The Great Potato Search: The Effects of Visual Context on Users' Feature Search and Recognition Abilities in an IVR Scene Interactive Poster

Cullen D. Jackson David B. Karelitz Sean A. Cannella David H. Laidlaw
Department of Computer Science, Brown University, Providence, RI 02912
{cj, dbk, scannell, dhl}@cs.brown.edu

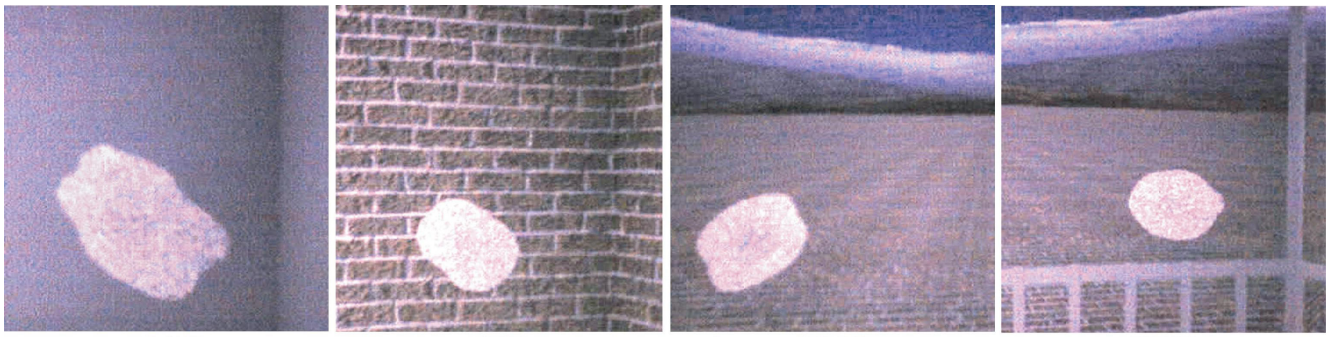

Figure 1. The four visual contexts, BLANK, BRICK, WORLD, and PORCH, displayed in our study of user search performance. The white objects in each context are examples of those displayed during the search task.

Abstract

We present a user study of the effects of visual context and interaction methods on visual search performance in an immersive virtual reality (IVR) system. Contrary to previous research on 2D visual search, our results suggest that visual context does not significantly contribute to a user's ability to find and identify features on an object in a scene. However, we did find that using a hand-held device to manipulate the object during search facilitated user performance more than simply walking around a static object.

1. Problem

Previous research suggests that visual context facilitates participants' ability to find and recognize objects in real-world scenes [1], [2]. These "real-world" studies suggest that humans use visual context to assess object congruence with the background, at least when the scene is briefly shown.

The current study aims to examine the role of visual context on a visual search task in an immersive virtual reality (IVR) environment. IVR allows users to fully interact in an actual 3D scene, replete with color, texture, and stereoscopic depth information. We can carefully control the parameters that describe each context, with the aim of understanding those variables that best optimize user performance. A secondary goal is to determine if there is a relationship between the visual context and the method of user interaction (walking around or using a hand-held device).

2. Hypotheses

We hypothesize that the previous results from the 2D visual search studies will generalize to a 3D setting, and that user performance will vary across visual context. Specifically, contexts which provide good motion parallax information and that are projected coincident to the physical CAVE walls will facilitate user performance. Furthermore, we hypothesize that users will be more accurate in finding and identifying features while using a hand-held device to manipulate the object (no locomotion) than while walking around the object; users are able to examine the objects at more orientations and viewpoints with the hand-held device than when the objects are in a fixed position in the environment, and do so more thoroughly.

3. Methods

Fourteen people ($mean_{age}$ = 24.14 ± 3.37 yrs.) participated in the study. The study took place in a CAVE [3] environment with three walls and a floor. The study was split into two 60-trial sets, one for each of the interaction types. Each set contained 12 practice and 48 test trials varied according to the context and the feature presented. In the BLANK context, all of the walls were the same shade of gray. A brick texture was drawn on the walls in the BRICK context. The WORLD context was composed of a brick floor (same as used in the BRICK context) extending to a horizon with mountains and trees and a blue sky above. The PORCH context was identical to the WORLD context with a white porch in the foreground, coincident with the physical walls of the CAVE (see Figure 1).

Users were instructed to find and identify a randomly placed feature (square or triangle) on the object. Each object had a noisy surface and floated in the middle of the CAVE in each trial. The features extended off the surface of each object at a fixed height while the heights of the noise components were random with a fixed maximum height. The shape of each object was determined by a series of spherical harmonics such that the objects were of similar 'lumpiness,' but not identical, between trials. Users received feedback during practice trials only. The two interaction methods used for the search task were walking around the object and using a hand-held device to manipulate the object while standing still.

