CAVE and Fishtank Virtual-Reality Displays: A Qualitative and Quantitative Comparison

Çağatay Demiralp, Cullen D. Jackson, *Member*, *IEEE*, David B. Karelitz, Song Zhang, and David H. Laidlaw

Abstract—We present the results from a qualitative and quantitative user study comparing fishtank virtual-reality (VR) and CAVE displays. The results of the qualitative study show that users preferred the fishtank VR display to the CAVE system for our scientific visualization application because of perceived higher resolution, brightness and crispness of imagery, and comfort of use. The results of the quantitative study show that users performed an abstract visual search task significantly more quickly and more accurately on the fishtank VR display system than in the CAVE. The same study also showed that visual context had no significant effect on task performance for either of the platforms. We suggest that fishtank VR displays are more effective than CAVEs for applications in which the task occurs outside the user's reference frame, the user views and manipulates the virtual world from the outside in, and the size of the virtual object that the user interacts with is smaller than the user's body and fits into the fishtank VR display. The results of both studies support this proposition.

Index Terms—User study, virtual reality, display, CAVE, fishtank VR, DT-MRI visualization.

1 Introduction

The scientific visualization community increasingly uses virtual reality display systems that employ new visualization and interaction techniques. Determining the relative merits of different VR display systems for different applications and tasks is thus important for developing effective visualization tools as well as new displays.

The purpose of the work reported here is to understand better how CAVE and fishtank VR display systems affect user performance. A CAVE is an immersive VR display system and fishtank VR display is a desktop display system [1], [2]. Both systems generate and update a stereoscopic view of a virtual world according to the user's head position and orientation. Interaction devices for CAVEs are mainly hand-held six-DOF tracked tools. For fishtank VR displays, possible interaction devices range from the standard mouse and keyboard to hand-held six-DOF tracked tools. The CAVE and the fishtank VR display used in this study are representative of current immersive and desktop display technologies. As a CRT-based desktop display system, the fishtank VR has higher angular resolution and brighter imagery than the projector-based, immersive display in the CAVE. On the other hand, the four walls of the CAVE provide more pixels than the fishtank VR display. As an initial attempt to compare user performance in these two environments, we conducted two consecutive user studies, one anecdotal (or qualitative) and one quantitative.

We performed the anecdotal study before the quantitative study for two specific reasons. First, we believe application-oriented user studies—using domain experts as subjects and their hypothesis-testing process as the task—can complement user studies involving abstract tasks and applications. Second, we wanted to gain insights for designing our quantitative study.

In the quantitative study, we measured users' time and accuracy in completing an abstract visual search task. The task was to find a feature on the noisy surface of a potato-like object shown in four different visual contexts. We developed two hypotheses prior to the study:

1. User performance would be better in the fishtank VR system than in the CAVE. Fishtank VR systems have such desirable characteristics as high angular resolution and bright, crisp imagery. Also, the fishtank VR system would be more effective for the type of task our study examined, since 1) the task occurred outside the user's reference frame, and the user viewed and manipulated the virtual world from the outside in, and 2) the virtual object that the user interacted with was smaller than the user's body and fit into the fishtank VR display. We give the details of the task in Section 3.2.
2. User performance would vary for different visual contexts. This hypothesis generalizes to a 3D setting previous results from 2D visual search studies. Previous 2D visual search studies found that presence and configuration of other objects in a scene influenced how quickly the target objects were detected or identified [3], [4]; the results of these studies suggest that humans use visual context to assess object congruence with the background and that visual context facilitates target detection or identification. In the case of 3D scenes, we believe that contexts that provide good motion-parallax information (better edge and figure-ground delineation) and are coincident with the physical display surface (better

- *The authors are with the Computer Science Department, Brown University, Providence, RI 02912.*
E-mail: {cad, cj, dbk, sz, dhl}@cs.brown.edu.

