Precision study on augmented reality-based visual guidance for facility management tasks

Fei Liu\textsuperscript{a,b}, Stefan Seipel\textsuperscript{b,b,*}

\textsuperscript{a} Department of Industrial Development, IT and Land Management, University of Gävle, Gävle, Sweden
\textsuperscript{b} Centre for Image Analysis, Uppsala University, Uppsala, Sweden

Abstract

One unique capability of augmented reality (AR) is to visualize hidden objects as a virtual overlay on real occluding objects. This “X-ray vision” visualization metaphor has proved to be invaluable for operation and maintenance tasks such as locating utilities behind a wall. Locating virtual occluded objects requires users to estimate the closest projected positions of the virtual objects upon their real occluders, which is generally under the influence of a parallax effect. In this paper we studied the task of locating virtual pipes behind a real wall with “X-ray vision” and the goal is to establish relationships between task performance and spatial factors causing parallax through different forms of visual augmentation. We introduced and validated a laser-based target designation method which is generally useful for AR-based interaction with augmented objects beyond arm’s reach. The main findings include that people can mentally compensate for the parallax error when extrapolating positions of virtual objects on the real surface given traditional 3D depth cues for spatial understanding. This capability is, however, unreliable especially in the presence of the increasing viewing offset between the users and the virtual objects as well as the increasing distance between the virtual objects and their occluders. Experiment results also show that positioning performance is greatly increased and unaffected by those factors if the AR support provides visual guides indicating the closest projected positions of virtual objects on the surfaces of their real occluders.

1. Introduction

Augmented reality (AR) supplements the real world with virtual information through computer displays in an interactive manner. With this technology, a user’s sensory perceptions of the real world are enhanced, which allows him or her to understand the surroundings more thoroughly and therefore perform tasks in hand more efficiently. The potential capability of AR to fundamentally change the way people access useful information closely related to the world around them has attracted a vast amount of research and developments in the field during the past two decades, which leads to the incessant extension of application areas. When Azuma published his influential survey paper on AR in 1997 [1], AR was mainly explored in medicine, manufacture of complex machinery and military. Thirteen years later, not only did the AR applications in those traditional areas become more sophisticated but it had also seeped into personal information systems, offices, entertainment and education, as summarized in the survey of Van Krevelen and Poelman [2].

1.1. AR and built environment

Among these rapidly growing application areas of AR, the Architecture, Engineering, Construction and Facility Management (AEC/FM) industry has attracted much attention lately. This can be seen from the overall increased trend of AR-related publications in AEC/FM reported by [3]. AEC/FM projects usually involve a lot of information. According to [4], a typical large-scale project (10 million US dollars or more) can generate 50 different types of documents with 56,000 pages in total, equivalent to 3000 MB of digital contents. In view of such a large amount of information as well as the intensive demand of accessing it, the AEC/FM community has been vigorously adopting information technology to digitally manage various phases of a construction project [5,6]. The recent development in this aspect is the Building Information Modeling (BIM) technology [4,7], which aims to overcome information fragmentations existing in both project phases and collaborations between different roles [8]. Although BIM and related technologies greatly bring down the costs and increase the efficiency of construction projects in the form of information integration, construction practitioners still need to mentally apply the digital...
information to the physical world when they are performing such tasks as on-site planning, management and inspection [9]. BIM and related technologies provide little answer to the challenge of seeing and manipulating virtual data directly in the physical world where they are related [10]. AR, on the other hand, shows substantial promise in tackling such a challenge by presenting a natural user interface to the virtual information which overlays on the real world directly and thus minimizes possible mistakes caused by the disconnection between the virtual and the real world.

As vital parts of FM, Operations and Maintenance (O&M) can especially benefit from AR technologies [11]. This is owing to a unique capability of AR, which is often referred to as “X-ray vision” visualization [12]. Within the context of built environment, there are plenty of facilities and installations that are not readily visible for maintenance workers, for example, electricity wires inside a wall or a ceiling, pipelines buried underneath roads. The application of AR allows workers to see through solid objects and visualize maintenance targets in situ. Consequently, the usefulness of AR in terms of O&M is rather obvious.

On the other hand, in order for an AR application to be useful, the accuracy of registration between the real and the virtual world is of utter importance [1]. This is particularly true for O&M tasks dealing with hidden infrastructure in built environment. Generating virtual overlay of the infrastructure which accurately aligns with the occluding physical object is the prerequisite for further maintenance procedures, such as locating leaking pipes behind a wall or drawing up a plan for replacing current HVAC (Heating, Ventilating and Air Conditioning) utilities of a room. Furthermore, inaccurate registration for underground infrastructure can lead to mis-located excavation operations which waste time, money and even cause life danger [13,14]. Therefore, constant research efforts have been made to perfect the tracking and registration aspect of AR systems, with improved hardware and software technologies regularly reported. Some examples of the latest results can be found in [15-17].

However, no matter how accurate an AR system is, in the end it has to be put into the hands of users to realize its purpose. In other words, if we regard an AR system as a user interface into the virtual information, there is an interaction loop involving users perceiving its output, mental cognition processing the output and finally determining a motor action [18]. Therefore, the output received by users contains errors from both themselves and the system. In this study, we aim to explore factors that affect users’ ability to correctly establish spatial relationships between virtual objects presented through AR “X-ray vision” and real objects in the O&M setting. To this end, we designed and conducted user experiments which looked into an intuitive application scenario of locating virtual pipes behind a real wall (for an illustration see Fig. 1a) via a hand-held video see-through AR system.

