An Evaluation of Graphical Context as a Means for Ameliorating the Effects of Registration Error

Cindy M. Robertson, Blair MacIntyre, and Bruce N. Walker

Abstract—An ongoing research problem in augmented reality (AR) is to improve tracking and display technology in order to minimize registration errors. However, perfect registration is not always necessary for users to understand the intent of an augmentation. This paper describes the results of an experiment to evaluate the effects of registration error in a Lego block placement task and the effectiveness of graphical context at ameliorating these effects. Three types of registration error were compared: no error, fixed error, and random error. These three errors were evaluated with no context present and some graphical context present. The results of this experiment indicated that adding graphical context to a scene in which some registration error is present can allow a person to effectively operate in such an environment, in this case, completing the Lego block placement task with a reduced number of errors made and in a shorter amount of time.

Index Terms—Artificial, augmented, and virtual realities, evaluation/methodology, communicative intent, augmented environments, human-computer interaction.

---

1 INTRODUCTION

AUGMENTED reality (AR), overlaying graphics on the physical world, has been shown to be a useful user-interface paradigm for various application domains. Many AR systems use see-through head-worn displays to merge the virtual elements into the user's view. The main benefit of using AR is that the graphics are displayed in situ and support hands-free interaction. However, registering the graphics with objects in the physical world poses a significant problem. In order to align the graphics with physical objects, both the user and the objects must be accurately tracked (at least with respect to each other), and the whole system (including the HMD) must be accurately calibrated. At this time, tracking and display technologies are not accurate enough to produce perfect registration between the graphical world and the physical world. In some domains, such as medicine, accurate registration is required. However, we believe that, in many situations, precise registration is not as critical.

Consider, for example, the need to visualize the location of one of the serial ports on the back of a computer workstation. There are many ports and plug-ins in the back of the computer that could be confused for the intended serial port. If there is registration error present in the AR system, how does the user know which serial port is the intended target of the visualization? However, if the port in

question is below a unique feature (such as a large button), adding a representation of the button to the augmentation may be enough to allow a person to choose the correct port.

In our previous work, we discussed the AIBAS system, an adaptive intent-based augmentation system designed to use the communicative intent [8] of an augmentation to simplify the creation of AR applications that work in real-world situations with "good enough" tracking [7]. Our goal was to enable programmers by providing them with a framework to create augmentations that function in the presence of registration error. Our group has also modified an open source scene graph (OpenSceneGraph [6]) to support the specification of uncertainty at its transformation nodes [1]. These values can then be used to estimate the registration error associated with the objects in the scene graph. Using this estimate, we can design augmentations that adapt to changing registration error.

We believe that the key to making adaptive augmentations work is the use of *visual context cues* (i.e., graphical information that provides visual cues about the relationship between the virtual and physical world). Visual context can be added to an augmentation to help the user understand the intent of the augmentation. We discuss this concept in detail in our previous publications (e.g., [7]). In this paper, we want to focus on evaluating the effectiveness of adding graphical context to an AR environment. The addition of this visual context raises many more questions. What is the best way to display the augmentations? Which augmentation should be used in which situation? How should transitions between different augmentations be handled? Would different ways of displaying the data be more effective than others? How much augmented information is enough? Is there a limit to how much information is helpful? Can too much information become intrusive?

In order to begin to answer even some of these questions, we first need to understand more about the effects of registration error and visual context on people submerged

---

- *C.M. Robertson and B. MacIntyre are with the School of Interactive Computing, Georgia Institute of Technology, 85 5th Street NW, Atlanta, GA 30308. E-mail: cindy.robertson@gatech.edu, blair@cc.gatech.edu.*
- *B.N. Walker is with the School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170. E-mail: bruce.walker@psych.gatech.edu.*

in an AR environment. The goal of this paper is to show that providing visual context is indeed a useful tool in battling the effects of uncertainty. We will begin by evaluating the various types of registration error and how users react to their presence in an AR system. We will then provide context, in the form of graphical augmentations, in the same setting and evaluate its effectiveness. This paper will show that adding context to a system can alleviate some of the problems caused by the registration error.

2 RELATED WORK

User-based experimentation in AR is an emerging field, so there are relatively few experiments described in the AR literature and even fewer that relate to this research. Tang et al. compared the effectiveness of augmented instructions in an assembly task [9]. That user study showed that the use of AR in the form of computer-assisted instruction projected on a head-mounted display can improve task performance and can relieve mental workload as compared to a printed manual and computer-assisted instruction using a monitor-based display.

Livingston et al. conducted a user study to determine which display attributes, including drawing style and opacity, best express occlusion relationships among far-field objects [4]. They found that response times for a task in which the users had to determine the location of a target were slower with a "wire" drawing style than for "fill" and "wire + fill" drawing styles. These later styles produced comparable response times. However, they found that participants made the fewest errors with the "wire + fill" task. They speculated that this style was most effective because it combines occlusion properties by using the "fill" style with wireframe outlines, which pronounce the targets' shapes. While we are not explicitly concerned with occlusion in our studies, we are concerned about the most effective drawing style for representing our augmentations. This study provided insight when designing our augmentations.

3 THEORY AND HYPOTHESES

AR has not been widely used for a variety of reasons, but we believe there are two main reasons. First, the focus has been placed on perfectly registering the graphics with the physical world. This is impractical for many reasons, ranging from the expense, or unavailability, of precise trackers in mobile situations. Second, there is a concern voiced occasionally about the inappropriateness of having computer graphics block a worker's view of the task space, thus interfering with their primary task. This could not only be annoying, but in certain tasks, it could also be very dangerous.

We believe that AR can actually be designed to work in both of these situations; in this paper, we focus on evaluating a possible solution to the first of them. We will evaluate the effectiveness of graphical context in combating registration error, showing that the assumption that graphics need to be perfectly registered is misguided.

3.1 Types of Registration Error

In order to prove that context can help ameliorate the effects of registration error, we first define the types of registration error used in this study because they are indicative of some of the more common types of error in AR. We categorize the registration errors that will be evaluated into three different types: *no error*, *fixed translation error*, and *random translation error*.

No error. When there is no visible misalignment between the graphics and the world, we can say there is no error. For our purposes, however, achieving absolutely no registration error in an AR system is impossible, so our *no error* case could actually be considered *unambiguous error* or *negligible error*. In the case of our experimental setup, if the amount of registration error is less than half of the size of one of the Lego pegs on the base plate, there is no question as to where the block should be placed; therefore, we consider this to represent no error. From here on, we will also refer to the no error case as perfect registration.

Fixed translation error. When the error consistently manifests itself in the same direction and offset, we call this fixed error. For example, if the offset is always up and to the right 2 Lego pegs, this is considered a fixed translation error.

Random translation error. When both the direction and magnitude of the error are completely unpredictable, we consider this to be random error. Given these three types of errors, we have six conditions: each of the three errors in a context-free environment and in an environment where some visual context is provided. We intentionally do not include orientation errors in this study because comparing the results when there are both orientation and translation errors would be difficult.

3.2 Hypotheses

When exposed to different types of error in a scene in which no context is provided, there will be a different user response for each type of error. Therefore, we predict:

1. When exposed to a fixed error, users will gradually learn how to compensate for the error; however, when exposed to random error, the task will become a guessing game as to where the block should actually be placed.

If context has the effect of providing useful clues as to the relationship between the physical and virtual worlds, placement tasks should be able to be performed with fewer errors; however, the cognitive processing of the context information might increase the trial times. Therefore, we predict:

2. When context is added to a scene, whether fixed or random error is present, the time per trial will increase, but the total amount of errors will decrease.

