Hybrid Object and Screen Stabilized Visualization Techniques for an AR Assembly Support System

Bui Minh Khuong*1, Kiyoshi Kiyokawa*2, Tomohiro Mashita*1*3, and Haruo Takemura*1*3

Abstract – This study proposes and evaluates the effectiveness of hybrid object and screen stabilized visualization techniques for an assembly support system using augmented reality (AR). Object stabilized visualization techniques are frequently used in AR-based assembly support systems but are not always available for the types of tasks, large sizes of assembled models, and narrow field-of-view head mounted displays (HMDs) examined in this paper. Based on a pilot study that investigated the best display locations and content sizes for displaying guidance information on an HMD screen, we propose two hybrid object and screen stabilized visualization modes, and then evaluate the two modes with an object stabilized visualization mode that we studied in the previous works (the side-by-side mode). Our experimental results indicate that the hybrid mode showing target assembly objects at a fixed position on the HMD screen with object stabilized orientation yields the best performance and subjective rating.

Keywords : augmented reality, assembly support system, visualization modes, reference frames.

1 Introduction

Augmented Reality (AR) is an innovative technology that blends computer generated virtual objects with the real environment [3]. One very promising AR application is in the traditional manufacturing assembly domain. In manufacturing, while some assembly operations are automated, there are still a significant number of assembly operations that require manual human effort. There are several reasons why AR may improve manual assembly. AR technology makes it possible to display digital information in the assembler's field of view, such as step-by-step instructions that are essential for the task. Such visualization helps to reduce search time in assembly. AR also makes it possible to reduce head and eye movements [7] and attention switching and to improve spatial perception, and thus it can increase productivity [9].

In assembly support systems using AR, visualization techniques have a very important role because they directly affect the efficiency of assembly tasks and a user’s mental workload. In previous work [8], we proposed visualization techniques for AR-based assembly support systems and evaluated their effectiveness in assembly tasks. We found that an object stabilized visualization mode that shows virtual target status next to the real objects of concern (the side-by-side mode, see Figure 1) outperforms traditional direct overlay under moderate registration accuracy with marker-based head tracking and RGBD camera-based object tracking. However, the side-by-side mode is not always available. In the context of large assembly models assembled with narrow field of view head mounted displays (HMDs), like those commonly used in assembly support systems, virtual guidance information is often out of the viewport of the HMD screen and assembly subjects need to move their head out of the workspace to find the guidance. This increases head and eye movement, cost of attention switching, and difficulties for spatial perception and thus it likely impacts the completion time of assembly tasks. Large assembly models in this context are models that do not fit within the field of view of the HMD at a normal reaching distance in assembly.

In this paper, we take advantage of the head (screen) stabilized information display and propose hybrid object and screen stabilized visualization techniques as a solution for the problem above. In screen stabilized visualization techniques, information stays in the user’s view regardless of his or her head motion [6]. Thus, it may decrease head and eye movement and cost of attention switching, and it may be an appropriate presentation technique for information that is crucial for the task or needs to be noticed very quickly. Our experimental results indi-
cate that one of the two hybrid visualization modes yields better performance and user preferences than the side-by-side mode, the best performing mode in our prior series of experiments [8]. The results also provide useful insight into the design of AR visualization techniques for assembly support systems.

2 Related Work

2.1 AR Assembly Support Systems

One of the most well-known applications of AR in the assembly domain is the assembly support system for cable harnesses at Boeing [3]. Their augmented reality project was designed to display pertinent instructions and diagrams in front of manufacturing workers, who used the information to work on or assemble pieces of the aircraft. Reiners et al. [14] developed an AR prototype system for assembling door locks on cars which used an optical see-through (OST) interface for presenting instructions. Zauner et al. [17] developed a prototype system for assembling of furniture with AR in which step-by-step instructions were given to users, who successfully completed the assembly task. Salonen and Sääskä [15] designed an AR system for assembling a 3D puzzle. They also used this for simulating industrial problems like putting together a power unit accessory of a tractor using CAD models as graphical objects. BMW has also experimented with AR to improve welding processes on their cars [5]. Pentenrieder et al. [11] show how Volkswagen uses AR in the field of construction to analyze interfering edges, plan production lines and workshops, compare variance and verify parts. Radkowski and Stritzke [13] studied the use of AR for assembly by using virtual parts instead of real ones, and they designed a solution based on a Microsoft Kinect video camera. In our previous work [8] we introduced a test-bed system that also uses a Microsoft Kinect for real time object tracking as well as a video see-through HMD to display guidance information and error detection information corresponding to the current detected assembly status. This Kinect system is also used as the test-bed system for the study in this paper.

