(2):241–252, 2002.

[75] Xiaozhou Zhao, Julia A Bennell, Tolga Bektaş, and Kath Dowsland. A comparative review of 3d container loading algorithms. *International Transactions in Operational Research*, 23(1-2):287–320, 2016.

[76] Silvano Martello, David Pisinger, and Daniele Vigo. The three-dimensional bin packing problem. *Operations Research*, 48(2):256–267, 2000.

[77] Arash Sharif. Packit4me api, 2014. Retrieved September 25, 2019 from `http://www.packit4me.com/api`.

[78] Optimum Logistics LLC. Loadplanner - advanced load planning and optimization software solution, 2004. Retrieved September 25, 2019 from `http://www.loadplanner.com`.

[79] 3DBINPACKING SERVICE. 3d bin packing - be intelligent with packing and money, 2012. Retrieved September 25, 2019 from `http://www.3dbinpacking.com`.

[80] Jessica A Tsimeris. Visual cues for the instructed arrangement of physical objects using spatial augmented reality (sar). *University of South Australia*, 2010.

[81] Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (pcl). In *Robotics and automation (ICRA), 2011 IEEE International Conference on*, pages 1–4. IEEE, 2011.

[82] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[83] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.