When there is no error in an AR system, the user does not need additional context to perform a placement task correctly. Therefore, we predict:

3. When there is no registration error associated with the system, adding context neither increases nor decreases the number of errors.
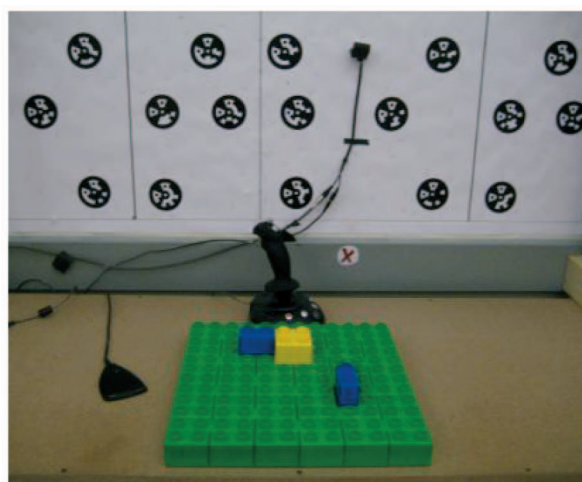
Fig. 1. The setup.

4 METHODOLOGY

To maximize statistical power and experimental efficiency, a within-subjects experiment was conducted. There were two independent variables, the type of error presented (none, fixed, and random) and the amount of context presented (no context and some context). The dependent variables include time to complete each task, the number of errors, and perceived mental workload.

4.1 Participants

Twenty-six participants took part in this study. Participants were solicited via email to on-campus mailing lists as well as by requesting student volunteers from classes. The participant pool consisted of 26 participants, 14 male and 12 female. The ages of the participants ranged from 19 to 56. As compensation for their time, the participants received $5/half hour.

4.2 The Setup

Fig. 1 shows pictures of the experiment setup we used in this experiment. The participants stood next to the desk, shown in Fig. 1a, on which a Lego base plate was located in a fixed position relative to fiducial markers that were hung on the wall in front of them. They wore the biocular

head-mounted display, shown in Fig. 1b, that contained an Intersense IS-1200 tracker, a 60-frames-per-second Point Grey Flea camera, and a Sony Glasstron optical see-through display which was modified to work as a video see-through display. The camera is mounted above a right angle prism, moving the optical center of projection of the camera closer to the participant's eyes than would otherwise be possible, with the intent of reducing the parallax offset of the video-mixed head-worn display.

4.3 Session Information

Participants were asked to complete an introductory questionnaire to provide some background information at the beginning of each session, including age, experience in AR systems, video game experience, how well they understood the concept of registration error in AR, etc. They were then asked to complete two tasks: an Edinburgh [5] handedness test and the Spatial Learning Ability Test [2]. We used the handedness test to ensure that they used their dominant hand to complete the experiment. The spatial ability test was given to evaluate the relationship between spatial abilities and successful task completion. Participants were then trained with an error training document as well as a training exercise to familiarize them with wearing an HMD, how to correctly perform the block placement task (with both no context and context present), and how to ensure that they maintain proper tracking throughout their trials. They were reminded that they would be evaluated based on the amount of time that it would take them to place each block as well as the number of errors (if any) they would make while placing a block; therefore, it was important for them to work as quickly and as accurately as possible.

4.4 The Placement Task

The task consisted of the following: picking up the yellow block, pushing a button to start the trial, placing the block, and pushing the same button to end the trial. After each trial, the participant was asked how confident they were with the block placement. The following five-point Likert scale was used:

1. I think the block is in the wrong place.
2. I think the block might be in the wrong place.
3. I don't know.
4. I think the block might be in the correct place.
5. I think the block is in the correct place.

If the block was indeed placed correctly, the participant was informed as such and advanced to placing the block in the next location. If the block was placed incorrectly, the participant was informed as such, and was instructed to attempt to place the block in the correct location again. The steps repeated until the block was correctly placed. After all of the trials were completed, the participant answered a survey questionnaire about their experience, including portions of the NASA TLX rating questionnaire [3]. It is important to note that we did not do a complete NASA TLX evaluation as Hart and Staveland did in our analysis. Instead, we used their six Likert scales for mental demand, physical demand, effort, frustration, temporal demand, and
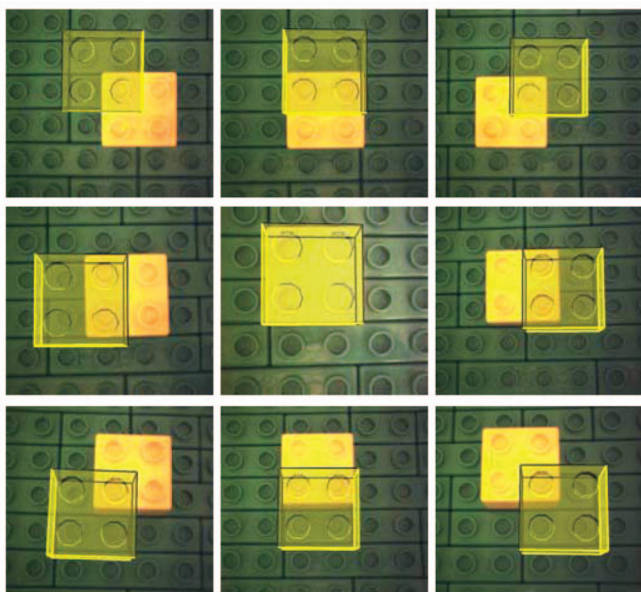
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS,   VOL. 15,   NO. 2,   MARCH/APRIL 2009

Fig. 2. The nine possible locations for block placement.

(a)

(b)   (c)

Fig. 3. Context blocks. (a) No context. (b) Context. (c) Context with error.

perceived performance as a model for evaluating subjective workload in our study.

While this study can be considered a follow-up study to the study done by Tang et al. to compare the effectiveness of augmented instructions in an assembly task, we have chosen to have each of the tasks be individual block placements rather than one large building task [9]. We designed the experiment this way because we did not want the errors to compound, as they would have if they were completing a building task. An error made in one trial does not affect the next trial as it does in Tang et al.'s experiment.

In our initial design of the study, there were a total of six blocks of trials: no error and no context, fixed error and no context, random error and no context, no error with context, fixed error with context, and random error with context. Whether there was no context or some context present in the block of trials, the following are descriptions of the types of error tested:

No error. The participants were presented with 18 targets in the correct target location.

Fixed error. The participants were presented with 18 targets that were offset by a fixed error. The fixed error is by a single Lego peg and can be in one of the nine possible locations illustrated in Fig. 2. All of the participants were exposed to a different ordering of the offsets, and each of the nine offsets was tested three times, with one (the perfectly aligned case shown in the middle of Fig. 2) only tested twice.