2.2 Visualization Techniques for Assembly Support Systems

There has also been a rich body of work on visualization techniques for assembly support systems using AR. For example, Caudell and Mizell [2] use a small head-mounted video camera to detect visual markers on the workpiece, a computer estimates the relative pose information, and the diagram and text are subsequently displayed on the workpiece. Reiners [14] and his colleagues demonstrated a prototype AR system that uses passive retroreflective markers illuminated by IR sources to augment a mechanic’s natural view with text, labels, arrows, and animated sequences designed to facilitate task comprehension, location, and execution. Translational and rotational animations of visual graphics in the object’s coordinate system allow for those graphics to convey additional meaning in intuitive ways, such as the insertion of one part into another, twisting of a tool, or other manual interactions.

Visual information in AR is typically defined and displayed within in a particular reference frame. For example, Billinghurst et al. introduced three types of reference frame stabilizations for information displays; head (screen), body and world stabilized [6]. Piekarsky et al. added location-relative coordinates and discussed the characteristics of four reference frames [12]. When guidance information is overlaid onto an object that is being tracked, it is no longer world stabilized, but object stabilized. When a target object is affixed to the real world, world and object stabilized information are essentially the same.

In this sense, most of visualization techniques introduced in the assembly support systems mentioned above are object stabilized. In object stabilized visualization techniques, information is affixed to the assembly objects themselves and its apparent position on screen changes as the user moves his or her head (see Figure 2). This requires the user’s viewpoint position and orientation to be tracked. Object stabilized information presentation also
enables annotation of the real world with context dependent visual and audio data, creating information enriched environments. This can increase the intuitiveness of the real world tasks [6].

Despite the advantages of object stabilized visualization techniques, some assembly support systems only use head stabilized (screen stabilized) information display (see Figure 3). Baird and Barfield [4] presented a system with screen fixed instructions on untracked monocular OST and opaque HMDs to support a computer motherboard assembly task. In the screen stabilized visualization techniques, information is fixed to the user’s screen and it does not change as the user moves his or her viewpoint. Therefore, guidance information is always available to users and he or she can refer to it very quickly. However, the poses of virtual 3D guidance information are not normally updated in real-time to match to those of their real object counterparts. Thus the user needs to mentally rotate guidance information to the corresponding real object.

3 Hybrid Visualization Modes

3.1 Design Concept

Considering the pros and cons of each visualization technique, in this study we propose hybrid visualization techniques by combining both object and screen stabilized visualizations to take advantage of the two (see Figure 4).

As discussed in the previous section, object stabilized visualization can be advantageous for intuitive and coherent information display by referring to real objects. However, the guidance information is not available when the user looks away from the objects to which it refers. On the other hand, screen stabilized visualization is advantageous for immediate access to the guidance information regardless of the user’s head orientation at the expense of the necessity of its mental rotation to match to the corresponding object.

To incorporate the advantages of both approaches, our idea is to fix the guidance information in the screen coordinates, and to update its orientation in real-time to match to that of the referring objects. In other words, orientation of the guidance information is object stabilized, but its position is screen stabilized. This way, we hypothesized that the guidance information would be always immediately accessible and understandable without the need for additional mental rotation.

3.2 Proposed Techniques

With the above design kept in mind, we propose two hybrid object and screen stabilized visualization modes (see Figure 5). In the first mode (called Hybrid fixed mode, Figure 5.A), the virtual guidance information is displayed at a corner of the screen (indicated in red). Its center of gravity (COG) is always aligned with the center of the information area, while its 3D position in world coordinates is determined by the intersection of the ray from
the camera to the COG and the desktop surface (see Figure 6). Thus the virtual guidance moves along the desktop surface depending on viewing direction. Its orientation is also updated in real-time to match to that of the corresponding real target object. Its size in space is shrunk to fit into the maximum screen area available. If there is no intersection between the ray and the desktop surface, its position and orientation remain the same.

