

Viewpoint Usability for Desktop Augmented Reality



Seokhee Jeon, Hyeongseop Shim and Gerard J. Kim

Abstract—In this paper, we have investigated the comparative usability among three different viewing configurations of augmented reality (AR) system that uses a desktop monitor instead of a head mounted display. In many cases, due to operational or cost reasons, the use of head mounted displays may not be viable. Such a configuration is bound to cause usability problems because of the mismatch in the user's proprioception, scale, hand eye coordination, and the reduced 3D depth perception. We asked a pool of subjects to carry out an object manipulation task in three different desktop AR set ups. We measured the subject's task performance and surveyed for the perceived usability and preference. Our results indicated that placing a fixed camera in the back of the user was the best option for convenience and attaching a camera on the user's head for task performance. The results should provide a valuable guide for designing desktop augmented reality systems without head mounted displays.

Index Terms—Augmented Reality, Desktop, Interface, Usability, View Point.

I. INTRODUCTION: MOTIVATION

Augmented reality (AR) system combines the images from a real environment with properly registered virtual objects. As such, AR systems usually require see through head mounted displays (HMD). One popular form is using the video see through head mounted display in which the real images are obtained through a video camera mounted on the HMD positioned to be near, and aligned with the direction of the user's eyes on the HMD. This configuration can produce a naturally combined imagery in which the physical view of the user (nearly) matches the view into the actual mixed reality space.

However, in many cases, due to operational reasons (e.g. cost, maintenance problems), the use of head mounted displays may not be viable. In addition, if the interaction of the task is confined to a narrow area, thus requiring only limited head turning, one can resort to using a desktop monitor instead. In such a case, the resulting imagery will be viewed at a direction different from the view direction into the actual mixed reality

space where the interaction is occurring. Such configurations are expected to suffer from usability problems and deteriorate task performances, because of the mismatch in the user's proprioception, scale, and hand eye coordination, and of the reduced 3D depth perception.

Another consideration related to the trade-off between convenience and usability of an AR set up is the placement position of the camera for the video see through functionality. We have come up with three options for the camera placement.

The first option is to mount the camera on the user's head (e.g. on a cap worn by the user without an HMD). This approach, while expected to cause less usability problem than wearing a heavier and delicate-to-handle HMD, still could be a nuisance to the user. The second strategy is to use a fixed camera as opposed to the moving camera on the user's head. This set up is expected to be effective especially if an interaction task is confined to a small area. A natural choice is to place a fixed camera as close to the user's eyes as possible, e.g. in the back of the user at approximately the height of the user's eyes. This configuration surely improves the convenience of using an AR system by removing the need for wearing any devices. However it has a drawback in terms of system set up because a separate fixture is needed and it must not get in the way of user's physical interaction. The third option, which is conceived to be the most convenient place to put the camera, is at the top of the desktop display facing the user (some computers even have cameras integrated in such a location). However, in that case, the imagery would be reversed as the view direction of the camera is opposite from the user's viewing direction. The imagery can be mirrored both horizontally and vertically to provide the right directionality but the resulting image's looks as if it was taken from an incomprehensibly odd angle. One way to alleviate this problem is to warp the image to correct for the distorted perspective as much as possible.

In this paper, we have investigated the comparative usability among three different configurations of augmented reality system that uses a desktop monitor instead of a head mounted display. We asked a pool of subjects to carry out an object manipulation task in an AR environment using a desktop display that provided views of the mixed reality space from the three different camera positions (See Fig. 1): (1) user's head (moving), (2) back of the user (fixed), (3) front of the user facing toward the user (fixed, but with the warped view to provide a more correct understanding of the working environment). We measured the subject's task performance and surveyed for perceived usability and preference. We hope

Manuscript Received on September 28, 2006.