Random error. The participants were presented with 18 targets, each of which was offset by a different random error. We used a blocked random design, so that the participants were exposed to the nine different offsets shown in Fig. 2 in a random order and then again exposed to the same nine offsets in a different random order. Each participant was presented a different random ordering of the offsets. Again, for the purposes of this study, the magnitude of the error was set at 1 Lego peg.
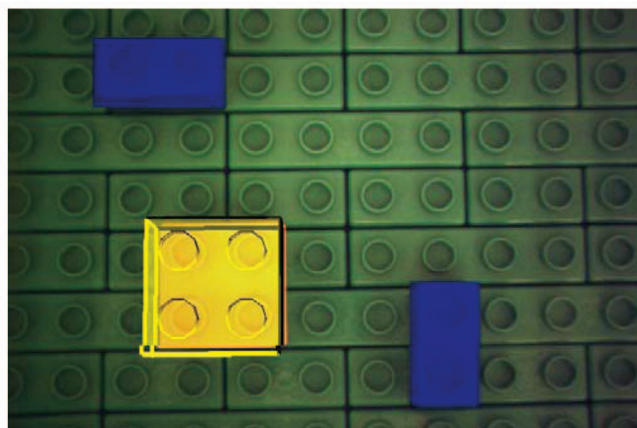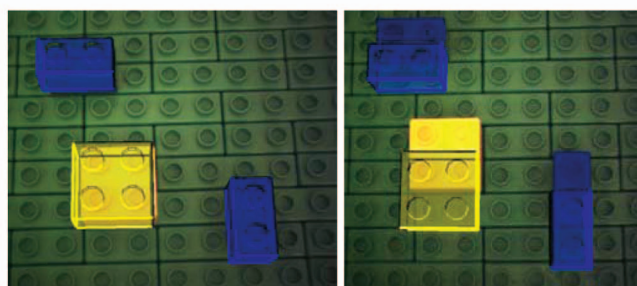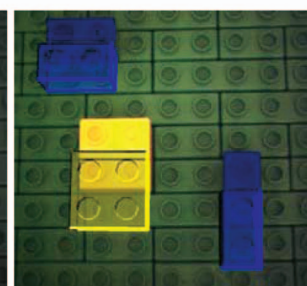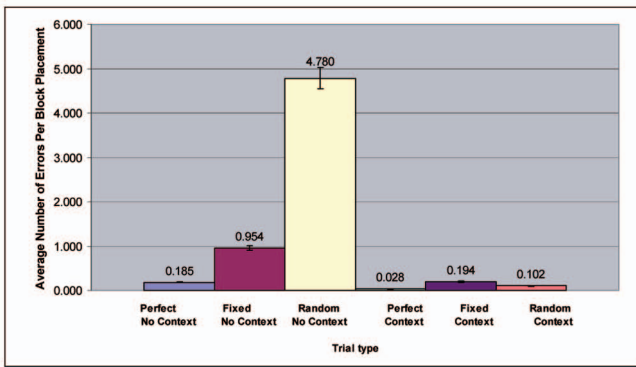
In half of the experimental trials the participants experienced, some virtual context was displayed on the head-mounted display. For the purposes of this study, the context took the form of two virtual blue Lego blocks that represented two physical blue Lego blocks that existed on the Lego base plate. Fig. 3 shows the context that was provided in some trials. Fig. 3a shows the physical blue context blocks. Fig. 3b shows both the physical blue blocks and the virtual blue context blocks that were provided on the head-mounted display when there was no registration error in the system. Fig. 3c shows both the physical blue blocks and the virtual blue context blocks that were provided on the head-mounted display; however, in this case, there is registration error in the system causing the virtual world and the physical world to be misaligned.
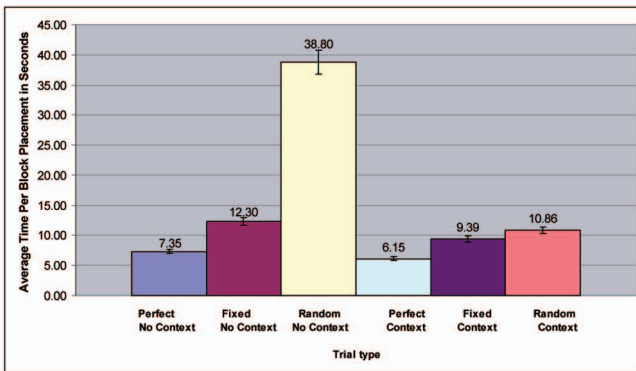
4.5   The Pilot Study

We ran six participants in our pilot study, ranging in age from 25 to 35 years old. The majority of these participants were members of our research group.

As previously mentioned, there were a total of six blocks of trials for this study: no error and no context, fixed error and no context, random error and no context, no error with context, fixed error with context, and random error with context. Half of the participants were presented with the no context trials first and the other half were presented with the context included trials first. In order to eliminate any order biasing within each of the context trials, a 3×3 Latin Square was used to determine the order of presentation of the three types of error.

(a)

(b)

Fig. 4. Pilot study: average errors and time per block placement.
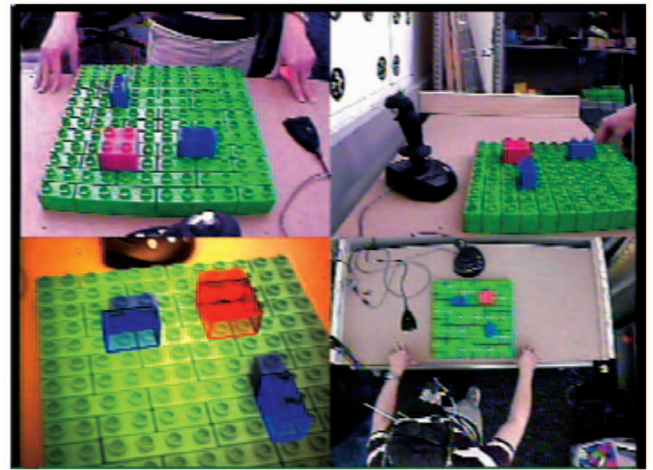

Fig. 5. The quad view of video data collected. (The block is red instead of yellow because red was used in the pilot study and we did not want to include an image of a real participant in this paper.)

4.6 Data Recorded

Several types of data were recorded during the experiment in addition to the questionnaires. Trial data including block data (color, size), how many times the participant attempted to place each block, and the time to complete each block placement was collected. We also recorded the tracker data for each trial in case the participants lost tracking and we needed to more closely evaluate the time it took for the participant to complete that particular trial. However, this ended up being unnecessary.

In addition, video data was collected including a view of what the participant was seeing, a view from above, and a frontal and side view of the Lego base plate to see the participant's hands and to see how and where they placed the block. Fig. 5 shows the quad view recordings of the various video data we collected.
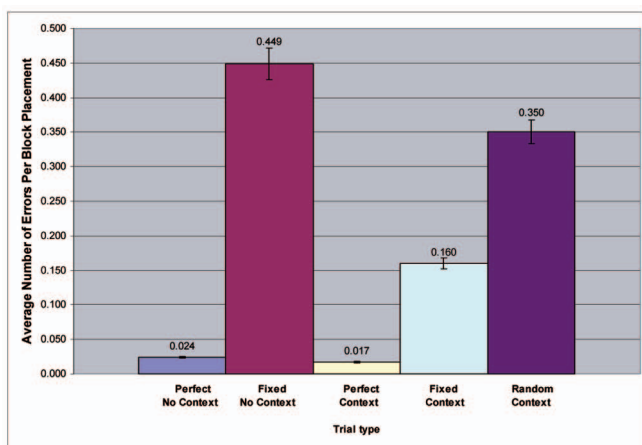
5 RESULTS

As previously stated, this was a within-subjects experiment, so each participant was asked to complete all of the blocks of trials. An alpha level of 0.05 was used for all statistical tests. In some of the graphs provided in this section, we have included the random no context case data that were evaluated in the pilot study, but discarded from the main study, to illustrate the vast difference between this case and the other five cases.
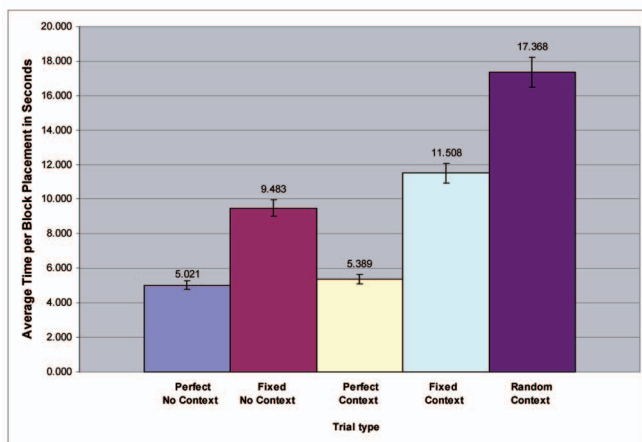
5.1 Descriptive Statistics

Fig. 6 illustrates the average number of errors per block placement and the average time per block placement in each of the five conditions in this study. Fig. 6a shows that the perfect registration cases produce the least amount of error, while the fixed error with no context case produces the largest amount of errors. It also shows that adding context to any of the three error cases reduces the amount of errors made, even in the case of perfect, or negligible, registration error.

Fig. 6b shows that the perfect registration cases have a better performance rate in terms of time per block placement and the random case has the longest time per block placement. Fig. 6 also shows that while the blocks of

However, after running our pilot study, we found that the participants were merely guessing as to where to place the blocks in the random error and no context case. We found that most of our participants adopted some sort of strategy for placing the blocks in this trial. Some tried each of the possible locations in a clockwise fashion and some in a counterclockwise fashion. Others followed the placements using rows and columns. Regardless of their method, we found that the use of these strategic approaches produced an average of 4.780 errors, as shown in Fig. 4a, for the nine possible block placement locations. This is close to what you would expect (4.5 errors) with random placement. More seriously, as shown in Fig. 4b, these trials took some participants a significant amount of time and were very tiring. Therefore, in the full study, we elected to dismiss the random error and no context case.