Manuscript received 16 Dec. 2004; revised 22 Sept. 2005; accepted 6 Oct. 2005; published online 10 Mar. 2006.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCG-0257-1204.

stereopsis signal) will enhance user performance because they permit the user to segregate the foreground target object from the background context easily [5], [6].

The next section summarizes related work. We explain the setup of both experiments and their results in Section 3. Possible interpretations of the results and related issues are discussed in Section 4, and our conclusions are given in Section 5.

2 Related Work

We are not aware of any work that directly compares CAVEs with fishtank VR displays. However, studies have compared different VR systems with one another as well as with conventional monoscopic desktop systems. Arthur et al., comparing a fishtank VR display to a monitor-based desktop system, reported that fishtank VR significantly improved user performance in a tree-tracing task [7]. They also found that head tracking enhanced user performance more than static stereo. In a more recent study, Ware et al. showed that stereo combined with motion gave the best user performance in 3D visualization of graphs [6], and that passive stereo was more effective than head tracking. Pausch et al. did a qualitative user study to compare a head-mounted display (HMD) with a conventional desktop system for a generic search task [8]. They found that the HMD improved user performance significantly when the target—the virtual object that users were asked to find—was absent. They found no significant user performance difference between the two systems when the target was present. A later study showed, however, that the findings of Pausch et al. did not apply to desktop VR (a conventional desktop system without stereo and head tracking that uses animated interactive 3D graphics to build virtual worlds) [9]. This study showed that fishtank VR and desktop VR have a significant advantage over HMD VR in performing a visual search task.

Bowman et al. recently compared HMD with tabletop (workbench) and CAVE systems for search and rotation tasks, respectively [10]. They found that HMD users performed significantly better than CAVE users for a natural rotation task (turning a corner by rotating the head or body rather than by rotating the virtual world manually using a joystick). They also showed that subjects performed a difficult search task differently depending on which display they used first. Our work builds on previous work by comparing CAVE and fishtank VR platforms directly using both quantitative and qualitative user studies.

3 Methods and Results

The CAVE we used is a four-sided, $8'$ TAN VR-Cube. It has rear-projected front and side walls and a front-projected floor, and each of its four displays has resolution $1,024 \times 768$ pixels. Our fishtank VR display setup has a $22''$ ($20''$ viewable) Mitsubishi Diamond Pro 2070-SB desktop monitor. The display runs at two different resolution modes: a lower resolution of $1,024 \times 768$ pixels and a higher resolution of $1,280 \times 1,024$ pixels. We henceforth refer to the fishtank VR run at the higher resolution mode as the higher-resolution fishtank (or HR fishtank) and to the fishtank VR run at the lower resolution mode as the lower-resolution fishtank (or LR fishtank).

(a) (b)

Fig. 1. The visualization application running in the CAVE (a) and on the fishtank VR display (b).

In both systems, users wear a pair of LCD shutter glasses that support active stereo viewing; a tracker attached to the glasses relays their position and orientation to the computer. We used the most reliable tracker available in each environment at the time of the experiments: an InterSense IS-900 tracker in the CAVE and a Polhemus 3Space Fastrak for the fishtank VR display. The IS-900 is more accurate than the Polhemus but was only available in the CAVE environment.

While we conducted the anecdotal study with only the higher-resolution fishtank VR, we used both resolution modes in the quantitative study.

3.1 Anecdotal Experiment

3.1.1 Application

We used a diffusion tensor magnetic resonance imaging (DT-MRI) visualization application for our anecdotal work. DT-MRI is an imaging modality with the potential to measure fiber-tract trajectories in fibrous soft tissues such as nerves and muscles. The application, based on work by Zhang et al. [11], visualized DT-MRI brain data as 3D streamtube and streamsurface geometries in conjunction with 2D T2-weighted MRI sections. We ran the application both in the CAVE and on the fishtank VR display. The size and placement of 2D sections with respect to the brain model used were the same in both environments. Since our brain models occupied the same relative screen space on both displays, 2D sections displayed in the CAVE were larger than 2D sections on the fishtank VR display (see Fig. 1).