Seokhee Jeon, Dept. of Computer Science and Engineering, Pohang, Korea, yeager@postech.ac.kr

Hyeongseop Shim, Dept. of Computer Science and Engineering, Pohang, Korea, yeager@postech.ac.kr

Gerard J. Kim, Dept. of Computer Science and Engineering, Korea University, gjkim@korea.ac.kr (Corresponding author)

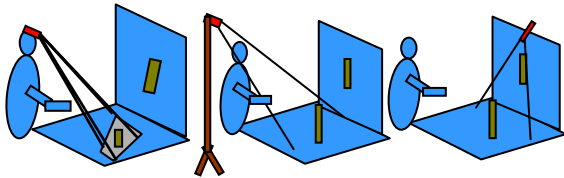


Fig. 1. The three viewing/display conditions: (1) with head mounted camera (left), (2) with a fixed camera in the (middle), (3) back with a fixed camera in front and warped and mirrored image (right).

that the results will provide a valuable guideline for designing desktop augmented reality systems without head mounted displays that provide the right balance between usability and system set up convenience.

II. RELATED WORK

Few AR applications use desktop displays as its major display by design. It is suspected that using desktop displays (upright) is implicitly regarded non-ideal compared to HMD's most probably due to the viewpoint mismatch. Field et al. used an upright rear projection screen and a camera looking toward the user and the interaction space in their Augmented Chemistry application. However, they did not specify in details why such a display was used other than it was meant to give the feeling of a mirror image of the interaction space [1]. The term "Desktop AR" is in fact more commonly used to indicate that the desktop is used as the interaction space. Tian et al. presented a desktop AR system using a small LCD screen (held over and the camera looking toward the interaction space) [2], and Knowlton used a half mirror overlaid upon the desktop interaction space [3]. Bimber et al. developed a projection-based augmented reality system in which a real object and a virtual object could be seen simultaneously without using HMD on the desktop [4].

Usability issues with regards to using upright displays for AR have not attracted much attention either. Kraut et al. compared different display configurations for video based remote collaboration. In their work, the authors compared the level of collaboration with displays, for instance, coming from the collaborator's head mounted camera (giving a small view of the interaction space), a fixed camera located a step back of the collaborator (to give a wider view), or both. Not much difference in performances was found among the test conditions in this study. Furthermore, this study focused on the degrees in the ease of collaboration than in the task performance itself [5]. Biocca et al. has showed the importance of hand-eye coordination in using a video based see-through HMD in AR systems [6]. Their study investigated at the effects caused by a HMD design that displaced the user's "virtual" eye position (by the position of the camera). The position of the cameras created images of the world that were slightly downward and inward from the normal. The study revealed significant reduction in task performance in terms of time and accuracy, and despite some evidence of adaptation by the users, the performance did not recover back to the baseline level.

They also found that when subjects removed the see-through HMD, their hand-eye coordination had been altered.

III. EXPERIMENT

The purpose of this study is to assess the comparative usability of the desktop AR systems that does not employ head mounted displays. As such the experiment does not include a test environment in which a HMD is used. The basic approach is to have the subjects carry out a particular AR task in the three different viewing/display configurations as defined in Section 1 and measure their task performance and subjective feelings and judgments about the respective environments.

A. Experimental Task

The desktop AR usability was measured in two ways: by having the subjects perform a simple manipulation task and measuring their task completion time, and by surveying the subjective usability using a short questionnaire. The experimental task was to move a virtual teapot (whose position was set randomly), using a hand held marker, to a randomly generated target position and match its orientation at the same time (see Fig. 2). We used the ARToolkit by Kato et al. [7] to track the position of the hand-held marker. To remove an effect of the difference in the fields of view among three display configurations, the randomly generated target virtual teapot was always positioned within the camera's field of view. The amount of required rotation was also controlled because it was not possible to make full rotation with the 2D markers (the marker would not be visible to the camera at severe angles). A single matching task was deemed successful when the position and orientation was within pre-defined threshold accuracy. The task was carried out in two difficulty levels: one at 70% accuracy (easy) and the other at 90% accuracy (difficult).