With the exception of the random error with no context case, all of the cases we ran in our pilot study seemed appropriate for the full study. And most of the data that we collected seemed to align with our hypotheses and had sufficient variability, despite the fact that we ran somewhat "expert" users in the pilot. Therefore, other than dismissing the random error with no context case, we did not change any of the other cases in this study.

By eliminating the random error and no context case, our full study was left with five blocks of trials. We used a 5×5 Latin square to determine the order of presentation of the remaining five cases.

(a)

(b)

Fig. 6. Full study: average errors and time per block placement.

trials with context produce a smaller amount of error as opposed to their no context counterparts, the amount of time per block placement for both perfect and fixed registration when context is provided actually increases. We attribute this occurrence to the fact that the participants have to take the time to mentally process the context before they can place the block, whereas in the no context cases, this mental process does not exist.

5.2 Correlations

While analyzing the data we collected, we discovered several interesting relationships, some of which we expected to find and some of which we decided to look at only after observing the participants complete the Lego block placement tasks.

5.2.1 Spatial Abilities

We found some significant relationships between spatial abilities and both time per block placement and number of errors made. We found significant negative correlations between spatial abilities and block placement times for the perfect no context $(r = -0.619, p = 0.001)$, fixed no context $(r = -0.579, p = 0.002)$, fixed with context $(r = -0.446, p = 0.22)$, and random with context cases $(r = -0.650, p = 0.000)$. These results imply that people with higher spatial abilities can complete these tasks more

quickly. It also suggests that high spatial abilities were not needed in order to complete a task in which no registration error was present but context was provided anyway. Basically, this implies that people with lower spatial abilities performed as well as people with high spatial abilities in this case.

However, there is only a significant relationship between number of errors made and spatial abilities for the fixed no context $(r = -0.577, p = 0.002)$ and random with context $(r = -0.657, p = 0.000)$ cases. These results suggest that people with low spatial abilities can perform just as well as people with high spatial abilities for the perfect no context, perfect with context, and fixed with context cases. We attribute the significance of the fixed no context and random with context cases to the observation that these cases seemed to be the most difficult cases that we studied. The fixed no context case required the participants to fully understand the concept of registration error because it did not provide any contextual clues and it required them to remember which locations on the Lego base plate that they had already tried to place the block in order to keep the number of errors they made to a minimum. This proved to be difficult for some participants, especially those that scored lower on the spatial ability test. The random with context case was equally as difficult because it required the participants to make a mental model of the relationship between the physical context block and the virtual context block, and reverse that model to place the physical yellow block in the correct location with respect to the virtual yellow block.

We found the most frequent mistake that participants made was not reversing the relationship between the physical and virtual worlds when trying to determine where to place the block. This reversal was very difficult and frustrating for many of the participants in this study, and as proof of that, when we looked at the workload data we collected, there were no significant correlations between spatial abilities and any of the NASA TLX categories except for perceived frustration in the random no context case. The lower the participant's score on the spatial ability test, the more frustrated they were with the random with context case. $(r = -0.468, p = 0.16)$ However a chi-square test showed that spatial abilities do not predict frustration $(chi-square = 2.746, p = 0.098, df = 1)$. These results could indicate that the sample size might not be large enough to make such a prediction, that there were range restrictions in the spatial ability scores evaluated, or that spatial ability cannot actually predict frustration levels in tasks such as those in this experiment.

Another interesting observation that we made was that a participant's profession did not have any bearing on their spatial abilities. The majority of our participants were in technology fields, and scored high on the spatial abilities test. However, two participants who had majored in business in college (and had related jobs) had two of the highest scores on the test.

5.2.2 Confidence

There is a negative correlation between the time per block placement and the average confidence in each of the cases. However, there is only a significant correlation in the no context cases (perfect no context $r = -0.493$, $p = 0.01$, fixed

TABLE 1
Correlations between Total Number of
Errors Made and Perceived Performance

Condition	Pearson Correlation	Sig. (2-tailed) * significant at 0.05 ** significant at 0.01
Perfect No Context	-0.075	0.715
Fixed No Context	0.532	0.005 **
Perfect with Context	0.624	0.001 **
Fixed with Context	0.548	0.004 **
Random with Context	0.704	0.000 **

TABLE 3
Correlations between Frustration Level and Mental Demand

Condition	Pearson Correlation	Sig. (2-tailed) * significant at 0.05 ** significant at 0.01
Perfect No Context	0.539	0.005 **
Fixed No Context	0.449	0.022 *
Perfect with Context	0.419	0.033 *
Fixed with Context	0.634	0.001 **
Random with Context	0.610	0.001 **

no context $r = -0.491$, $p = 0.11$). We expected to see confidence levels rise as the time per block placement decreased, especially in the no context cases where the participants had to rely on their instincts for proper placement. We believe that there was not a significant correlation between time per block placement and the average confidence in the context cases for two possible reasons. First, block placement times tended to be longer in the context cases because the participants had contextual clues to decipher and, second, because the participants were being more careful and checking and rechecking their block placements.

The only significant relationship between number of errors made and confidence was in the random error with context case ($r = -0.508$, $p = 0.008$). In general, this was the most frustrating case, as previously discussed; the more errors people made, the less confident they were.

5.2.3 Subjective Workload

We found many correlations when we looked at the workload data we collected using the six NASA TLX Likert scales as a guide for evaluating subjective workload. Tables 1, 2, and 3 show the correlations and significance values for each of the comparisons we made. Table 1 shows the correlations between total number of errors made and the perceived performance of the participants. The significant positive correlations imply that the more errors the participants made, the less successful they felt they were in accomplishing the goals set out for the task. The only case in which there was not a significant correlation was the perfect no context case, but people made so few errors that their perceived performance tended to always be good, with very little variation.

Table 2 shows the correlations between perceived performance and frustration levels. Again, these significant

relationships implied that the less successful people felt they were at completing the task, the more insecure, discouraged, irritated, stressed, and annoyed they became. This relationship held for all five of the conditions.

Table 3 shows the correlations between frustration levels and mental demand. They imply that the more frustrated the participants became, the more mentally demanding the tasks became.

The subjective workload correlations implied that the participants knew how they were performing in the tasks. And because they were aware of how poorly (or well) they were doing, they were able to take measures to either fix any mistakes they might have been making (if they were doing poorly) or ensure that they kept doing what they were doing (if they were doing well).

Despite the above correlations, mental demand and total number of errors made did not correlate. This makes sense because there were trials in which the participants exerted mental effort to ensure the correctness of their block placements and did not make errors as a result.

5.3 Planned Contrasts

A multivariate analysis using repeated measures was used to analyze the data. The within-subjects factors were the five conditions that the participants experienced, and the two dependent measures we were evaluating were number of errors made per block placement and time per block placement. The results of the tests of within-subjects contrasts for errors made and time per block placement can be found in Tables 4 and 5.

The simple contrasts for number of errors made showed some significant differences. In terms of errors, there was no significant difference in the perfect no context case and the perfect with context case. This implies that when there is no error in an AR system, adding context does not significantly reduce the number of errors. However, there is a significant difference in terms of errors between both of the fixed error cases. This implies that if there is registration error, adding context to an AR system significantly reduces the number of errors a person will make. The contrasts also showed no significant difference between the fixed no context case and the random with context case in terms of errors. Since these two cases were the most difficult, this result is not surprising.

The simple contrasts between the different conditions in terms of time per block placement produced different results. In terms of time per block placement, there were no significant differences between the context and no context cases for both the perfect registration and fixed error cases.