1.2. Research objectives

In order to locate any 3D building component represented as a visual overlay upon real solid surfaces (e.g., walls), users of an AR system need to mentally establish its true spatial position in terms of its actual 3D position behind the real solid surface or, equivalently, in terms of its closest projected position on this surface. In our case of virtual pipes, these positions are the black lines on the wall plane as illustrated in Fig. 1a, or as indicated by O in Fig. 1b. Due to a spatial distance between the pipe and the wall, a naive determination of the pipe’s closest location on the wall would be exactly where the line of sight intersects with the wall, namely at point Q in Fig. 1b, which results in a horizontal parallax error $E_p$. According to Fig. 1b, the expected $E_p$ can in this case be predicted as

$$E_p = \frac{d - h}{d + w}$$

where $h$ is the horizontal offset from the viewer in front of the wall with respect to the pipe, $w$ is the distance between the viewer and the wall and $d$ is the depth of the pipe behind the wall. In a real situation, however, users of AR “X-ray vision” are expected to minimize $E_p$ based on the awareness of their spatial relation within the real environment and the understanding of the pipe’s spatial relation with the wall. Hence, we hypothesize that given sufficient depth cues for spatial understanding in AR “X-ray vision” visualization, users are able to mentally compensate for $E_p$. In addition, according to Eq. (1), the horizontal offset of the viewer position and pipe depth are expected to affect user’s performance in establishing the closest pipe location on the wall, which leads us to further hypothesize that lateral errors in designating pipe positions on the wall, if not fully compensated for by users, depend on $h$ and $d$.

Our experiments, which are described in Section 5, attempt to test the aforementioned hypotheses and investigate in what way different kinds of visual guides will affect users’ performance in determining the closest projected position of the pipe on the real wall through establishing the spatial relation between the pipe and the wall. Section 3 presents the technical details of our AR system developed to conduct the experiments.

2. Related work

AR has been reported to have positive impacts on tasks which demand high precision and/or accuracy from various application domains. Henderson and Feiner [19] found that subjects were significantly faster and more accurate with AR-based dynamic, prescriptive instructions than the ones presented through traditional 3D graphics and a stationary LCD monitor when it comes to psychomotor tasks, which are common in manufacturing and maintenance. AR has also been adopted to project visual cues on vehicle panels to improve precision and accuracy of manual welding in automotive manufacturing [20]. Physio@Home [21] is an AR prototype system for guiding proper physical therapy exercises. The authors reported that test participants performed most accurately with the visual guides.

--

\[Fig. 1. \] (a) Closest projected positions of some pipes on the wall (indicated by black lines); (b) top view of the parallax error $E_p$ for pipe position $P$.\]
provided by the system. Kytö et al. [22] proposed to use auxiliary attached to a wireless mouse.

Recognizing the potential benefits AR can bring, the AEC/FM field has seen plentiful AR-related research in recent years. Several published works have thus far provided theoretical frameworks for integrating AR into various phases of a building’s life cycle, which also serve as guides for AR research in the AEC/FM domain. Shin and Dunston [9] identified eight work tasks (layout, excavation, positioning, inspection, coordination, supervision, commenting and strategizing) in AEC whose performance can potentially be improved by the application of AR. Dunston and Wang [18] developed a methodology for breaking down AEC operations into five levels from a user-centered perspective in order to identify opportunities for applying AR technologies. Localization, natural user interface, cloud computing and mobile devices were recognized as the research trends of AR in AEC/FM through a summary of 101 research works in [23]. Comprehensive reviews of recent progress in AEC/FM-related AR research can be found in [3,24,25].

Compared with other areas of built environment, applications of AR in FM have not yet received much attention. The study in [11] demonstrates that AR-based user interfaces have the potential to improve maintenance fieldwork of building HVAC systems by reducing or eliminating information-related difficulties. Realizing the same performance bottleneck in facility maintenance, Koch et al. [26] also proposed AR as an alleviating measure and employed common indoor signs as natural markers for their AR maintenance system. Later they improved the navigation method of their system in [27]. A mobile AR tool called InfoSPOT [28] can enhance the situation awareness of facility managers by displaying information about facilities in question. Obrich et al. [29] present an AR framework for on-site BIM information access and collaborative user annotations during the O&M phase of a building’s life cycle.

In outdoor maintenance scenario, AR has mainly been adopted to visualize underground infrastructures for planning new installation, repairation or rehabilitation. The system described in [13] can superimpose geometrical models of subterranean objects (pipelines, electricity cables, etc.) on the view of an excavator driver according to inputs from GPS and geospatial databases in order to prevent him or her from damaging the subsurface utilities. Additionally, Smart Vidente [14] is a hand-held AR system for utility companies to survey and plan geospatial data on-site. According to the authors, the system satisfies the high demands of ergonomics, performance, accuracy and interaction for outdoor use. The work of Hou et al. [30] realized and studied a plant management system which utilizes augmented virtuality and AR to display facility data modeled and stored on a cloud server. Although it is intended specially for the oil industry, the system provides yet another real example of how AR can offer simple but effective information support for complex FM practice.