In the second mode (called Hybrid dynamic mode, Figure 5B), the system first detects the largest free rectangle area on the screen for every frame. It then automatically moves the virtual guidance information to that area with a smooth transition animation to minimize occlusion between the guidance information and the real objects. The 3D position of the virtual guidance is determined based on the intersection of the ray from the camera to the COG and the desktop surface.

3.3 Pilot Study to Determine Best Screen Position

To determine the best position and size of the guidance information area, we conducted a pilot study using the test bed system described in Section 4.1. We found that displaying guidance information on the top left area with a size of 50% by 50% of the entire screen yielded the best user preference among 12 participants, which would show the virtual guidance in approximately 60% in size of its real counterpart in our experiment. Four positions on the HMD screen (top left (TL), top right (TR), bottom left (BL), and bottom right (BR)) and three levels of size of guidance information (40%, 50%, 60% of the real counterpart) were compared (see Figure 7 and Figure 8). Other size options of guidance information (10%, 20%, 30% or 70%, 80%, 90%, etc.) were too small or too large when displaying on a narrow field-of-view HMD screen so we did not consider them in the scope of this pilot study. At each level of position, we respectively showed each level of size of the virtual guidance information and participants were asked to manipulate, translate, and rotate the model in the workspace. Then they were asked to fill out a questionnaire for feedback about the ease with which they could see the guidance information. The questionnaire consisted of a 7-point ordinal scale response, with 1 and 7 indicating the most negative and the most positive responses, respectively.

4 Main Evaluation

In the main evaluation, we evaluate the two hybrid object and screen stabilized visualization modes, Hybrid fixed mode and Hybrid dynamic mode, in comparison to

Side-by-side mode [8].

4.1 The Test-Bed System

The hardware setup of the test-bed system is shown in Figure 9 and Figure 10. The task was to assemble a building block structure from the table up, layer by layer, while the system highlights the next layer to build on the virtual representation of the real model being assembled. The system uses depth information captured in real-time by a depth sensor, the Microsoft Kinect, for acquiring and tracking. A video see-through head mounted display (Vuzix Wrap 920AR) (see Table 1) is used to display assembly guidance information to the user. All of our experiments were conducted on a desktop computer with an Intel core 2 Duo E7400 2.80GHz x 2 processors, an NVIDIA GeForce GT230 GPU and three gigabytes of memory.

4.2 Hypotheses

In the previous work [8], we found that the side-by-side mode had the best performance for task completion.
Khuong · Kiyokawa · Mashita · Takemura: Hybrid Object and Screen Stabilized Visualization Techniques for an AR Assembly Support System

Compared position and size of the guidance information area in the pilot study.

Conditions rated on ease with which participants could see them in the pilot study.

In Hybrid fixed mode, users can see both the virtual guidance information and the real object in the HMD’s viewing field at the same time, even if the real object is large. Automatic scaling of the virtual guidance helps users see the entire model and allows them to confirm assembly details on demand.

In Hybrid dynamic mode, although the system automatically moves guidance information to avoid occlusion of the object being assembled, the user may lose focus and feel stress when he or she frequently has to search for guidance information. This may also increase completion time for assembly tasks.

Therefore, we formulated the following hypotheses:

- **H1**: Hybrid fixed mode will achieve the best task completion time among the visualization modes.
- **H2**: Hybrid fixed mode will achieve the highest subjective ratings for: ease of understanding, ease with which users can see information, stress level, familiarity, satisfaction and usefulness.

### 4.3 Experiment Design and Procedure

We recruited 24 people (12 males, 12 females) from several departments of a college campus for participation.

- **H1**: Hybrid fixed mode will achieve the best task completion time among the visualization modes.
- **H2**: Hybrid fixed mode will achieve the highest subjective ratings for: ease of understanding, ease with which users can see information, stress level, familiarity, satisfaction and usefulness.