TABLE 2
Correlations between Perceived
Performance and Frustration Level

Condition	Pearson Correlation	Sig. (2-tailed) * significant at 0.05 ** significant at 0.01
Perfect No Context	0.708	0.000 **
Fixed No Context	0.630	0.001 **
Perfect with Context	0.398	0.044 *
Fixed with Context	0.470	0.016 *
Random with Context	0.780	0.000 **

TABLE 4
Simple Contrast Matrix for Number
of Errors Made per Block Placement

	Perfect No Context	Fixed No Context	Perfect with Context	Fixed with Context	Random with Context
Perfect No Context	1				
Fixed No Context	F = 30.661 Sig. = 0.000	1			
Perfect with Context	F = 0.088 Sig. = 0.770	F = 26.181 Sig. = 0.000	1		
Fixed with Context	F = 19.076 Sig. = 0.000	F = 10.777 Sig. = 0.003	F = 20.327 Sig. = 0.000	1	
Random with Context	F = 24.663 Sig. = 0.000	F = 1.476 Sig. = 0.236	F = 21.157 Sig. = 0.000	F = 5.640 Sig. = 0.026	1

This implied that adding context in each of these error conditions did not help users perform their tasks more quickly.

When looking at both of these contrasts, it is interesting to note that context does not help reduce errors or help quicken task completion when there is no error in the system. However, when there is fixed error in the system, context does help reduce the number of errors, but it takes relatively the same amount of time to perform the tasks. This phenomenon can be explained by the fact that despite the time savings in making less errors when there is context provided, there is a significant amount of time that is

TABLE 5
Simple Contrast Matrix for Time per Block Placement

	Perfect No Context	Fixed No Context	Perfect with Context	Fixed with Context	Random with Context
Perfect No Context	1				
Fixed No Context	F = 16.225 Sig. = 0.000	1			
Perfect with Context	F = 0.732 Sig. = 0.400	F = 11.225 Sig. = 0.003	1		
Fixed with Context	F = 24.297 Sig. = 0.00	F = 1.871 Sig. = 0.184	F = 18.724 Sig. = 0.000	1	
Random with Context	F = 24.955 Sig. = 0.000	F = 10.649 Sig. = 0.003	F = 22.140 Sig. = 0.000	F = 4.900 Sig. = 0.036	1
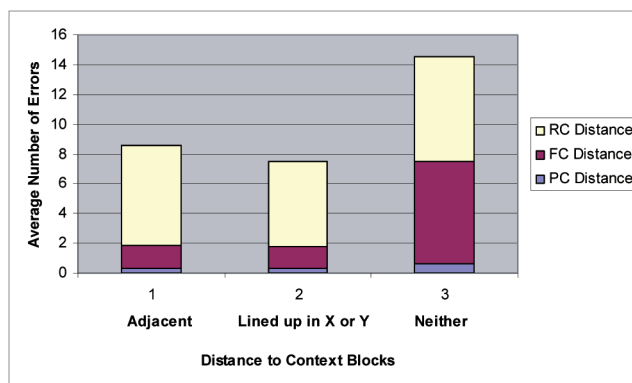
(a)
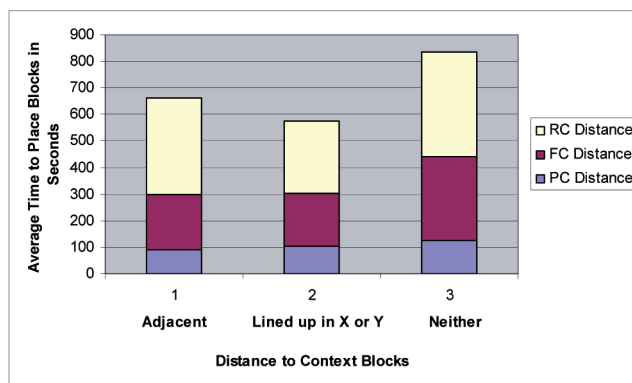
(b)

Fig. 7. Average number of errors and average time to place blocks as they relate to distance to context blocks.

devoted to understanding and using the visual context blocks provided.

We did not include the results from the pilot data in the above analysis, so we were not able to address the effect of context in alleviating random registration error. However, we did a similar analysis of our pilot data and found that context significantly reduces the number of errors made ($F = 28.803, p = 0.003$) and the block placement time ($F = 11.061, p = 0.021$) when there is random error in the system. The results were obvious when watching the participants struggle to complete the task by basically guessing where to place the block.

5.4 Distance to Context Blocks

As we were running participants through our study, we noticed that people seemed to have an easier time placing the yellow blocks when they were located adjacent to the blue context blocks. Because of this observation, we decided to see if there was any significance to this observation.

Additionally, we observed that when the yellow blocks were lined up with the blue context blocks either along the X direction or the Y direction, block placement was slightly easier. Therefore, we decided to divide the block placement tasks into three types: adjacent to the context, lined up in X or Y with the context, and neither, meaning the yellow blocks were being placed somewhere else on the board, but had no adjacency or linear alignment with the context blocks. Fig. 7 shows the number of errors and total block

placement times for each of the three distances from the context blocks mentioned above. Looking at the graphs, we can see that having the yellow target block location adjacent to or linearly aligned with the blue context blocks produced the fewest number of errors and the tasks were completed in the least amount of time. However, these results are not statistically significant, likely because of the small size of the data set (there were only 13 adjacent placements, 19 linearly aligned placements, and 22 nonaligned; the errors and time are the averages for that block across all participants). Also, we had expected to see the adjacent cases produce fewer errors and take less time than the linearly aligned cases, but that is not what is shown in the graphs; again, this is likely due to the small data set.

We performed several univariate analyses of variance (ANOVA) to compare the total errors made and the total time taken to the distance from the context blocks. We found that in the perfect with context cases, the total number of errors made did not differ significantly between the different distance cases. However, with respect to total time taken to place the blocks, the difference between the times did significantly differ. We found that both the adjacent and linearly aligned cases differed significantly from the neither case (adjacent $F = 48.312$, $p = 0.029$; linearly aligned $F = 35.066$, $p = 0.037$), but the adjacent case and the linearly aligned case were not significantly different.

In the fixed error cases, the results varied from the perfect cases. We found that the total times did not vary significantly between the different distances, but there were some significant differences with respect to total errors. We found that there was a significant difference in number of errors made between the adjacent distance case and the neither case ($F = 5.375, p = 0.049$). And the difference between the linearly aligned case and neither case was approaching significance, but was not quite significant ($F = 5.375, p = 0.078$). Again, there was no significant difference between the adjacent and linearly aligned cases.

In the random with context cases, we found no significant differences in the context block locations. We expected to see more significant differences between the different types of distances in the random case as well as in the perfect and fixed error cases than we actually saw. Again, we think that the sample sizes for each of the distance types was too small to produce more significant results.

6 DISCUSSION

In this section, we discuss the relationship between our findings and our previously stated hypotheses. Again, in the graphs provided in this section, we have included the random no context case data that were evaluated in the pilot study, but discarded from the main study, to illustrate the vast difference between this case and the other five cases as well as to discuss how that case pertains to some of our hypotheses. The random data also provide the worst-case baseline to contrast with the perfect case baseline.
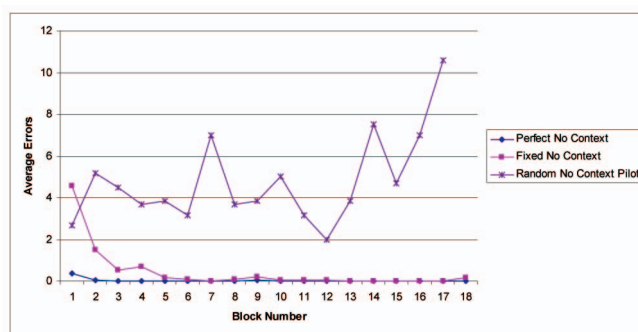
(a)
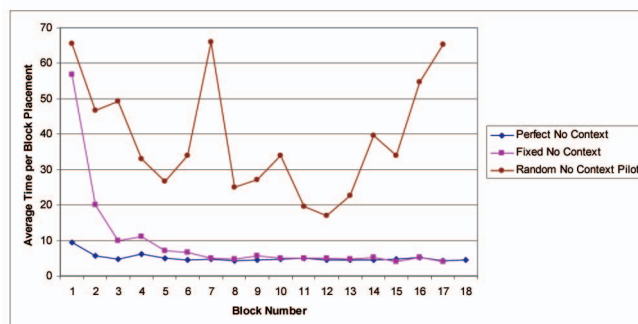
(b)

Fig. 8. Average number of errors and time per block placement when no context is given. (Note: The random no context data are from the pilot study and are included only for reference.)