3.1.2 Participant Pool

Five domain-expert users were asked to use the application in the CAVE and on the fishtank VR display. Four of the users were male and one was female. Two of the users started with the fishtank version of the application and the others started with the CAVE version.

3.1.3 Task

Each user had his/her own task (or scientific hypothesis to be tested). They described their hypotheses to us, and we asked them to compare how well the two platforms suited their tasks. The users expressed their opinions by talking to us while they used the application. We offered comments on and counterarguments to their observations in order to

TABLE 1
Advantages Reported for CAVE and Fishtank VR

	User 1	2	3	4	5
Advantages reported for CAVE:					
Has bigger models, one can see more.		✓	✓	✓	✓
Has larger field of view.		✓			
More suitable for gestural expression and natural interaction.		✓			
Possible to walk around.			✓		
Advantages reported for the fishtank VR:					
Has sharper and crisper images.	✓		✓	✓	✓
Displays the data more compactly; spatial relationships between the structures are easier to see.	✓				
Feels more comfortable: not claustrophobic and sitting is better than standing.	✓	✓			✓
Works better for collaboration, especially with two people.	✓				
Easier to point objects on the screen.		✓		✓	
More time-efficient to use; doctors prefer to work-and-go.					✓
Would work better for telemedicine-like collaboration.			✓		
More intuitive for surgery planning because doctors are used to working with brain models at full scale or smaller.			✓		
Overall preferred display:					
The CAVE.		✓			
The fishtank VR.	✓		✓	✓	✓

explore the reasoning behind the users' observations. The users were then asked to give their overall preference for one of the two VR systems.

3.1.4 Results

One user preferred the CAVE and four preferred the fishtank VR display. Table 1 summarizes the users' comments on the relative advantages of CAVE and fishtank VR systems.

Our first user was a neurosurgeon who had used the application before. He uses DT-MRI data to study obsessive-compulsive disorder (OCD) patients and was particularly interested in studying changes that occur after radiation surgery, which ablates an important white-matter region. He wanted to see the relationship between neurofiber connectivity and linear diffusion (streamtubes) in the brain. He strongly preferred using fishtank VR and found no relative advantages of the CAVE.

Our second user was a biologist who was also trying to see correlations between white-matter structure and linear diffusion in the brain. His interests were not confined to a specific anatomical region. He was the only user who preferred the CAVE over the fishtank display.

Our third user was a doctor and a medical school instructor with an undergraduate degree in computer science. She evaluated the application from teaching and learning perspectives.

Our fourth user was a computer science graduate student with an undergraduate degree in neuroscience. Like our second user, he looked at the application to see correlations between white-matter structures and linear diffusion in the brain. He preferred the fishtank VR because

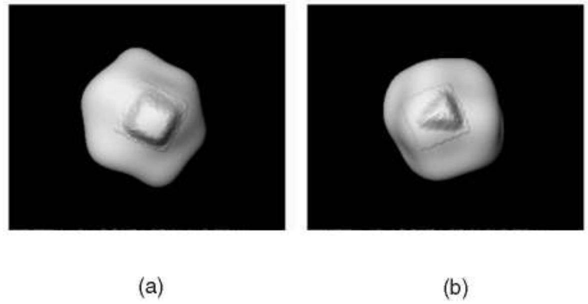

Fig. 2. Two "potato" objects before noise was added. (a) The potato on the left has a rectangular feature while (b) has a triangular feature.

the 2D sections had higher resolution and the models looked crisper on the screen, attributes that helped him see the correlations more easily.

Our last user was a neuroradiologist working on multiple sclerosis who wanted to see the 3D course of the neural fibers in the corpus callosum. He was able to see what he was looking for using both display platforms.

All users found 2D sections to be very helpful in both platforms. They said they were familiar with looking at 2D sections, which helped them to correlate and orient the 3D geometries representing diffusion with the brain anatomy.