#### 4.3.1 Experiment Setup

We used the Vuzix Wrap 920AR for the assembly tasks. The specifications of the HMD are as follows:

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display Resolution</td>
<td>Twin high-resolution 640 × 480 LCD displays</td>
</tr>
<tr>
<td>Screen Size</td>
<td>Equivalent to a 67-inch screen as viewed at ten feet (approximately 3 m)</td>
</tr>
<tr>
<td>Refresh Rate</td>
<td>60 Hz progressive scan update rate</td>
</tr>
<tr>
<td>Field of View</td>
<td>31-degree diagonal field of view</td>
</tr>
<tr>
<td>Color Depth</td>
<td>24-bit true color (16 million colors)</td>
</tr>
<tr>
<td>Cameras</td>
<td>Two discrete VGA (640 × 480) video cameras</td>
</tr>
<tr>
<td>Camera Frame Rate</td>
<td>30 frames per second video capture at 640 × 480</td>
</tr>
</tbody>
</table>

The prototype system for guided assembly of building block (LEGO) structures.
in the evaluation. Participant age ranged from 22 to 30 years old, and only 5 participants had prior experience with AR.

We used a one way within-subjects experimental design, where the independent variable was the visualization mode for presenting assembly instructions, and dependent variables were time taken to complete the tasks, subjective scores on ease of understanding, ease with which they could see information, stress level, familiarity, satisfaction, and usefulness. The independent variable included the following modes: Hybrid fixed, Hybrid dynamic and Side-by-side. Each participant was subjected to all three modes in a randomized order. For each level, the participants were asked to assemble four Duplo models in a randomized order (see Figure 11).

Before the experiment, each participant was given a tutorial on each of the levels. The completion time in each assembly task was recorded by an experimenter. When a participant finished an assembly task, he or she was asked to fill out the questionnaire shown in Table 2. Each question was rated with a 7-point ordinal scale, with 1 and 7 indicating the most negative and positive responses, respectively.

4.4 Analysis on Quantitative Data

Figure 12 illustrates the mean task completion time for each mode in the evaluation. A star indicates a significance level of $p < 0.05$. On average, Hybrid fixed mode had the shortest completion time, while Hybrid dynamic mode had the longest. We conducted a repeated measures ANOVA test and post-hoc pairwise t-tests and found differences between Hybrid fixed mode and Hybrid dynamic mode ($t_{23} = -2.80, p = 0.010$). We found a near significant difference between Hybrid fixed mode and Side-by-side mode ($t_{23} = -1.70, p = 0.10$) but no significant difference between Side-by-side and Hybrid dynamic mode ($t_{23} = -0.13, p = 0.90$).

We also conducted a repeated measure ANOVA test and post-hoc pairwise t-tests on the task completion time data of only the large models (models 3 and 4 in Figure 11), and found significant differences between Hybrid fixed mode and Hybrid dynamic mode ($t_{23} = -2.67, p = 0.014$). We found a marginally significant difference between Hybrid fixed mode and Side-by-side mode ($t_{23} = -1.8794, p = 0.07$) but no significant difference between Side-by-side mode and Hybrid dynamic mode ($t_{23} = 0.34, p = 0.734$).

4.5 Analysis of Questionnaire Data

Figure 13 shows the participants’ ratings for the questionnaire for each condition in each aspect: ease of understanding, ease with which they could see information, stress level, familiarity, satisfaction level and usefulness. Based on the ratings elicited from the participants, the Hybrid fixed mode was the most preferred among three visualization modes in every qualitative aspect (see Table 3). We used a non-parametric Friedman test to check for significant differences in each subjective measure among the conditions, followed by Wilcoxon signed rank tests. One, two and three stars indicate significant differences of $p < 0.05, p < 0.01$, and $p < 0.001$, respectively.

4.6 Discussion

In the first hypothesis $H1$, we hypothesized that Hybrid fixed mode would achieve the best task completion time among the visualization modes considered. The results of the experiment supported this hypothesis.

In Hybrid fixed mode, users can see both the virtual guidance information whose pose is updated in real-time, and real counterpart objects being assembled in the same viewpoint on the HMD screen, even when the real objects

![Figure 12: The mean task completion time (sec) of each mode in the main evaluation. Error bars indicate 95% confidence intervals.](image-url)
being assembled were large. The guidance information also automatically changes the size on demand when the user moves his or her head closer to or further away from the work table. This helps the user reduce head and eye movements, at the cost of attention switching and spatial perception.