6.1 Effect of Type of Registration Error on Average Time and Average Number of Errors

When exposed to different types of error in a scene in which no context is provided, there will be a different user response for each type of error. We predicted that when exposed to a fixed error, users will gradually learn how to compensate for the error; however, when exposed to random error, the task will become a guessing game as to where the block should actually be placed.

Fig. 8 illustrates the number of errors made per block placement and the average time per block placement for each of the three conditions in this study when no context was provided. In this figure, we have included the data for the random no context case included in the pilot study but not in the full study to show the enormous difference between this case and the others. Fig. 8a shows that when the graphics and the real world are perfectly aligned, there were almost no errors made by the participants. Two participants did make a few errors in the perfect case, but these errors can be accounted for. Participant 18 was trying to complete the perfect task very quickly and got careless throughout that block of trials. She made two errors because she did not really look around the Lego base plate to confirm the block was in the correct location. The angle between her and the base plate was not ideal for correctly placing the block. Despite being instructed during the training phase of the study to look around, this participant stood still and tried to place the blocks quickly. After she made those two mistakes, she slowed down and looked around more and made no additional mistakes.

In contrast, Participant 14 made nine errors on his first block placement because he had completed only error cases before completing the perfect case, so he did not trust the graphics. He did not even try the correct location that the system was showing him until he tried all of the other eight possible locations. Once he realized that the graphics aligned perfectly, he made no further mistakes.

Fig. 8a also illustrated some interesting results for both the fixed error and random error cases. When there was fixed error in the system, the average number of errors started off high and gradually approached zero errors as the participants learned how to deal with this type of registration error. In the random error pilot case, users never came up with a strategy to deal with this type of error; they merely guessed until they found the correct location. This supports our first hypothesis.

Fig. 8b shows similar results, but in terms of time per block placement rather than number of errors per block placement. Again, in this case, the time per block placement for the fixed error case approached the time for the perfect registration case as the task progressed and the time per block placement in the random registration error pilot case varied drastically.

6.2   Effect of Context on Average Time and Average Number of Errors

If context has the effect of providing useful relationship information, placement tasks should be able to be performed with fewer errors; however, the cognitive processing of the context information might increase trial times.

Therefore, we predicted that when context is added to a scene, whether fixed or random error is present, the time per trial will increase, but the total amount of errors will decrease. Fig. 9 illustrates the number of errors made per block placement and the average time per block placement for each of the three conditions in this study when context was provided. When comparing Figs. 8a and 9a, the difference in the number of errors in the fixed and random cases is very apparent. In the fixed error case, context seemed to quicken the learning curve involved in figuring out how to adapt to the registration error, thereby reducing the number of errors made by 35.7 percent. However, when comparing Figs. 8b and 9b, there is a noticeable increase in the average time taken to place the block, even though there is not such a drastic learning curve involved in figuring out where to place the block in the context case. In fact, the block placement times increased by 21.4 percent. We attribute this increase to the increased cognitive load required to comprehend the context provided.

In the case of random error in the pilot study, the context was so helpful in completing the task successfully, that the context improved the average number of errors per block placement by 92.6 percent but decreased the average time per block placement by 55.2 percent. In short, without the context, participants could not successfully complete the placement task when there is random error; they merely guessed where to place the block until they guessed correctly. We did not expect to see such a drastic improvement in time per block placement for the random
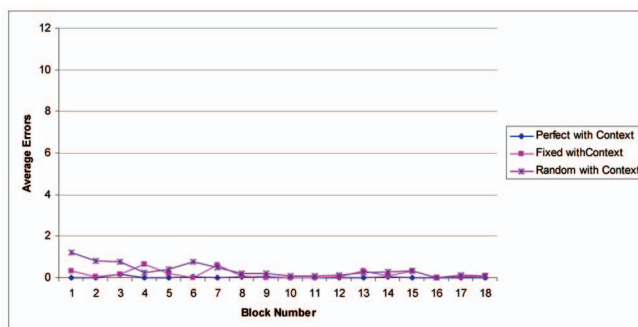
(a)
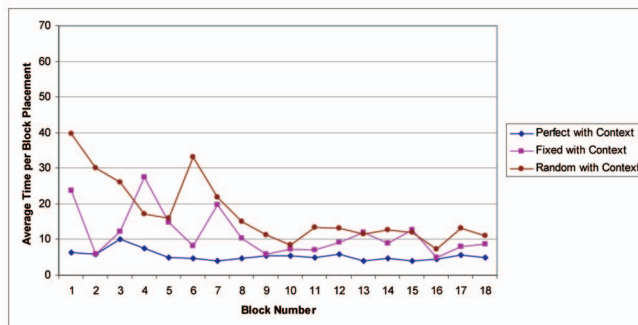
(b)

Fig. 9. Average number of errors and time per block placement when context is provided.
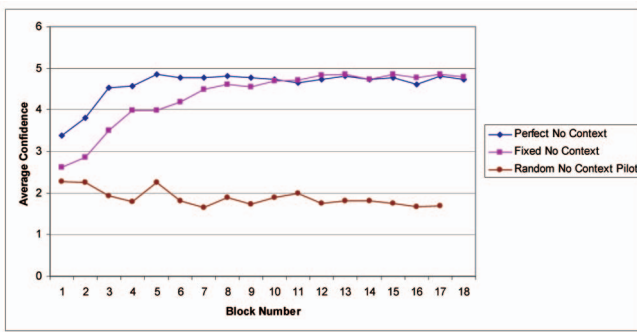
case, but quickly realized that portion of our hypothesis was incorrect.

Therefore, we found that our hypothesis with regards to number of errors made and block placement times holds true for fixed error, but not for random error. Only our hypothesis that the number of errors made would reduce holds true for the random case; the hypothesis concerning time per block placement does not.
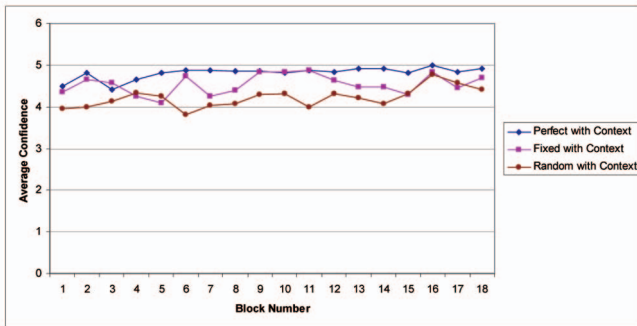
6.3   Effect of Context on Perfect Registration

When there is no error in an augmented system, the user does not need additional context to perform a placement task correctly. Therefore, we predicted that when there is no registration error associated with the system, adding context neither increases nor decreases the number of errors. We did not anticipate any errors being made in the perfect cases, but we neglected to factor in participants rushing through the tasks and being careless. We also did not anticipate anyone not trying the target location shown because they had only been exposed to error cases up until that point. Fig. 6a shows that there were a few errors made in both of the perfect error cases and there was a slight difference in the number of errors made between the no context and context cases. In fact, there were fewer errors made in the perfect with context case.

However, despite the slight, yet visible, differences between the context and no context cases when registration was perfect, there was no significant difference between the number of errors made ($F = 0.088, p = 0.770$) and there was no significant difference between the time per block placement ($F = 0.732, p = 0.400$). Therefore, this implies

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ROBERTSON ET AL.: AN EVALUATION OF GRAPHICAL CONTEXT AS A MEANS FOR AMELIORATING THE EFFECTS OF REGISTRATION...   11

(a)

(b)

Fig. 10. Average confidence per block. (Note: The random no context data are from the pilot study and are included only for reference.)

that when there is perfect registration, context does not give us any added benefit, nor does it hinder performance.