3.2 Quantitative Experiment

In this experiment, we compared the two systems by measuring user performance on a visual search task.

3.2.1 Application

Users were asked to identify a feature on a potato-like object with a noisy surface in different visual contexts. The users searched for either a rectangular or triangular feature protruding from the surface of the object; the height of the feature was fixed across trials and exactly one feature was present in each trial.

We generated the potato-like objects in three steps. First, we generated lumpy shapes using spherical harmonics. Each lumpy shape p was formed by scaling a unit sphere s with a combination of harmonic offsets Y_{lm}:

$$p = s \sum_{l \in \{0,2,4,6\}} \sum_{m=-l}^{l} Y_{lm} n_{lm}, \tag{1}$$

where n_{lm} is a normally distributed random number with mean 0 and standard deviation ranging from 1 for the lowest degree, Y_{00}, to 0.27 for the highest degree, $Y_{6,-6}$. The drop-off of the standard deviation was tuned to create subjectively lumpy objects that were not too distorted to conceal the protruding shape.

Second, we created triangular or rectangular extrusions from random positions on the surfaces of these shapes (see Fig. 2). Last, we added noise all over the surface of each object. Fig. 3 shows a potato object with three different noise levels.

Potato objects were shown to users in four different visual contexts or scenes: *blank*, *brick*, *world*, and *porch* (see Fig. 4). The visual contexts varied in content and depth disparity within the environment. In the *blank* context, all

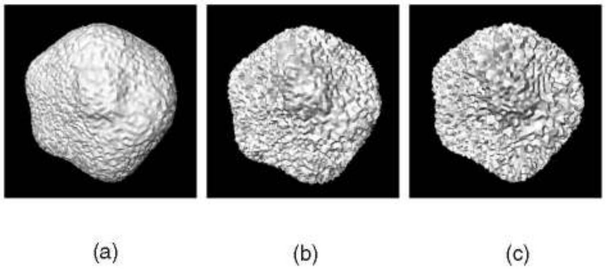

Fig. 3. A potato object with three different noise levels, (a) 0.015, (b) 0.035, and (c) 0.045, in ascending order from left to right. The feature in all three is rectangular.

walls were the same shade of gray. A brick texture was drawn on the walls for the *brick* context. The *world* context had of a brick floor (the same texture as in the *brick* context) extending to a horizon drawn to appear about 100 miles away with mountains and trees and a blue sky above. The *porch* context was identical to the *world* except that a white porch was drawn in the foreground coincident with the physical walls of the CAVE. Note that all the contexts completely covered all the walls of the CAVE.

These four contexts let us test our hypotheses on the benefits of good motion parallax and stereopsis. The *blank* context provides no cues for using motion to segregate the target from the potato object. The *brick* context provides a strong stereo cue (particularly since it is displayed coincident to the wall) and an emphatic texture that should facilitate target-object separation through motion parallax. The *world* context provides only textural elements to promote target-object segregation, but these elements are weak since they are projected to appear far away relative to the object and user. The *porch* context is a hybrid between the features of the *brick* and *world* contexts: The porch provides some near-depth textural information relative to the object, but not as much as the *brick* context. All of the contexts, except for the *blank* context, use the same brick texture on the floor.

In order to exclude brightness as a variable, we adjusted the overall brightness of the different contexts to be about the same. We checked the luminosity values by opening the contexts in Photoshop and by using a photometer.

3.2.2 Participant Pool

Forty-one volunteers participated in the study, 20 of whom were male. They were mostly undergraduate and graduate students with different backgrounds (science, art, and humanities). Fourteen of the participants performed the experiment in the CAVE, another 14 used the higher-resolution fishtank VR system, and the remaining 13 used the lower-resolution fishtank VR system. The experiments all took place over the course of a few days. The experiments for various groups of participants—CAVE, HR fishtank, and LR fishtank—occurred in that order over the course of two years.