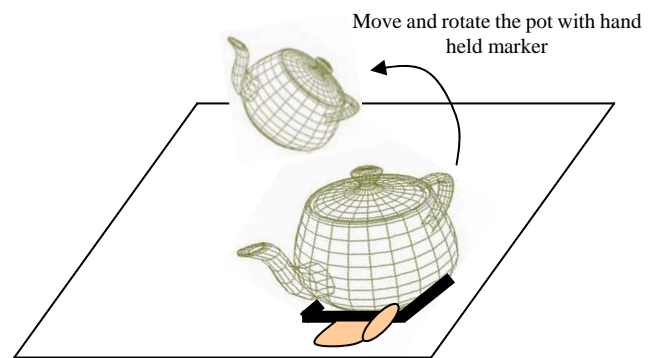


Fig. 2. The experimental task was to move and match a virtual teapot (using a hand held marker) to a randomly generated target position and pose.

Pre-runs of the experiment indicated that the size of the virtual teapot (manipulation object) was an important factor in effective task performance. Thus, the sizes (as displayed on the monitor) of the objects (between the three test display configurations) were adjusted according to the viewing distances so that they were perceived to be approximately the

same size. Thus, as for the second and third viewing conditions, it was not necessary to position the cameras exactly at the height of the user. The difference in the resulting image resolution was negligible.

The task was chosen particularly to observe the subject's sense of directionality in the different conditions. Table 1 shows the short usability questionnaire used in the experiment. The first questionnaire was given to the subject after each treatment. The second questionnaire was answered after all treatments, asking the overall subjective feeling about the viewing condition. The questions were answered in the 7-Likert scale.

TABLE 1:USABILITY QUESTIONNAIRE.

Questions for each treatment
1. How easy/convenient was it to carry out the task in this environment? (Answer in the scale of 1 to 7, 1 being the most inconvenient)
2. If you felt the task was difficult or inconvenient, please choose reasons why it was so from the following or write out the reasons. <ol style="list-style-type: none"> a. I could not adapt myself to the environment in time. b. The display did not match my sense of the view. c. I could not get the sense of depth from the display. d. The display did not provide sufficient resolution. e. The virtual object did not look realistic. f. The virtual object was too jittery. g. It was just too hard and not natural to manipulate the object using the hand held marker. h. The camera mounted on the head was a nuisance (only asked when the head mounted camera was used). i. Others (write down separately).
3. How much fatigue (physical or mental) did you experience in this environment? (Answer in the scale of 1 to 7, 1 being the most tiring)
4. Please rate the overall test environment (Answer in the scale of 1 to 7, 1 being the worst).
Questions after all treatments
1. Please give your order of preference (in the increasing order) among the test environments you have just tried out.
2. Please write down any of your opinions regarding the experiment.

B. Experimental Set Up

The task described above was carried out under three different viewing/display conditions. Fig.3. shows the three different viewing and display set ups. All of the set ups do not use a head mounted display as the purpose of the study is to assess the usability of the different non-HMD AR configurations for interactive tasks.

The first viewing/display configuration is to mount the camera on the user's moving head. The second is to place a camera at fixed position in the back of the user, such that the camera is as close to and aligned with as the user's eye. The third configuration is to place a camera at a fixed position in the front, facing toward the user.