4. Results and Discussion

Our hypotheses suggest that the BRICK context should facilitate user search performance more than the other contexts, followed by the PORCH and WORLD contexts respectively. The BLANK context, with a lack of any texture or depth information (sans the physical CAVE walls) should not facilitate user performance at all.

Results are reported for 9 of the 14 participants. The data of five participants were excluded because they timed out of more than eight of the 48 total test trials on at least one of the test sets (average number of timed-out trials was 12.5 ± 3.6).

No relationship was found between the main factors of the experiment (context, interaction method, and feature type) and there was no effect of context on user performance (shown by a four-way 4x2x2x2 ANOVA). However, there was a significant increase in recognition performance when using the wand to manipulate the object instead of walking around the object [$F(1, 7) = 6.37$, $MS_e = 250.39$, $p < .05$]. The relationship between the three within-subject factors is shown in Figure 2. In this figure, note the difference in performance across the two plots, each of which represents an interaction type. The mean accuracy for the two interaction types are shown as dashed lines in Figure 2 ($mean_{walk} = 79.1\%$, $mean_{wand} = 86.4\%$).

These results are contrary to our initial hypotheses concerning the visual contexts; no differences were found between the contexts. One reason may be that the dependent measure was not sensitive enough to capture subtleties in user performance. Another reason may be that users were focused only on the object. We might find that context influences performance in a more exocentric task that requires users to investigate more than one area of the virtual environment.

Assuming a sensitive enough measure, context *did not matter* in terms of this visual search task. This result is contrary to 2D studies showing that context aids object search and recognition. However, those tasks were performed with 2D images and over brief presentation times. The current study utilized 3D scenes, over much longer presentation times, with motion.

The results of this study suggest several things. First, motion cues provide robust configural information about the scene, as well as providing strong depth cues. Stereo also helps disambiguate information in the scene, particularly the structure of the feature with respect to the object. While we did not explicitly test the effects of motion cues or stereo viewing on the task (artifacts of head-tracked stereo IVR), previous research suggests that users make fewer errors comprehending visual data while using hand-coupled or head-coupled motion with stereo viewing [4]. Also, long presentation times permit the user to integrate a large amount of the available visual information in the scene. We believe that these strong cues, lacking in previous 2D visual search studies, likely contributed to the absence of differences between the four visual contexts in this study.

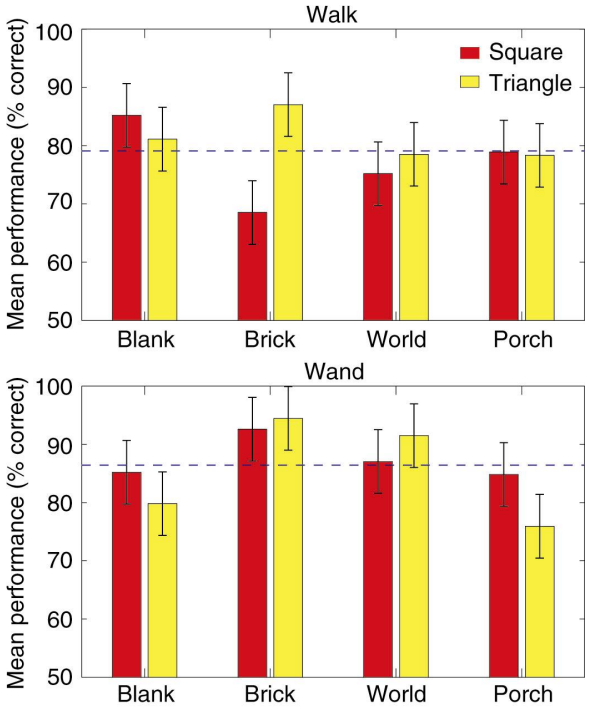

Figure 2. Influence of context, feature type, and interaction method on task performance. Context did not affect performance, while interaction type did. Dashed lines show mean performance over interaction type ($mean_{walk}$ = 79.1%, $mean_{wand}$ = 86.4%). The error bars are ± 1 SEM.

5. Acknowledgements

We wish to thank Morriah Horani for her helpful comments. This work was partially supported by a NSF ITR grant (CCR-0086065) and the Human Brain Project (NIDA and NIMH).

References

[1] Biederman, I., Glass, A. L., and Stacy, E. W. Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*(1), 1973, pp. 22-27.

[2] Boyce, S. J., Pollatsek, A., and Rayner, K. Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 1989, pp. 556-566.

[3] Cruz-Neira, C., Sandin, D. J., and DeFanti, T. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In the *Proceedings of ACM SIGGRAPH '93*, 1993, pp. 135-142.

[4] Ware, C. and Franck, G. Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics, 15*(2), 1996, pp. 121-139.