3.2.3 Task

Each participant group completed the same set of 60 trials—12 practice and 48 test trials. The order of the test trials was randomized across users. To ensure the best stereoscopic experience, the interocular distance of each user was used in generating the stereo imagery. Users had 30 seconds to complete the task in each trial; trials taking longer than 30 seconds (timed-out trials) were excluded from analysis. On average, the CAVE participants timed out 5.1 percent of the time, the lower-resolution fishtank VR participants timed out 2.4 percent of the time, and the higher-resolution fishtank VR participants timed out 1.8 percent of the time. (These percentages are calculated using only the data from the participants included in the statistical analysis in Section 3.2.5.)

We presented a different context-potato-feature combination at each trial. Each context had the same set of potato-feature combinations, which had equal numbers of triangular and rectangular features. Participants received auditory and visual feedback during practice trials to indicate feature position and response accuracy, but received no feedback during test trials.

The users manipulated the object with a hand-held device, a wand, using their dominant hands. Another wand (three-button) served as a response box (left = square, right = triangle).

The users stood still in the CAVE, but they sat on a chair in front of the fishtank VR display. Interaction in the CAVE was based on an "object-on-a-stick" metaphor: The potato object was fixed at the far end of a virtual stick and coupled with the wand's rotation and translation. The fishtank VR had no virtual stick. Instead, the object was centered and fixed at the user's eye level, and its orientation was coupled to that of the wand. In both environments, we tried to use an appropriate interaction technique. "Object-on-a-stick"

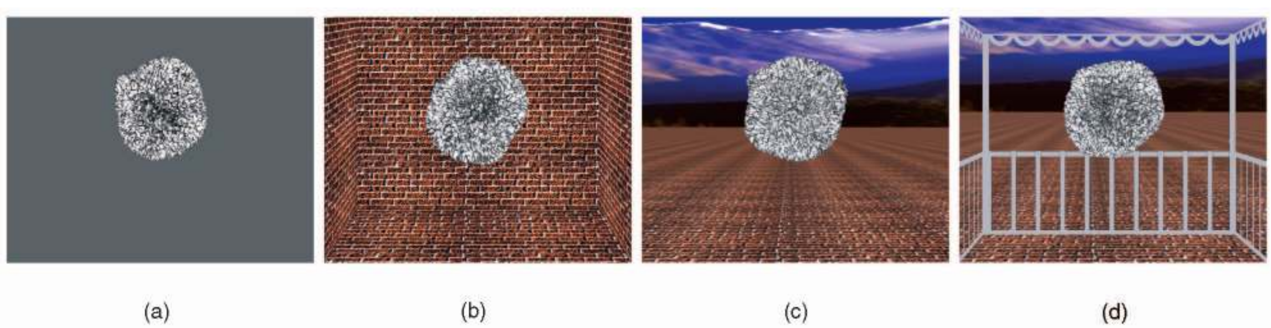

Fig. 4. The four visual contexts, *blank*, *brick*, *world*, and *porch* in our study of user search performance. The potato objects in each context are similar to those displayed during the search task. Note that these visual contexts exploit the CAVE's immersive capabilities by covering of its all four display walls. (a) *blank*, (b) *brick*, (c) *world*, and (d) *porch*.

TABLE 2
Between-Subjects ANOVA Results

Factor	Measure	F Statistic	p Value
display type	time	$F(2,36) = 6.889$	0.003
	error rate	$F(2,36) = 9.129$	0.001

works well in the CAVE, but not for the fishtank VR system: Participants cannot visually track objects off-screen, which would naturally occur when manipulating a potato skewered on the end of a virtual stick. We also noticed that most users in the CAVE held the potato in front of their faces in a relatively fixed position while performing the search task. Therefore, limiting users' ability to translate the potato object in the fishtank environment replicates to some extent the actual visual stimuli of users in the CAVE system.