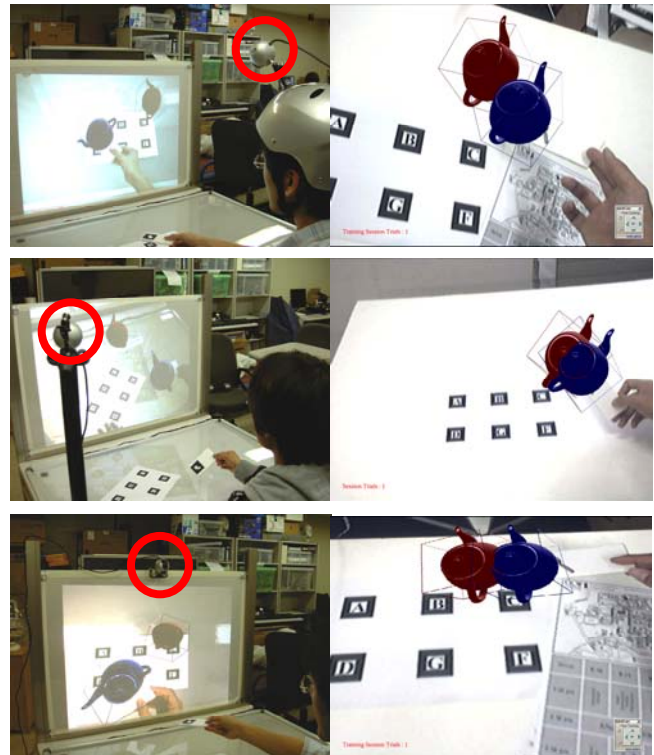


Fig. 3. Three tested augmented reality views (all displays are positioned upright in front of the user). The user is to move the blue teapot and match it with the red one. Top: With the camera mounted on the user's head (near eye view), middle: with the camera at a fixed position in the back of the user, and bottom: with the camera on the opposite side but with the image mirrored horizontally and vertically and warped. See Color Plate 13.

In all cases, the subjects were instructed to perform the task looking at a display situated in front of the user in an upright position (see Fig.1. again). Note that we excluded the case where the camera was placed on top of the interaction space providing a top down view. This configuration is deemed similar to placing the camera in the back. Table 2 summarizes and compares the characteristics of the three viewing/display conditions. In all cases, the view into the display does not match the view into the interaction space. As for the match between the head-neck proprioceptive cue and the view, the first configuration is the most natural, then the second, then the last. Thus, it was expected that the third display configuration would clearly be the worst (among the three) in all aspects except for its convenience in set up (e.g. just placing the camera on top of the display, some laptops already come with cameras mounted in that position).

In hopes of making this configuration a bit more competitive to the other two, we applied image warping to improve the proprioceptive match and reduce the user's difficulty in handling the directional tasks (see Section III(B)). We later also conducted a separate supplemental experiment to further isolate the two factors (simple mirroring and adding image warping). The results from this experiment showed that adding image warping did not really improve the situation contrary to our wish, nor did it worsen it to invalidate our main experiment (see Section III(B)).

TABLE 2: CHARACTERISTICS OF THE THREE TESTED VIEWING/DISPLAY CONDITIONS

Camera Position	View	Setup	Proprioceptive match	Image
Head Mounted	Moving	Major	High	As it is
Back	Fixed	Medium	High	As it is
Front	Fixed	Minor	Low	Warped and mirrored

1) Image Warping (for experimental condition 1)

First we warped the scene of the interaction space (assuming it is a flat table) using the homographic relationship as shown in Fig.4. We assume that the location of the plane M (the interaction table surface) and the virtual camera (from which we would like to simulate the view) are known. Then, we can compute the vertices of a quadrilateral, R , formed by the intersection of the view frustum of the real camera and the plane M . Similarly, we can compute the quadrilateral, V , the intersection of the view frustum of the virtual camera and M . We can establish a homographic relationship between R and V with the four corresponding points and map the original image R onto V . By doing so creates an image as if it was viewed from the virtual camera (located in the back of the user as in the second of viewing configuration).

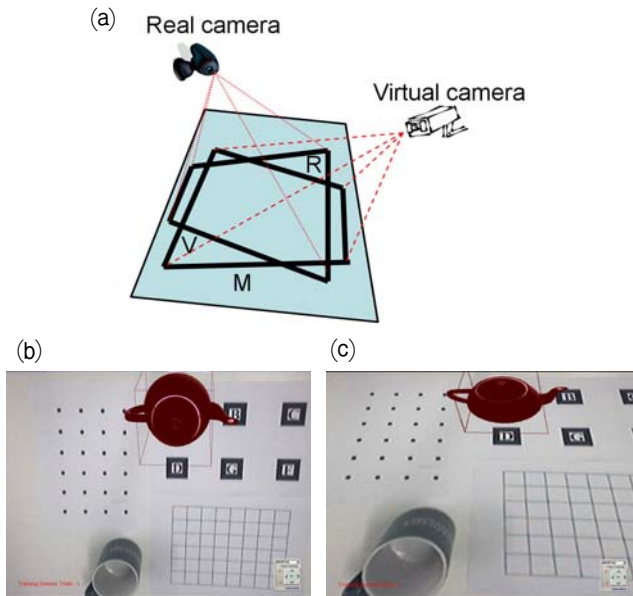


Fig. 4. Creating an image for virtual camera (a). Image from the front camera mirrored horizontally and vertically (b), image of the top right warped to provide a view from a different perspective (c). Because of the assumption of a planar interaction space (background), the 3D cup is displayed inconsistent to the virtual camera view.