All the research presented above conveys an idea that AR is a new information interaction paradigm which can greatly boost the efficiency of FM activities in the built environment and the proposed systems along with the related tests have upheld this idea. However, none of them but [14] have provided quantitative error studies of their systems and/or users. These errors may have crucial influence on the performance of actual tasks, e.g., avoiding damaging underground utilities during excavation and understanding the errors can help mitigate them during the system design and the operation. On the other hand, although the system accuracy was reported in [14], the work does not include errors caused by user perception and cognition.

Locating pipes behind a wall through AR, like some of the systems discussed above [13,14,29], makes use of AR “X-ray vision” visualization. Being able to see hidden structures within our surroundings has many applications. Besides O&M tasks, this visualization metaphor has also been employed in medical diagnosis and surgery [31], outdoor navigation [32] and surveillance [33]. Since AR “X-ray vision” is such an unnatural way to view the real world, one of the challenges of realizing it is how to correctly convey the depth information, especially ordinal depth, of the augmented scene [34]. Clearly, a simple superposition of hidden objects with their occluders will only make them appear to float above the occluding objects. Hence, visualization techniques such as translucency [35,36], “cutaway” [37] and virtual tunnel [12] have long been employed to help people better understand the depth relationship of scene objects. More recent approaches [32,38,39] draw on image-based techniques to derive important regions of the occluding objects, which are represented by edges and/or visual saliency. These important regions are rendered after occluded objects to provide very convincing sense of occlusion. The subject of AR “X-ray vision” is more thoroughly treated in [40]. It is not our intention in this study to propose new visualization techniques for depth perception but rather we will investigate some of these well-established techniques when designing our experimental AR application, which is covered in the following section.

3. AR-based positioning system

Our test system comprises an AR application and a position tracking system (Fig. 2). The AR application presents co-registered virtual objects, e.g., a pipe, to users while the position tracking subsystem allows them to point at the virtual pipe and subsequently records the designated position.

3.1. AR application

The main function of the AR application is to offer an “X-ray vision” for users to see virtual pipes behind a solid wall. It was developed in the Unity game engine (version 4.6.3) with Qualcomm Vuforia mobile vision platform (version 3.0.9) and runs on a Microsoft Surface Pro 3 CI5 tablet. As we discussed in Section 2, rendering the minimal amount of virtual objects, in this case only the pipes, would poorly convey the spatial relationship between them and the real wall. To aid the depth perception of users, we added a ground-plane grid [35] along with a skirting board to represent the inner floor while making both the pipes and the ground-plane grid translucent. Additionally, two red ribbons were drawn on the imaginary wall plane to further enhance the sense that the pipe is behind the wall and also function as a marking zone, wherein pipe positions are expected to be marked up by subjects (see Fig. 3a). As another experiment condition, we provide an enhanced visual augmentation to facilitate better spatial comprehension. This enhanced visual augmentation is represented by the perpendicular
projection of a pipe upon the wall rendered explicitly in the form of a yellow line shown in Fig. 3b. These two experiment conditions will be elaborated in Section 5. All the virtual objects and a true-scale model of our lab room were created in Autodesk AutoCAD and then imported into Unity. The room model provided a visual reference as well as a world coordinate system for placing our virtual objects and was removed from the scene once the virtual objects were correctly situated in it. Fig. 3 also shows the layout of four virtual fiducial markers on the imaginary wall plane and their spatial relation with the other virtual objects in terms of the world coordinate system. Successfully tracking any of these markers is sufficient for Vuforia to register the virtual objects with the wall. The redundancy was introduced here to counter possible occlusion of markers during the experiments. Similar to the room model, the virtual markers were displayed for scene modeling purpose and were not rendered in the final application.

The AR application displays one pipe at any given time at one of several predetermined positions. For the purpose of the experiments, users can use a pair of buttons, labeled "Previous" and "Next" respectively, in the top right corner of the screen to scroll through the set of pre-determined positions and the system will render the pipe at the new position accordingly. Meanwhile, the index of the current position, e.g. "Pipe9", will also be shown at the top center of the screen. Fig. 4 is a picture of the AR-based interface as seen by the user and it illustrates how the visual overlay is co-registered with camera view of the real world. Notice the arrangement of real markers on the wall. They were put up at exactly the same places as their virtual counterparts in the world coordinate system to ensure accurate registration. We will discuss experiment procedures along with the detailed description of setting up pipe positions in Section 5.

3.2. Target designation and position tracking

Recording pipe positions estimated by users requires marking up those positions on the wall and then measuring them. Considering the size of the augmented object (the entire wall), the AR interface has to be used from a distance to contain its entirety in the view, which hinders users from marking up positions on the wall through immediate physical contact. Hence, some means need to be provided for users to designate target positions. To this end, a laser pointer was employed in the system. We attached it on top of a tripod which can only rotate horizontally (marked 3 in Fig. 2). The height of the tripod was adjusted such that the laser dot always appeared between the two red ribbons in the AR application. Once a user points the laser at the intended location on the wall, we need to obtain the coordinates of this point with respect to the world coordinate system. This was fulfilled through the Vizard precision position tracking (PPT) system, which is an optical tracking system consisting of four high speed cameras (marked 1 in Fig. 2), special light emitting probes (Fig. 2 inset), as well as software for calibration and tracking. The cameras track the blue light from the probe and after calibration, the accompanying software is able to report the probe’s position in the tracking coordinate system \( C_T \). Therefore, the current coordinate of the laser dot on the wall can be retrieved by simply aligning the probe with it (see Fig. 5). Additionally, in order to signal the tracking software to record a coordinate, we bond the probe...
to a wireless mouse and programmed the software to read in the coordinate whenever the mouse is clicked.