3.2.4 Factors and Measures

Our quantitative experiment used a mixed-model design (also called a cross-plot or, in biomedical research, cross-sectional design) that contains both between-subjects and within-subjects factors. In our case, display type (CAVE, lower-resolution fishtank VR, higher-resolution fishtank VR) was a between-subjects factor, and feature type (triangular, rectangular) and context (*blank*, *brick*, *world*, *and porch*) were within-subjects factors. Users participated in only one of the independent levels of the between-subjects factor but participated in all the within-subjects factor levels. We recorded the users' task-completion times and task-accuracy values as measures of performance.

3.2.5 Data Analysis

Two users, one from the CAVE and one from the higher-resolution fishtank VR, were omitted from the statistical analysis because they timed out on more than 10 percent of the trials. We thus analyzed the data from 13 different users for each display environment using multifactor mixed ANOVA in SPSS, a commercial statistics package.

3.2.6 Plotting Confidence Intervals

We display *inferential confidence intervals* around means in our bar graphs. These intervals are calculated directly from the ANOVA analysis. Using *inferential confidence intervals* simplifies drawing pairwise statistical inferences from data plots of ANOVA results [12]. The figures indicate a statistically significant difference to the $p = 0.05$ level between any two means if the intervals around the two means do not overlap. We used Bonferroni's correction in calculating pairwise comparisons.

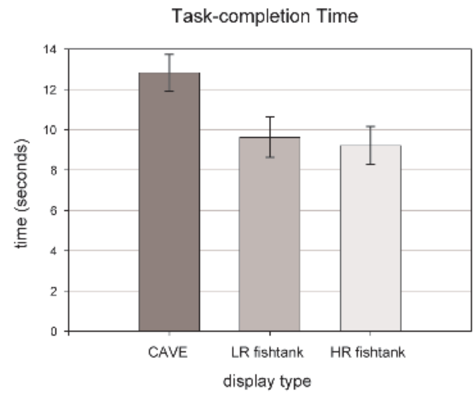

Fig. 5. Mean task-completion times for the CAVE, LR fishtank VR, and HR fishtank VR displays. Users identified the features faster on the HR fishtank and LR fishtank VR displays than in the CAVE.

3.2.7 Results

Users were significantly faster and more accurate on the fishtank VR system at both resolution modes than in the CAVE. There was no statistically significant difference in user performance between the HR fishtank VR and the LR fishtank VR (see Table 2 and Table 3).

Also, disproving our hypothesis, visual context had no significant effect on accuracy or speed in either environment ($F_{time}(3,108) = 2.002$, $p_{time} = 0.118$, $F_{error\,rate}(3,108) = 1.025$, $p_{error\,rate} = 0.385$).

Fig. 5 and Fig. 6 show mean task-completion time and error rate values for the CAVE, LR fishtank and HR fishtank. Error bars on the graphs indicate statistical significance: Overlapping error bars for any measurement indicate that the difference between the measurements is not statistically significant.

4 Discussion

4.1 Looking-In versus Looking-Out Tasks

The results of both studies are consistent with our hypothesis that fishtank VR displays are more conducive to *looking-in* tasks than CAVEs. In a *looking-in* task, the user views and manipulates a virtual world from the outside in, and interacts with a virtual object that is smaller than his/her body and fits in the fishtank VR display. The potato search task is a *looking-in* task. Interaction with our DT-MRI brain application also provides a *looking-in* perspective, as does brain surgery.