The interaction object (co-located with the marker) is handled and drawn separately from the background image. The location and pose relative to the real camera is obtained from the AR system (we use the ARToolkit by Kato et al. [7]). And by the known relationship (i.e. both positions/orientations are known) between the real camera and the virtual camera, the object can be easily drawn from the perspective of the virtual camera.

$$T_{virtualCamera}^{marker} = T_{realCamera}^{marker} \times T_{virtualCamera}^{realCamera} \quad (1)$$

Fig.4(b) shows the resulting image in contrast to the image that is merely mirrored. Because formulation assumes that interaction space is planar, the planar table surface, planar objects on the table (e.g. the grid and the marker pattern) and interaction object look approximately right, but the 3D object on the table (e.g. the cup or the hand) is somewhat distorted.

A. Experimental Design and Procedure

A two factor within-subject experimental design was used: the first factor (independent variable) was the type of view/display configuration with three levels and the second factor was the difficulty of the task (See Section III (C)). The dependent variables were the task completion time and other assorted answers to the usability questionnaire. Subjects experienced the six test conditions shown in Table 3.

TABLE 3: SIX TEST CONDITIONS IN THE EXPERIMENT.

View/Display \ Difficulty	Head mount	Back	Front
Easy	H1	B1	F1
Difficult	H2	B2	F2

The subject was first briefed for the main purpose of the experiment and his vital information was collected such as age, gender, background, power of vision, color blindness, and experiences with 3D games or AR/VR systems. As the given task was a 3D manipulation task (including rotation), the subjects were assessed for their spatial abilities using the standard paper and pencil mental rotation test [8] (subjects were given 24 questions and asked to stop after 3 minutes). The subjects experienced the six test conditions in an order generated by Diagram-Balanced Latin Square method. The subject performed four practice trials (for each test condition), then sixteen main trials in a row without resting. After filling out a questionnaire and brief rest, the subject went on to the next combination. After finishing all the six test conditions, they answered the final two items in the questionnaire. The overall experiment (for one subject) lasted about an hour. Fig.3 shows snapshots from the experiment. Twenty four subjects participated in the study. There were 18 males and 6 females. All the subjects were engineering students with an average age of around 22.

IV. RESULTS

Fig.5 shows the results of the ANOVA on task performance among the three test conditions. According to the Student-Newman-Keuls grouping, for both types of the task (difficult and easy), the viewing conditions of head mounted camera and back camera (H1, H2, B1, B2) were grouped together separately to the front camera (F1, F2). In summary, the task completion time was different with statistical

significance between the head-mount/back (shorter) and front camera (longer).

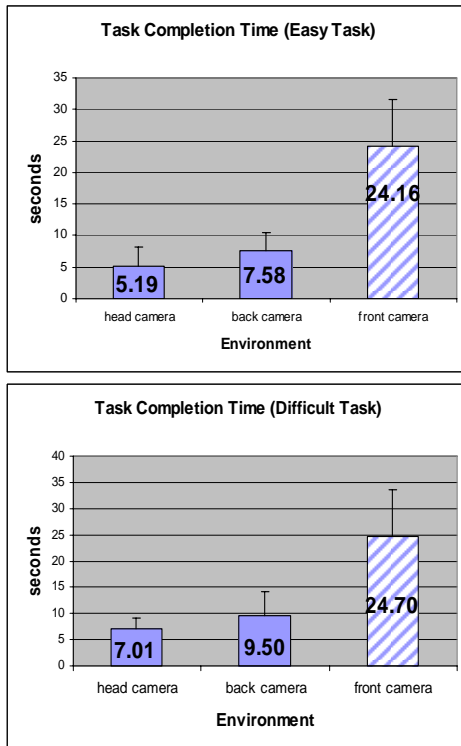


Fig. 5. The task completion time. Top: difficult task, $F(71,0) = 81.97$, $p < 0.0001$. Bottom: easy task, $F(71,0) = 120.99$, $p < 0.0001$.