To transform \( C_T \) to the world coordinate system \( C_W \) in which the pipe positions are expressed, we introduced an intermediate coordinate system \( C_I \) based on 3 points on the markers, namely point \( O \), point \( P_1 \) and point \( P_2 \) illustrated in Fig. 6. Through aligning the probe with them respectively, we can obtain their coordinates in \( C_T \) which allows us to construct a \( 4 \times 4 \) matrix \( M_{T \rightarrow I} \). Meanwhile, we manually measured their coordinates in \( C_W \) with a ruler and similarly, another \( 4 \times 4 \) matrix \( M_{I \rightarrow W} \) can be constructed. With these two matrices, the coordinate transformation of any point reported by the tracking system into \( C_W \) can be expressed as
\[
P_W = P_T M_{T \rightarrow I} M_{I \rightarrow W}
\]
where \( P_W \) and \( P_T \) are both 4D row vectors representing the homogeneous coordinates of the point in \( C_W \) and \( C_T \) respectively. The whole transformation process was implemented as additional scripts running alongside the tracking software on a desktop computer. The final transformation results are stored in a text file for analysis later on.

4. Precision of position designation and acquisition

Prior to the actual experiments, we established the tracking precision in terms of a fixed reference position. To carry out this measurement, the probe was fixed on the wall within the area of interaction for the experiments. We then let the Vizard PPT optical tracking system capture 500 coordinate samples of the probe’s position. In this setting, we aim to identify the baseline precision of the camera-based tracking system under the specific environmental conditions in the experiment room.

Next, we incorporated the manual target designation to determine the precision limits of the whole target designation and acquisition process, which is used in the later experiments. For this measurement, a fixed reference position on the wall (also within the area of interaction for the experiments), i.e. the target, was designated by a test person with the laser pointer and then another test person retrieved the designated position with the probe in the manner described in the previous section. This measurement procedure was repeated 100 times with that fixed reference position and after each set of designation and acquisition, the laser was pointed away from the reference position randomly before the next iteration started. This process is affected by several factors comprising a) user’s performance in visually establishing target positions on the wall through the AR interface; b) user’s performance in manually positioning the laser dot with that target; c) the experimenter’s performance in manually aligning the probe with the laser dot; and finally d) the precision of the tracking system established above.

Fig. 7 shows a scatter plot of reported positions from both foregoing measurements with respective medians subtracted. As expected, there is much less variation in the data from the automatic recording of the fixed probe; but there seems to be some dependency between \( y \) and \( x \), which is most likely due to some systematic errors intrinsic to the multicamera optical tracking system. On the other hand, data from the manual procedure show a cluster with a less salient pattern. Three out of the hundred samples are not plotted because they are at distances farther than 8 mm from the center of the plotting area. For quantitative assessment of the precision, we further computed the Euclidean distance in terms of the centroid for each dataset and Fig. 8 summarizes the statistics. For the fixed probe, the median deviation of 500 samples was 0.13 mm. The upper quartile was at 0.27 mm and the 95% quantile was 0.56 mm, confirming the high precision of the tracking system. In regard to the manual repeated target designation and acquisition procedure, the median deviation of 100 repetitions was 0.75 mm.
upper quartile was at 1.0 mm and the deviation for 95% of the data was less than 1.6 mm from the centroid, which seems to be also very precise considering the various human-related sources of errors in the whole process.

5. Experiment

5.1. Task, conditions and stimuli

The general objective of our experiments was to evaluate users’ performance in establishing positions of structures (pipes) hidden behind walls when using AR-based guidance of different forms. As described in Section 1.2, the position on the wall means the shortest distance to the structure behind the wall, which is a valid assumption for a vast majority of tasks related to building construction and FM. For a systematic analysis of errors, we designed a layout of 12 pipe positions that are equally spaced both horizontally (x) and in depth (z), as illustrated in Fig. 9. The distance between neighboring positions is 0.2 unit (meter) along the z axis and 0.5 unit (meter) along the x axis. Meanwhile, we designate Position 1 as the reference point for implementing the layout in the world, whose x and z coordinates are -3.84 and -8.7115 respectively in Cw. In the vertical direction (y) pipes were aligned with the ground plane which coincides with the real floor.

The users’ task involved viewing the wall through the AR interface (see Fig. 4), judging the horizontal pipe position on the wall and designating this position using the laser pointer (Fig. 10). In our study we chose two conditions which differed in their levels of augmentation: in the weak augmentation condition we only visualized the ground-plane grid and the two red ribbons for providing 3D depth cues in addition to a virtual pipe (refer to Fig. 3a), whereas in the strong augmentation condition we also superimposed a yellow line on the wall indicating the perpendicular projection of the pipe on the wall plane (refer to Fig. 3b). A stimulus in the AR interface therefore consisted of the virtual pipe and its additional visual guide on the wall plane during the strong augmentation.