However, this mode still has occlusion problems between virtual guidance information and the real counterpart objects being assembled. The visualization of guidance information becomes harder when models become larger and more complex due to the limited screen space as well as narrow field of view and low resolution of the HMD screen. This might increase time to confirm parts’ positions as well as their relationship at each assembly step, and thus overall completion time may also increase. The Hybrid fixed time was not clearly faster than the Side-by-side (see Figure 14), presumably due to this issue.

In Hybrid dynamic mode, the users reported that although the system automatically detects and moves the guidance information to the position with the least occlusion between the guidance information and the real counterpart objects, controlling movement of the guidance information to a desirable position was difficult. They had to chase the movement of the guidance information continuously, causing them to lose focus. They had to spend more time to find and confirm guidance information during the assembly process, and thus this mode had the worst task completion time among all visualization modes.

In the second hypothesis H2, we stated that Hybrid fixed mode would achieve the best user preference in the following aspects: ease of understanding, ease with which users could see information, stress level, familiarity, satisfaction level and usefulness. The results of the experiment supported this hypothesis.

In Hybrid fixed mode we believed that the ability to display virtual guidance information at a suitable size and rotation at a fixed position at screen corners would help users see and understand the relationship between the guidance and its real counterpart. This subsequently reduced perceived workload during assembly tasks. This mode also resulted in a higher feeling of satisfaction and usefulness in comparison to the other visualization modes. This subjective data shows that this mode was preferred over the other visualizations, and these differences were all statistically significant for all aspects.

5 Conclusion

In this study, we proposed and evaluated the effectiveness of hybrid object and screen stabilized visualization modes for AR-based assembly support systems. Our experimental results indicate that one of our proposed visualization modes, Hybrid fixed mode, is preferred for many aspects, such as ease of understanding, stress level,
and usefulness, among all visualization modes tested. We also found that the Hybrid fixed mode yielded the fastest average task completion time with a marginally significant difference. These results provide useful insights into the design of AR visualization techniques for assembly support systems. The proposed hybrid visualization techniques assume a horizontal work surface, so we would like to extend these techniques to function in different work layouts, such as assembly tasks on a vertical wall. Although visual guidance based on geometric models has been solely investigated in this study, other forms of visual assistance [16][10] are also an interesting area to explore.

Acknowledgments


Minh-Khuong Bui

He is a PhD student at Graduate School of Information and Science, Osaka University. He received his ME degree in the same faculty in 2013. His interests include augmented reality, computer vision, and 3D user interfaces.

Kiyoshi Kiyokawa

He received his ME and PhD degrees in information systems from Nara Institute of Science and Technology (NAIST) in 1996 and 1998, respectively. He is currently a Professor at Graduate School of Information Science, NAIST. He worked for Communications Research Laboratory from 1999 to 2002. He was a visiting researcher at Human Interface Technology Laboratory at the University of Washington from 2001 to 2002. He was an Associate Professor at Cybermedia Center, Osaka University from 2002 to 2017. His research interests include virtual reality, augmented reality, human augmentation, 3D user interface, and CSCW. He is a member of the IEICE, the IPSJ, the VRSJ, the HIS, and the ACM.
Tomohiro Mashita  （正会員）

Tomohiro Mashita graduated from Osaka University in 2001 and completed the M.S. and doctoral programs in 2003 and 2006, respectively. He was a postdoctoral fellow at Osaka University from 2006 to 2008. He was a senior research fellow at Graz University of Technology from 2012 to 2013. He is currently an Assistant Professor at Cybermedia Center, Osaka University. He is research interest includes Computer Vision, Pattern Recognition, and Human Interface. He is a member of the IEICE, IPSJ, HIS, and IEEE.

Haruo Takemura  （正会員）

He received his BE, ME, and PhD degrees from Osaka University in 1982, 1984, and 1987, respectively. In 1987, he joined Advanced Telecommunication Research Institute, International. In 1994, he joined Nara Institute of Science and Technology, as an associate professor in the Graduate School of Information Science and Technology. From 1998 to 1999, he was a visiting associate professor at the University of Toronto, Ontario, Canada. He is a Professor at Cybermedia Center, Osaka University since 2001. His research interests include interactive computer graphics, human-computer interaction, and mixed reality. He is a member of the IEICE, the IPSJ, the VRSJ, the HIS, the IEEE, and the ACM.