Users also expressed their dislikes (asked comparatively after all the combinations were tried) for the front camera view/display condition as shown in Fig.6(a). Within-Subject non-parametric ANOVA (Friedman Test) revealed that there were significant differences in user preference among three conditions ($F(74,0) = 51.19$, $p < 0.0001$). For the post hoc comparison, the SNK (Student-Newman-Keuls) grouping was performed and it showed that the back and head camera configurations were not significantly different from each other, while front configuration was.

Fig.6(b) and (c) show the results for the other usability survey questions: user convenience and user fatigue respectively. The results as to the user convenience was similar to the user preference ($F(74,0) = 51.19$, $p < 0.0001$). However, the three viewing/display conditions exhibited three different levels of fatigue ($F(74,0) = 46.45$, $p < 0.0001$): the most tiring configuration being the front, the back, then head mounted. Fig.5 shows the task difficulty was not an important factor, hence, the task difficulty was not taken into account in the above three analysis.

We also looked at the experimental results between two groups characterized by their spatial ability based on the results from the MRT test: ones that belong to the high 30 percentile and the low. Fig.7 compares the task completion time between the two for the three test conditions. One can immediately observe the significant difference between the two groups for the case when the front camera was used, but not much so for the other cases. In addition, we performed an unpaired t-test

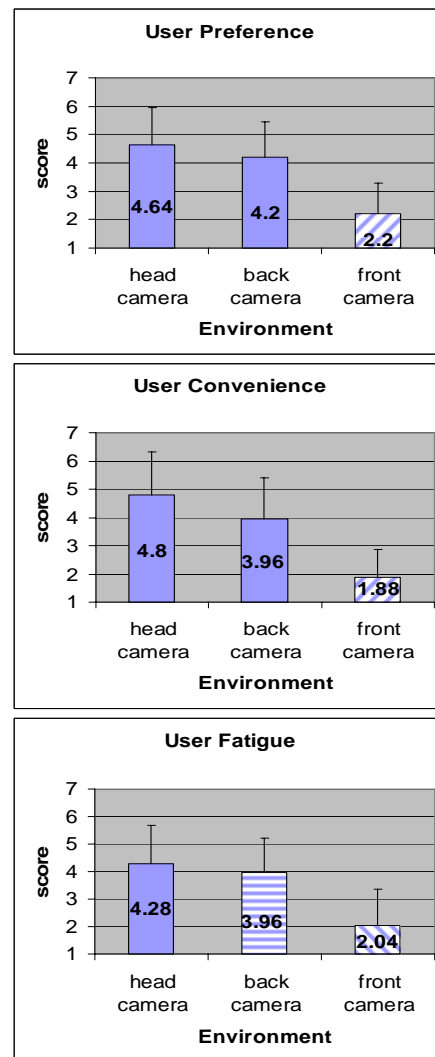


Fig. 6. Usability questions answers. Top: user preference ($F(74,0) = 51.19$, $p < 0.0001$), middle: user convenience ($F(74,0) = 51.19$, $p < 0.0001$), bottom: user fatigue ($F(74,0) = 46.45$, $p < 0.0001$).

(one-tail) to see whether there was a difference in task performance between experts (in using 3D user interfaces) and novices. The result of the test showed there was indeed a statistically significant difference between the two groups when the front camera was used ($p = 0.028$), but none when the head camera and the back camera was used ($p = 0.49$ for the head camera, $p = 0.33$ for the back camera), similarly to the results of the MRT.