5.2. Experimental design and procedure

Participants in our study had to solve the task in both augmentation conditions, that is, two sessions one with weak and one with strong augmentation. In every session, they made 48 attempts, meaning four attempts for each of the 12 stimuli (pipe positions). In order to counteract learning and to avoid biasing effects based on position, we ensured that any two subsequent stimuli varied both in depth and horizontal position. This led us to four carefully designed groups of stimulus sequences which were presented in the same order to all participants. These four stimulus sequences are listed in Table 1. Also, we balanced the order of sessions such that half of the participants in our study, determined by random, started with the weak augmentation session while the other half started with the strong augmentation.

Before commencing the actual experimental procedure, the experimenter introduced subjects to the AR application and the task they were about to perform. After that, the subjects had the opportunity to familiarize themselves with the system and the task by practicing a demo session which included both the weak and the strong augmentation conditions. When the subjects felt ready, their first session of 48 attempts began. At each attempt, the subjects designated the pipe position and then signaled the experimenter to record the designated position with the probe. The transition to the next stimulus was initiated by the subjects through the AR interface. Meanwhile, our tracking software was programmed to record the time elapsed between each attempt automatically. Both sessions (96 attempts) were done in one run without any significant interruption.

5.3. Participants and background

We recruited 20 voluntary subjects for our experiments, 11 males and 9 females. They were either students or staff from the university. The ages ranged from 19 to 63, with an average of 32. Before a subject entered into the experiment, we conducted a brief structured interview based on a questionnaire. Apart from some basic demographic information such as gender and age, the questions focused on inquiring subjects’ previous experience with AR as well as 2D and 3D spatial perception-related tasks (e.g. gaming, 3D modeling). We also tested subjects’ capabilities to make metric spatial judgments by asking them to estimate the length of a metal pipe placed on the floor roughly 1 m away. After completing the experiments, the subjects were asked to rate the difficulty of the positioning task, both under the weak and the strong augmentation conditions on a Likert scale with six levels where one means very difficult and six means very easy.

<table>
<thead>
<tr>
<th>Group</th>
<th>Position sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>4, 6, 11, 8, 3, 12, 7, 1, 10, 5, 2, 9</td>
</tr>
<tr>
<td>Group 2</td>
<td>1, 7, 10, 5, 2, 9, 6, 4, 11, 8, 3, 12</td>
</tr>
<tr>
<td>Group 3</td>
<td>9, 2, 7, 12, 3, 5, 10, 8, 11, 4, 6, 1</td>
</tr>
<tr>
<td>Group 4</td>
<td>12, 3, 6, 9, 2, 8, 11, 5, 10, 1, 7, 4</td>
</tr>
</tbody>
</table>
6. Results

The primary data gathered from our experiments comprise response times for every attempt from subjects and the designated position retrieved by the experimenter. Since this study investigates errors in spatial positioning, reported positions were transformed into a deviation measure which is the horizontal distance $D$ between the designated position and the corresponding actual pipe position for every attempt, namely,

$$D = X - \bar{X}$$  \hspace{1cm} (3)

where $X$ is the x component of the actual pipe position while $\bar{X}$ is the x component of the designated position. Positive values of $D$ indicate deviations to the right of the true position as seen by the viewer looking at the wall. The main factor investigated here is the form of augmentation, weak and strong as described in the previous section. In the subsequent analysis we also grouped the data by other factors such as depth of the pipes behind the wall and lateral displacement, or offset for short, of the pipe position in relation to the viewer position (refer to Fig. 9). Data analysis and statistical tests were carried out using the statistical software R.

Initial data exploration using Q-Q plots revealed that the distribution of observed deviations as well as times was highly skewed in all settings, which led us to adopt non-parametric statistical methods where applicable. It was also found that one participant in the test quite obviously misunderstood the task and produced systematically deviating measurements. Those data were excluded from further analysis which leaves us with a total of 19 (subjects) × 4 (offsets) × 3 (depths) × 4 (attempts per pipe position) = 912 recorded positions (or deviations) and response times in each of the two sessions (weak and strong augmentation).

Since our experiments followed a within-subject design, we used in the following data analysis pairwise comparisons and Wilcoxon signed-rank tests (test statistic V) for factors augmentation, depth, and offset. On the other hand, the analysis of order effects (Section 6.2) and subject-related effects (Section 6.3) implied comparisons between different groups and hence we used Mann-Whitney-Wilcoxon tests (test statistic W) there.

6.1. Deviations of reported positions

An initial contrast of lateral deviations $D$ between the two main conditions (weak and strong augmentation) did not show any shift in location for the entire dataset. But a more detailed analysis of the data revealed huge differences in the distributions of reported positions and factors other than augmentation affected users' precision.

6.1.1. Deviation for different viewing offsets

In Fig. 11 data was first grouped by horizontal pipe positions (viewing offsets) and the boxplot summarizes the statistics for each of the four offsets in the two augmentation conditions. The statistics shown in each box is based on 228 (912/4) data points.

Fig. 11a shows deviations from true pipe positions when weak augmentation was used. Quite evidently, the interquartile range (IQR) increases remarkably as the viewing offset increases to the side; that is, the reproducibility of measurements decreases. A test for homogeneity of variances using a Fligner-Killeen test lets us reject the hypothesis of equal variances ($\chi^2 = 47.5226$, $df = 3$, $p < 0.0001$) confirming significance of differences in dispersion stated above. Lateral deviations $D$ occur at both sides of the true pipe positions, but statistically (in terms of median deviation, see Table 2) they are positively shifted for all offsets, which is in agreement with what the model (Eq. (1)) predicts when parallax errors are not entirely compensated for by the user. The difference of means is statistically significant ($p < 0.001$) for all pairwise comparisons between offsets, except when contrasting offset at 1.0 m and 1.5 m.