6.4 Effect of Context on Confidence

Although we did not have any hypotheses related to confidence, we did make a few interesting observations with regard to confidence. Fig. 10a shows the average confidence per block for the no context cases, including the random no context case only tested in the pilot study. It shows that the participants gradually became more confident as the placement task progressed for both the perfect and fixed error cases. This shows that when there is no context present, participants can eventually figure out how to complete the task successfully, thus boosting their confidence. This figure also shows that in the pilot study, confidence levels decreased as the task progressed for the random case. The decrease implies that as people began to realize that they just had to guess to try to find the correct location for the block, they became less confident.

Fig. 10b shows the average confidence per block for the context cases. In general, the average confidence for all three cases started higher than the no context cases. In the case of perfect registration, the confidence levels were higher on average for the context case (average confidence = 4.814) than in the no context case (average confidence = 4.603), thus showing that even when context does not significantly help reduce errors or block placement time, it does help people feel more confident about their performance. In the fixed error no context case, there was a bit more fluctuation in the confidence, although the average confidence in the context case (average confidence = 4.542) was still higher than in the

no context case (average confidence = 4.324). We attribute this fluctuation to an observation we made while conducting this study. We noticed that in many cases, despite the fact that the error was in the same direction and magnitude for each of the block placements, a large number of the participants did not notice that occurrence and continued to try to figure out the relationship between the real world and the virtual world each time there was a new block. This caused their confidence to fluctuate much as it did in the random error with context case.

While it is important to note that the confidence levels did fluctuate in the random error with context case, the confidence levels on average for the context case (average confidence = 4.212) were a great deal higher than in the no context cases conducted during the pilot study (average confidence = 1.880). Basically, this implies that adding context turned a task that people felt they were constantly failing at into a task that they felt they were quite successful in completing.

Therefore, on average, we found that context helped people feel more confident when completing these tasks.

7 OBSERVATIONS

While collecting data in this study, we noticed some very interesting participants' comments and behaviors. We heard many comments, such as "Why are they moving?" in the random error case when people had not quite figured out which case they were in. We also heard comments like, "Oh, I get it now" or "I see" or "I've got the pattern now" when the particular trial finally made sense to them. We even heard comments like "I don't trust you" when they were in the error cases. We also noticed many differences in the strategies the participants used for placing the blocks as well as other higher level observations.

7.1 Block Placement Strategies

As we previously mentioned, participants used many different strategies for placing the physical blocks. In general, some participants were extremely confident in the placements and immediately placed the physical block in the correct location, but usually they developed their own strategy to use.

When no context was present, the strategies were more like guessing games. Some participants immediately tried the shown target location no matter what, while others who had been through some error trials before did not trust where the system was telling them to place the block and would try all of the other eight locations before trying the shown target location. As previously mentioned, we noticed that people tended to search for the correct location around the virtual target by either going through the nine possible locations in a clockwise or counterclockwise fashion, while a few others tried their placements using rows or columns. A few less successful participants randomly tried the possible locations, but quickly forgot where they had previously tried and ended up repeatedly trying the same locations.

The placement strategies also differed if there was context present. When the virtual blue context blocks were present, some participants used the information they

provided for every single block placement no matter what type of error was present in the system, usually costing the participants more time. However, others quickly learned the correct offset (if there was an offset) for the no error and fixed error cases and completely ignored the context blocks after they learned the offset. This learned behavior actually saved those participants time in placing their blocks. In the random error case, the participants were forced to use the context for every block placement. In addition, when there was context present and it was located next to or just a peg or two away from the target location, participants tended to more quickly place the blocks without having to think too much about the correct displacement. They could just see where they needed to place it. However, the larger the distance between the context blocks and the target location, the more the above strategies were used.

The participants differed on how they used the context information. Some would mentally do the mapping between the virtual world and the physical world. Others used some sort of hands-on strategy. Some participants would place the physical block in the location that the virtual block was shown and then move the block once they figured out the proper translation. Some would count the pegs in both directions from the virtual context blocks to the physical context blocks and count that same number of pegs from the virtual target location to decide where to place the physical target block.

7.2 Common Mistakes

The most common mistakes people made were not related to the system at all, but were related to inexperience with an AR system in general. It sometimes took quite a while for people to feel comfortable enough with the system to move around and look at the Lego base plate from different perspectives. Without these different perspectives, it was sometimes hard to see the proper location for the block, especially for people who were not as tall as others. It was very easy for them to mistake one row for the next.

Another common mistake the participants made was not trusting the system. Due to the Latin Squares design, some people received all of the error cases first before any of the no error cases. Once they got to the no error cases, many of these participants did not even try the location that the virtual block was showing them because they did not trust the information they were being given. These types of human error mistakes accounted for quite a few of the errors that were recorded.

The biggest conceptual mistake that we noticed people making was reverse mapping the relationship between the virtual context and the real world. If the scene showed that the blue virtual context block was up and to the right of the physical blue context block, numerous participants would attempt to place their physical yellow block down and to the left of where the yellow virtual target was pointing them to place the block. It sometimes took several trials before the participants realized their mistakes, but once they did make the correct association, they very rarely made that same mistake again.
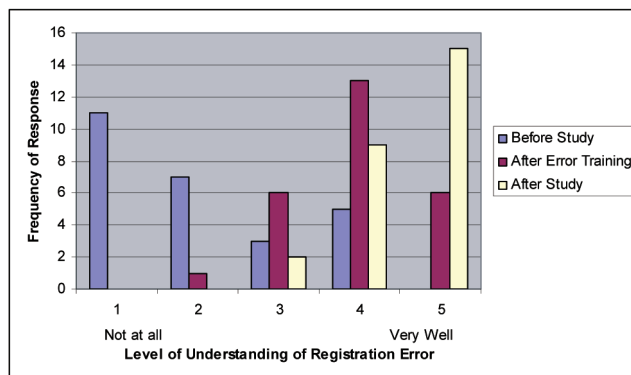

Fig. 11. Frequencies of participant understanding of the concept of registration error in AR.

7.3 Participant Understanding of Registration Error

We asked the participants how well they understood the concept of registration error when they were filling out the introductory questionnaire before they participated in the study, after they read the error training document, and after they participated in the study when they were filling out their final study questionnaire. A five-point Likert scale was used, with 1 meaning they did not understand the concept of registration error at all and 5 meaning that they understood the concept very well. We wondered how well our error training document had prepared them for their participation in the study as well as how much the hands-on experience taught them about registration error.

As seen in Fig. 11, the majority of the participants did not understand the concept of registration error before they took part in this study. We found that 18 out of 26 people reported either a 1 or 2 out of 5 (69 percent) in response to our survey. After reading the error training document, the results varied drastically. We found that 19 out of 26 people (73 percent) reported either a 4 or a 5 for their level of understanding. This showed us that the error training document we provided seemed to illustrate the concept of registration error well enough for the participants to successfully complete the tasks. And after completing all of the tasks, the number of participants reporting an understanding level of either 4 or 5 jumped to 24 out of 26 people (92 percent).

While these results do not have any statistical value, it was interesting to see how people's perceived understanding of registration error increased during their participation in our study.

7.4 General Opinions

The participants varied in their opinions about the study. Some of the participants loved having the context blocks present, no matter which type of error was present, because the context gave them more confidence in their block placements. Other participants only found the context blocks necessary in the random error case because they felt like they could complete the tasks quicker without feeling the need to interpret any context that was present.

In the general comments section of our questionnaire, we received several comments about the shape of the context blocks that we had not anticipated. Some participants

thought that they could have mapped the difference between the virtual blue context blocks and the virtual yellow target block better if they had been the same shape, while other participants liked the fact that they were a different shape than the physical block they were trying to place. In fact, a couple of participants commented on how helpful the orientation of the 1×2 peg Lego blocks were for context because we aligned one of the context blocks to have its longer side on the vertical plane and one on the horizontal plane, as shown in Fig. 3. They felt that this planar information gave them an additional type of context to use.