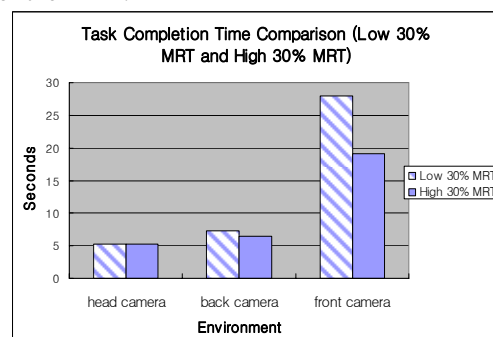


Fig. 7. Comparing between the high MRT score group and the low MRT score group.

A. Supplemental Experiment

While results from our main experiment indicate that the front view/display configuration was the worst in most aspects, it was unclear whether this could have been the result by the image warping, and not just the mirroring. As for the third display configuration, when the camera is located on the opposite side, merely mirroring camera-captured images horizontally and vertically is not sufficient to provide a natural view (See Fig. 4(b)), even though the directionality problem is only solved partially. That is, when the real hand moves from up, it looks as if the hand moves toward oneself and vice versa) to down and vice versa, violating our sense of proprioception.

To reveal the effect of the image warping technique in fixed front camera configuration, we asked 10 subjects to carry out the task described in Section III(A) in an mirrored-only (vertically and horizontally) configuration (See Fig.4(b)), and in mirrored+warped configuration (See Fig. 4(c)). The subjects task performance and the system usability was measured in a similar manner to the main experiment. We omit the detailed experimental procedure for lack of space.

Contrary to our expectation, however, the results showed that there was not statistically significant advantage of the mirrored+warped configuration in terms of both the task completion time ($P = 0.18$, paired T-Test) and usability ($P = 0.11$, paired T-Test). The raw data shows that the mirrored-only configuration resulted in faster task completion time (but only marginally and with no statistical significance). Thus, even though the image warping did not produce the expected effect, we conclude that the image warping could not have invalidated the main results. That is, even if the main experiment proceeded with only mirrored effect for the front camera configuration, we expect the main results would be the same.

V. DISCUSSION

From the analysis data, the first viewing condition (head camera) provides the best option (among the three tested) in terms of task performance, usability and convenience. Although there was not statistically significant difference in task performance between the head mounted camera and the back camera case, all the results consistently rated the head mounted case the best in terms of task performance and most of the usability questions. But considering the effort of the system set up, the second viewing condition (back camera) is also a good choice for the AR system without HMD. That is, it exhibited the same level of task performance as the case when the camera was mounted on the user's head and scored reasonably across the usability questions.

Our basic hypothesis that the mismatch of view points and the lack of proprioceptive cue affecting the task performance and usability has been shown. Especially in the front camera configuration, a directional mismatch of a vertical movement (if a user moves an object upward from the desk, the object seems like moving toward the user) and directional (of object facing) mismatch (if there is a cube like object on the desk, the

opposite face of the cube from the face in front of the user can be shown in the screen) is the most serious mismatch from user's proprioceptive cue. The analysis with regards to the spatial ability of the subjects (Fig. 7) further supports this conclusion. The experiment revealed that the spatial ability greatly affected the performance when the front camera was used. This shows the user needs more cognitive load (or mental rotation) to overcome the mismatched proprioception situation.

Although not tested, we can project the display condition of using an HMD would have produced the highest task performance. The subjective assessment of one's depth perception (from subject briefing) was the highest with the head mounted camera, then with the fixed back, and then with the fixed front. We can also project that usability wise, using the HMD would exhibit somewhat similar results to the case of using the head mounted camera. Although the mental/physical fatigue scores do not indicate it, 30% of the users of head mounted camera complained about having to wear the cap.