When strong augmentation was used (see Fig. 11b), precision was generally much better in terms of lower IQR compared to weak augmentation. Additionally, IQR also decreases with increasing lateral viewing offset in strong augmentation. Observed differences in variance are statistically significant as a Fligner-Killeen test with offset as factor confirms ($\chi^2 = 4$, $df = 3$, $p < 0.0001$). Also, in a comparison between weak and strong augmentation, the dispersion (IQR) differs by an order of magnitude for large offsets (see Table 2). Looking at the levels of deviation, there is an interesting pattern with strong augmentation. For small viewing offsets ($\leq 0.5$ m), deviations of reported pipe positions from true positions are almost twice as large as for larger offsets, where they decrease to very low levels. See Table 2 for exact figures. Those differences of $D$ (medians) in the strong augmentation condition are relevant effects. Wilcoxon tests were highly significant ($p < 0.001$) for all possible pairwise comparisons between the four offsets.

6.1.2. Deviation for different pipe depths

In order to assess the potential influence of pipe depth on the errors of reported lateral pipe positions on the wall, we analyzed the data for depth in each augmentation condition. The boxplot in Fig. 12 summarizes the statistics. Each boxplot is based on 304 (912/3) observations. At first glance, a pattern similar to the variation for increasing offset is evident; i.e., there is substantially larger variation of deviations for the weak augmentation condition when compared with the strong one. Exact values for IQRs and medians of $D$ are shown in Table 3. They indicate that the variance doubles for every level of increasing depth for weak augmentation. This inhomogeneity of variances is significant, as a Fligner-Killeen test confirmed ($\chi^2 = 76.3539$, $df = 2$, $p < 0.0001$). For weak augmentation, there is also a clear positive shift of the deviation $D$ (medians) for increasing levels of depth. The increase is significant when we compare depth = 0 m with depth = 0.4 m ($V = 15706$, $p < 0.001$) and compare depth = 0.2 m with depth = 0.4 m ($V = 12616$, $p < 0.001$).

The strong augmentation condition results are notably different. Both the dispersion (IQR) and lateral deviation $D$ (medians) are apparently constant for all pipe depths (see Fig. 12b and Table 3) and no significant differences could be found.

6.2. Efficiency and learning

Analysis of user responses shows a difference between the two main experimental conditions regarding times used by subjects to set out pipe positions. The median of times for the weak augmentation was $\bar{t}_w = 13.1$ s and $\bar{t}_w = 10.6$ s for the strong augmentation. This time difference is significant ($V = 350921$, $p < 0.001$). Strong augmentation helped users to faster identify pipe positions.

Furthermore, subjects in our experiment needed more time for pipes that were positioned behind the wall at any depth > 0.0 m and the response times were also increasing for larger pipe offsets. Fig. 13 shows this interrelation. Among the different levels of depth, time differences are significant for depths between 0.0 m and 0.2 m ($V = 59557$, $p < 0.001$), between 0.0 m and 0.4 m ($V = 68486$, $p < 0.001$), as well as between 0.2 m and 0.4 m ($V = 105147$, $p = 0.004$). As for the offset, the increases in times are significant between every incremental level, which is listed in Table 4.

Contrasting responses by the session order reveals that regardless of the type of augmentation, users performed faster in the second session of the experiment, while deviations did not differ, i.e. users were equally precise in both sessions. The decrease in times between sessions is, on an absolute scale, small with $\bar{t}_{sw} = 12.0$ s and $\bar{t}_{sw} = 11.1$ s, yet statistically highly significant at $a = 0.025$ ($W = 482891$, $p < 0.0001$). A more stratified comparison of time differences by session order is presented in Fig. 14. It shows that learning effects occurred in both augmentation conditions with statistical significance. Detailed median time values for session orders and conditions as well as p-values are shown in Table 5.
6.3. Subject-related factors and responses

We associated the subject-related background factors collected from the interviews with observations of deviations and times for further analysis. These background factors included gender, age, prior experiences with AR, 3D games, 2D/3D content authoring applications and visual capabilities. Subsequent comparisons of times and deviations between various aforementioned factors revealed significant effects only for gender. Female participants in our study performed faster in target designation than male counterparts but as a group, male subjects performed with significantly lower errors (compare Table 6). In other words, using more time in this task made users to perform more precisely.

Participants in our study were also asked to rate the ease of designating pipe positions using the weak and strong augmentation based on a Likert-scale with six levels, where 1 means very difficult and 6 means very easy. The average grade obtained for the weak augmentation was 3.15 compared to 5.5 for the strong augmentation.

7. Discussions and conclusions

Our study investigated the usability of AR as a tool to support positioning tasks in FM. In this endeavor we first looked into the validity of the proposed solution to AR-based target designation. It addresses the need for remote target designation which arises from the limited field of view in video see-through AR systems thus forcing the user to

---

Table 2
Statistics for deviations \( D \) in different conditions for increasing viewing offset.