A few participants mentioned that they found it harder to place the blocks when their target locations were at the back of the Lego base plate, farther away from them. This is probably due to the fact that the participants had to bend over more to get a good overhead view of the target location. One participant suggested tilting the board on an angle to eliminate this difference between the front and back of the base plate.

We asked the participants if they had any ideas about different types of context to add to this system, and we received a variety of responses. As mentioned above, some wanted the context blocks to be the same size and shape as the physical target block. Others wanted the entire base plate to be drawn. Some wanted a graphical indication of where they had already tried to place the block previously, for instance, a grayed out virtual representation of the block in all of the locations that had been tried. Some suggested different colors of context for the different directions of error. Some suggested that we draw arrows to show the directions in which there was error, while others wanted both arrows and a number to indicate how many pegs in that direction that the block was off. Some wanted audio feedback, like a beep when you got close and a distinct bell when you were over the correct location. While some of these ideas are interesting and could be implemented in the future, most are not feasible in a real-life AR system because they rely on knowing exactly what the error is: if a system knew the error, it could just be corrected for.

8 CONCLUSIONS

This study has shown that adding context to a scene that is ambiguous because of registration error can help a user make sense of the ambiguity. We have shown that context is not really needed when there is perfect registration in an augmented environment, but it does help people feel more confident. We have also shown that context cannot only help reduce the number of errors that people make in a Lego block placement task when registration error is present, but it can also help to reduce the time it takes a user to complete the task when random registration error is present. In addition, in the case of random registration error, context can actually make a completely impossible task doable by almost anyone.

9 FUTURE WORK

There are several possible avenues for future work in this project, including some possible additions to this study to broaden the results as well as some additional studies to evaluate other types of context.

9.1 Broadening This Study

In this study, we recruited most of our participants from the undergraduate and graduate population at Georgia Tech. This meant that the majority of our participants were in technology-related fields and tended to be similar in age, spatial abilities, etc. We also had a huge majority of our participants that were right-handed (20 out of 26). It would be interesting to carry on with this study, and try to recruit a larger variety of participants. We would like to get a better range for spatial abilities, age, and handedness to see if any of these traits significantly affect a participant's ability to successfully complete these tasks. We were able to see some correlations in successful task completion in relation to spatial ability, but we could not see any causation. We believe that part of this is due to the skewed nature of our spatial ability data. This holds true as well for our handedness and age data. (Also, what other factors play into successful performance in these tasks?)

In addition, in this study, we chose a restricted class of errors (planar translations) so that we could study several specific cases and draw meaningful conclusions. It would be useful to evaluate more realistic forms of error such as translation error off the plane of the base plate, rotational error, three-dimensional error, and various kinds of jitter. These errors are more indicative of real-world errors and would be interesting to study. However, since these errors will likely move the augmentation away from the surface of the base plate, we may need to use a stereo display rather than a biocular display to avoid confounds arising due to the visual anomalies when the graphics are not aligned with physical objects in the world.

This brings up the question of whether a stereo display would change the results of our current study. Can people understand and adapt to registration error better, even in the simple situation of this study, when using a stereo display?

9.2 Evaluating Types and Amounts of Context

It would also be interesting to evaluate different amounts and types of graphical context to see what the best combination of context is for successful task completion. We could study the use of color and size of blocks to see if they play a part in success. It would be interesting to look at varied ways of displaying peg information on the base plate. For instance, we could draw only the pegs between the nearest virtual context block and the virtual target block to see if that extra information is helpful or just intrusive in the viewing plane. We could compare a fully rendered virtual representation of the base plate and blocks to just wire frame representations, and so on.

We noticed in this study that the relationship between the context block and the target augmentation was an important contextual cue. Would it help to draw lines between the context and the augmentation to gain the advantage of being "lined-up" with the context? We have noticed that as we study context, more questions emerge and more possibilities for future studies arise. The drawback to studies like these is that there is a limit to the

generality of the information gathered. Are the findings general, or would they be specific to the particular application the study was designed for? We believe to some degree the data evaluated would be application specific, but we also believe our results will be helpful in guiding the design of augmented environments.

9.3   Evaluating Nonregistered Context

We plan to run another study to evaluate context in less traditional forms of AR. Many researchers have proposed the use of AR for repair tasks. For example, Honda evaluated the Nomad Expert Technician System, a hands-free wearable display that provided access to vehicle history and repair information. But a common concern with HMDs is the risk of blocking important parts of the real world with the augmented graphics. We showed in this study that context is useful in registered AR, but we also want to prove the benefit of context in nonregistered AR as well as on a heads-up display (HUD). In particular, we want to try to more fully investigate the question "When provided enough context, is registration necessary?"

In order to test this question, we plan to compare four cases:

1. registered AR,
2. nonregistered AR with fixed position and orientation with respect to the Lego base plate,
3. HUD with variable orientation with respect to the user where the graphics remain in the user's field of view at all times [9, Condition 3], and
4. HUD with variable orientation with respect to the user where the user has to look to the side to see the graphics.

These cases are not only designed to evaluate the usefulness of registered verses nonregistered graphics, but we also want to see if orientation information is a good enough contextual clue when spatial registration is not possible. By this, we mean if an augmentation is not superimposed directly over the physical world, is displaying the graphics relative to the user's viewpoint on a HUD still useful?

ACKNOWLEDGMENTS

REFERENCES

[1] E.M. Coelho, B. MacIntyre, and S. Julier, "(osgAR): A Scenegraph with Uncertain Transformations," *Proc. Third Int'l Symp. Mixed and Augmented Reality (ISMAR '04),* pp. 6-15, Nov. 2004.

[2] S.E. Embretson, "The Factorial Validity of a Cognitively Designed Test: The Spatial Learning Ability Test," *Educational and Psychological Measurement,* vol. 57, pp. 99-107, 1997.

[3] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Human Mental Workload,* P. Hancock and N. Meshkati, eds., pp. 139-183, Elsevier, 1988.

[4] M.A. Livingston et al., "Resolving Multiple Occluded Layers in Augmented Reality," *Proc. Second Int'l Symp. Mixed and Augmented Reality (ISMAR '03),* pp. 56-65, Oct. 2003.

[5] R. Oldfield, "The Assessment and Analysis of Handedness: The Edinburg Inventory," *Neuropsychologia,* vol. 9, pp. 97-113, 1971.

[6] *Open Scene Graph,* http://www.openscenegraph.org, 2008.

[7] C. Robertson and B. MacIntyre, "Adapting to Registration Error in an Intent-Based Augmentation System," *Virtual and Augmented Reality Applications in Manufacturing,* A.Y.C. Nee and S.K. Ong, eds., pp. 143-163, Springer Verlag, 2003.

[8] D. Seligmann and S. Feiner, "Automated Generation of Intent-Based 3D Illustrations," *Computer Graphics, Proc. ACM SIGGRAPH,* vol. 25, no. 4, pp. 123-132, July 1991.

[9] A. Tang, C. Owen, F. Biocca, and W. Mou, "Comparative Effectiveness of Augmented Reality in Object Assembly," *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '03),* pp. 73-80, Apr. 2003.

Cindy M. Robertson received the PhD degree in computer science from the Georgia Institute of Technology. She is with the School of Interactive Computing, Georgia Institute of Technology, Atlanta. Her research interests include augmented reality, visualization, and human-computer interaction. She is interested in creating meaningful augmented reality applications in the face of changing registration error.

Blair MacIntyre received the BMath and MMath degrees from the University of Waterloo and the PhD degree from Columbia University in 1999. He is an associate professor in the School of Interactive Computing, College of Computing, Georgia Institute of Technology, Atlanta. His research interests include computer graphics and human-computer interaction with a focus on the design and implementation of augmented reality systems.

Bruce N. Walker is an associate professor in the School of Psychology and the School of Interactive Computing, Georgia Institute of Technology, Atlanta. His research in Sonification Lab focuses primarily on auditory displays and sonification within the context of the larger fields of human factors, engineering psychology, and human-computer interaction.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.