We also take a note that when the camera was mounted on the head, subjects reported problems, much more compared to the second test condition, with disappearing virtual objects (markers moving out of the camera's field of view and not being recognized anymore) due to the relative motion between the head and the manipulated markers. This is probably due to the head's "free" degree of freedom that is likely to cause more unintended relative motion unknowingly. The third configuration (also fixed camera, but in the front) also caused the same problem despite the fixation of the view, and we suspect that this was because of the reduced spatial capability by the distorted imagery (resulting in unintended motion of the manipulating hand). This prompts the need for a "video" mosaic or multi-camera system that can paint or cover a large interaction area and support significant head movements.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have compared the task performance and usability among three different viewing configurations of augmented reality system that uses a desktop monitor instead of a head mounted display. Our results indicated that mounting a camera on the user's head and placing a fixed camera in the back of the user was the best option in terms of task performance, user perceived performance and easiness of setting up the environment. The results should provide a valuable guide for designing desktop augmented reality systems without head mounted displays. We are currently continuing our study to compare our results to the usual AR systems that uses the HMD. Our other future work also includes the usability assessments with regards to other types of tasks and task conditions (e.g. larger interaction space, using stereoscopy, better warping). It is well known that there is another advantage of the head mounted camera configuration; wide potential field of view due to free head moving. We will address this effect in future work.

ACKNOWLEDGMENTS

This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment). We also thank Prof. Seungmoon Choi for giving us valuable comments and proofreading our manuscript.

interaction and innovative applications, presence, VR/AR authoring, experience based computing), and computer music.

REFERENCES

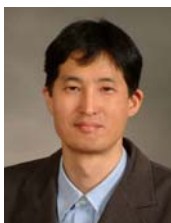
- [1] M. Fjeld and B. Voegtli. Augmented Chemistry: An Interactive Educational Workbench, in the video program of the International Symposium on Mixed and Augmented Reality (ISMAR), 2002.
- [2] J. Tian and D. Parkhurst. VSARD: A low-cost augmented reality system for desktop applications, Poster presented at the ACM SIGGRAPH Conference, Los Angeles, U.S.A., 2005.
- [3] K. Knowlton. Computer Displays Optically Superimposed on Input Devices, *The Bell System Technical Journal*, vol. 56, pp. 367-383, 1977.
- [4] O. Bimber, B. Fröhlich, D. Schmalstieg and L. M. Encarnação. The Virtual Showcase, *IEEE Computer Graphics & Applications*, vol. 21, pp. 48-55, 2001.
- [5] S. R. Fussell, L. D. Setlock and R. E. Kraut. Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks, presented at CHI'03: *Proceedings of the Conference on Human Factors in Computing Systems*, 2003.
- [6] B. Biocca and J. Rolland. Virtual Eyes Can Rearrange Your Body: Adaptation to Visual Displacement In *See-Through Head -Mounted Displays*, *Presence: Teleoperators & Virtual Environments*, vol. 7, pp. 262-278, 1998.
- [7] H. Kato and M. Billinghurst. Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System, in *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, San Francisco, USA, 1999.
- [8] S. Vandenberg and A. Kuse. Mental rotations: A Group Test of Three-dimensional Spatial Visualization, *Perceptual and Motor Skills*, vol. 47, pp. 599-604, 1978.



Seokhee Jeon (M'03, MS'05) Seokhee Jeon is a Ph.D candidate at Pohang University of Science and Technology(POSTECH) in Korea. He has completed both his undergraduate and master's degree at POSTECH as well. His main research interests are many topics in Virtual Reality and Mixed Reality, including haptic augmented reality, large immersive displays, augmented reality systems, and image based rendering.



Hyeongseop Shim is a B.S. candidate at Pohang University of Science and Technology(POSTECH) in Korea. His research interests are topics in augmented reality systems, peer-to-peer networks, and game theory.



Gerard Jounghyun Kim (M'87, MS'89, F'94) Gerard Jounghyun Kim is currently an associate professor in computer science and engineering at Korea University. Prior to joining Korea University, he has spent 10 years directing the Virtual Reality Laboratory at Pohang University of Science and Technology (POSTECH), Korea. He received a bachelor's degree in electrical and computer engineering at the Carnegie Mellon University (1987), and master's (computer engineering, 1989) and Ph.D. (computer science, 1994) degrees at the University of Southern California. Dr. Kim's main research interests include various topics in virtual and mixed reality and human computer interaction (3D