<table>
<thead>
<tr>
<th>Offset (m)</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>39.6</td>
<td>37.2</td>
<td>48.1</td>
<td>71.8</td>
<td>7.4</td>
<td>9.7</td>
<td>9.4</td>
<td>9.4</td>
</tr>
<tr>
<td>Strong</td>
<td>9.4</td>
<td>9.5</td>
<td>6.5</td>
<td>3.8</td>
<td>13.2</td>
<td>16.3</td>
<td>6.8</td>
<td>−0.6</td>
</tr>
</tbody>
</table>

Table 3
Statistics for deviations \( D \) in different conditions for increasing depth.

<table>
<thead>
<tr>
<th>Depth (m)</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>19.7</td>
<td>47.4</td>
<td>83.4</td>
<td>6.0</td>
<td>8.7</td>
<td>17.5</td>
</tr>
<tr>
<td>Strong</td>
<td>12.9</td>
<td>13.8</td>
<td>13.3</td>
<td>7.9</td>
<td>8.1</td>
<td>8.6</td>
</tr>
</tbody>
</table>

---

Fig. 11. Boxplots of reported horizontal deviations \( D \) from true pipe positions grouped by pipe offset with weak (a) and strong (b) augmentation.

Fig. 12. Boxplots of reported horizontal deviations \( D \) from true pipe positions grouped by pipe depth with weak (a) and strong (b) augmentation.
inspect large objects at a distance. We then evaluated quantitatively users’ capabilities of utilizing spatial information mediated through the AR interface for precise positioning tasks.

7.1. AR-based target designation

To enable users to mark up intended targets on physical objects which are beyond arm’s reach, we proposed a laser designation approach whereby the laser point is interactively aligned by the users with intended targets perceived through the AR interface. While this approach is low-cost, technically robust and intuitive to use, its potential benefit may be limited by environmental factors (e.g. illumination), AR system performance and users’ visual acuity. In our experiments it showed that for an indoor environment, contrast of the laser dot in AR images was excellent even at very intense illumination levels, which enabled easy on-screen identification. These premises however may vary for different hand-held AR platforms and furthermore such excellent visibility may not be obtained in all outdoor circumstances. During the design phase of the experiments we found that a much too intense laser pointer, instead, resulted in a saturated white spot in the AR images rather than in its original color, which made it difficult to discern against the white wall. On the other hand, the size of a pixel, as determined by the display and camera hardware as well as by the viewing distance from the augmented object will determine the actual size of the pixel footprint on the real object and hence delimit the smallest detail that can be distinguished by the user using the AR interface. Quantitatively, our measurements for repeated position designation showed that deviation was less than 1.6 mm in 95% of the cases (see Section 4), which is more precise than expected considering the variety of potential error sources. Comparing to the medians of various deviations in Tables 2 and 3, the aforementioned 1.6-mm deviation is at least an order of magnitude less than the observed deviations in the main experiments and thus accounts for little of the errors found there. Consequently, we can claim that the laser-based designation of target positions is a suitable method for many precision

Table 4
Pairwise comparisons of median response times (seconds) for increasing pipe offset (meters).

<table>
<thead>
<tr>
<th>Offset</th>
<th>Offset 0.5</th>
<th>Offset 1.0</th>
<th>Offset 1.5</th>
<th>V</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.86</td>
<td>10.97</td>
<td>39,850</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.97</td>
<td>11.59</td>
<td>39,997</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.59</td>
<td>12.43</td>
<td>34,407</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Median response times (seconds) in 1st and 2nd session for weak and strong augmentation.

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>W</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>14.5</td>
<td>12.1</td>
<td>127,159</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Strong</td>
<td>10.9</td>
<td>10.2</td>
<td>121,698</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>All</td>
<td>12.0</td>
<td>11.1</td>
<td>482,891</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Table 6
Differences in precision and completion time for gender.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>W</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\overline{d}[\text{mm}])</td>
<td>6.8</td>
<td>11.0</td>
<td>351,161</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>(\overline{t}[\text{s}])</td>
<td>11.9</td>
<td>10.9</td>
<td>493,694</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Fig. 13. Boxplots of response times for different depths (a) and horizontal offsets (b). Times are significantly longer for pipes at depths larger than 0.0 m and growing for increasing pipe offset.

Fig. 14. Boxplots of times for weak and strong augmentation in terms of session order.

Fig. 15. Boxplots of condition and order times for weak and strong augmentation.
positioning tasks where interaction through an AR interface beyond arm’s reach is required.

7.2. User task and experimental design

The authors in [11] pointed out that in-house FM services personnel had difficulty locating exact parts of HVAC systems such as pipes and ducts for maintenance while one of the motivations for developing the system in [14] was to enable employees of utility companies to visualize underground infrastructures such as water pipes, gas pipes and power cables for surveying purpose. These are the rationales of our choice for the user task in this study and locating hidden objects behind a solid object employing AR “X-ray vision” clearly has its real world applications in FM. On the other hand, reducing real world utilities to a di

When designing the user experiments, we had taken potential effects of learning into consideration. The two augmentation conditions were therefore carried out by subjects in separate sessions with counterbalanced order. In designing the sequences of pipes presented to the subjects, we ensured that no successive stimuli would have either the same horizontal offset or the same depth. Also, all four sequences were different. As the results in Section 6.2 show, learning effects in terms of reduced task completion times did overall exist for successive sessions, but our experimental design counteracted the carry-over effects in comparisons between augmentation, depths and offsets.

7.3. Parallax effect in AR “X-ray vision”

Prior to the experiments, we hypothesized that 1) given sufficient depth cues for spatial understanding in AR “X-ray vision” visualization, users are able to mentally compensate for $E_p$; 2) if $E_p$ is not fully compensated for by users, then it depends on $h$ and $d$. In order to verify these two hypotheses, we plotted stratified median deviations in terms of offset and depth respectively for the two augmentation conditions together with the theoretical deviations computed from Eq. (1) in Fig. 15. If we exclude the zero cases, the fact that parts of the two red dash lines (the weak augmentation) starting at 0.5 m offset are far below the corresponding parts of the blue lines (theoretical deviations) proves that such mental compensation exists. On the other hand, we can see that the red dash line representing $depth = 0.4 \, m$ situates above its counterpart of $depth = 0.2 \, m$ and in the meantime both lines rise as the offset increases. These two phenomena indicate that the median deviation increases as the offset or the depth increases, which supports our second hypothesis. Such dependency is also evident from the positioning precision expressed in terms of IQR of the lateral deviation. According to the results in Section 6, they clearly show that positioning precision decreases as either the offset or the depth increases for the weak augmentation condition. The lowest precision was observed at the extreme cases for both factors, namely, 71.8 mm at offset = 1.5 m and 83.4 mm at depth = 0.4 m. One exception to the increasing trend of IQRs happens at offset = 0 m. Given that there is no viewing offset, the subjects should have achieved a rather good positioning precision, similar to its counterpart at depth = 0 m, which is 19.7 mm shown in Table 3 and much less than the IQR of its neighboring depth level.

Fig. 15. Modeled and observed errors with various factors.

In reality, however, the IQR at offset = 0 m is 39.6 mm and even greater than the IQR of its neighbor (refer to Table 2). A possible explanation for this is that we never imparted any information on pipe positions to the subjects. Without knowing there were pipes at zero offset, they still tried to mentally compensate for a hypothetically existing parallax error when they encountered pipes at such positions. We did observe this kind of behaviors during the experiments especially when the depths were non-zero. Apparently, any compensation made in this case would contribute to the deviation. Whereas for the case of zero depth, there was a clear visual cue from the ground grid that the bottom of the pipe touched the skirting board and hence most subjects knew they should only need to point at the pipe itself.

Although people are capable of counteracting the parallax error to some degree when utilizing spatial 3D cues, as discussed above, position assessment is yet not reliable and thus questionable for tasks requiring high precision. In light of this, we added a strong visual guide, which is a yellow line on the wall plane representing the closest position of the pipe on the wall and studied its effect on the same task. The most conspicuous observation from the results in Section 6 is that the precision has then been greatly improved, no matter at which offset or depth the pipe was positioned. At the extreme positions, the precision (in terms of IQR) has increased from 71.8 mm to 3.8 mm for offset = 1.5 m and from 83.4 mm to 13.3 mm for depth = 0.4 m. Such an improvement in performance also manifests itself in the task completion time where the weak augmentation took 13.1 s while the strong augmentation needed 10.6 s on average, as reported in Section 6.2. Comparing the statistics of the strong augmentation condition in both Tables 2 and 3, we find that the statistics for the offset factor sees an increase at 0.5 m and drop rapidly afterwards whereas the ones for the depth factor remain almost constant for all levels. One likely reason is that when the pipe position is close to the perpendicular viewing direction, the pipe and the strong visual guide overlap or are close to each other, which causes user confusion. For increased offsets, the distance between the pipe and the strong guide grows so that they become more distinct from each other and therefore users are more determined in identifying the guide.

Plenty of research has pointed out the importance and the challenges of creating visual depth cues to help AR “X-ray vision” users with better spatial understanding of the scene, e.g. [40] and [34]. Based on the findings from this study, we would like to emphasize that the inevitable parallax error introduced by the distance between hidden objects and their occluders, with the presence of viewing offset, are influential factors when it comes to precision positioning tasks in the context of FM. A mobile AR setup can arguably mitigate some of the error sources to a great extent, but in situations where stationary AR-tools are preferred or when a perpendicular viewing position is physically impossible to access, AR tools have to provide an intrinsic means to counteract this parallax effect. Moreover, even with the mobile setup, it is not guaranteed that users can find the exact perpendicular view with respect to hidden objects in question. Therefore, a strong
visual guide representing the closest position of a hidden object on its occluder is almost necessary for AR “X-ray vision” solutions to FM tasks.

7.4. Conclusions

In this work, we first brought forward a laser-based target designation method which aims to solve the dilemma that AR users have to move away from a large augmented object in order to capture its entirety in the view while they would like to interact with the physical object at the same time. The pre-study of precision upholds its validity and suggests that the method is feasible for positioning tasks in FM.

Since utilities and equipment are not installed immediately behind a wall, under a road, etc., correctly locating their positions in relation to their occluder will need to factor in the parallax error. Through the user study, we have found that although people do mentally compensate for the parallax error given sufficient depth cues for spatial understanding, spatial assessment of target positions is very unreliable especially when their occluder will need to factor in the parallax error. Through the user move away from a large augmented object in order to capture its en...

Acknowledgments

The authors would like to thank Torsten Jonsson for the fruitful discussions on the experiment design and support during the experiments.

References


Ch. 1. Augmented reality visualization for laparoscopic surgery.