A *looking-out* task, by contrast, shares the user's frame of reference: The user views and manipulates the virtual

TABLE 3
Mean and Standard Error for Task Performance for Each Display Type

Measure	Display Type	Mean	Standard Error	Significant Differences[1]
time	CAVE	12.831	0.754	LR fishtank ($p = 0.015$), HR fishtank ($p = 0.005$)
	LR fishtank	9.212	0.754	CAVE ($p = 0.015$)
	HR fishtank	9.628	0.754	CAVE ($p = 0.005$)
error rate	CAVE	0.151	0.19	LR fishtank ($p = 0.001$), HR fishtank ($p = 0.01$)
	LR fishtank	0.044	0.19	CAVE ($p = 0.001$)
	HR fishtank	0.068	0.19	CAVE ($p = 0.01$)

[1] *Significant differences were determined through pairwise comparisons using Bonferroni's correction.*

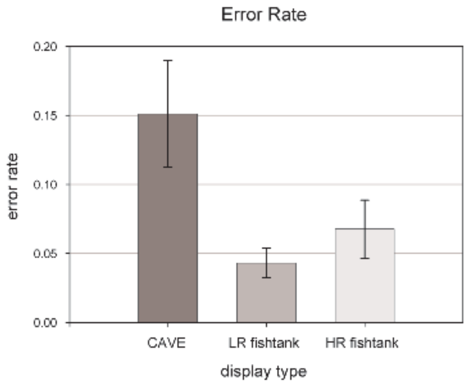

Fig. 6. Mean error rates for the CAVE, LR fishtank VR, and HR fishtank VR displays. Users were more accurate on the HR fishtank and LR fishtank VR displays than in the CAVE.

world from the inside out. The user interacts with a virtual object that is larger than his/her body and fills his/her foveal and peripheral vision. Thus, *looking-out* tasks require users to use their peripheral vision more than *looking-in* tasks.

We believe that fishtank VR systems are a natural fit for *looking-in* tasks because 1) a fishtank VR display physically separates the user's frame of reference from the virtual object's frame of reference, which forces the user to look into the virtual world, and 2) a fishtank VR display physically fits in the user's field of view. On the other hand, we expect *looking-out* tasks—such as architectural walkthroughs, navigation in unfamiliar terrain, or urban combat—to benefit more from CAVE-like displays, which fill the user's peripheral vision, than from fishtank VR displays.

4.2 Why No Significant Effect of Visual Context?

Contrary to our hypothesis, we did not find statistically significant differences between the visual contexts. One reason for this finding may be that the dependent measure was not sensitive enough to capture subtleties in user performance. If, however, our measure was sensitive enough, these findings are contrary to previous results in the 2D visual search domain. One possible reason for this is that many previous studies focus on semantic (relational) contextual influences between the background and the target objects in the scene [13], while we focused on the psychophysical traits of the contexts. Another reason may be that the properties of the potato object-target feature conjunction were so obvious as to let users perform the task without needing to rely on information external to the object (i.e., context) for object-target segregation; previous 2D studies purposely manipulated the signal-noise ratio between the target and distractors to gauge the extent to which similarity and edge ambiguity confounded search performance. In the present study, however, since the task required users to segregate and identify the target feature from the object, they may have ignored the context completely and focused their (foveal) vision and attention on the object. Furthermore, Wolfe et al. argue that separation of targets from a background scene is a preattentive step that allows subsequent focus of attention on candidate target items; if the segmentation of the potato object-feature pairs here is so salient as to be preattentive, one could argue that we should see no differences between the contexts [14].

We believe that strong motion and stereo cues, lacking in previous 2D visual search studies, masked any differences in performance due to the four different visual contexts in this study. The current study displayed the 3D scenes in stereo with hand-coupled and head-coupled motions together. Motion cues provide both robust configural information about the scene and strong depth cues. Along with stereo, these cues help disambiguate information in the scene, particularly the structure of the feature with respect to the object. Previous research also suggests that users make fewer errors in comprehending visual data when using hand-coupled or head-coupled motion together with stereo viewing [6]. Also, our study presented scenes for much longer viewing times than the studies performed with 2D images and thus allowed the user to explore the object more thoroughly. A final reason for the absence of differences may be the type of the task used: In *looking-in* tasks, the user's interaction with background and periphery is inherently limited.

Although we found no significant differences between the contexts, we believe that we may find contextual differences using a *looking-out* task if our suppositions about the nature of the contexts is correct. One of the possible reasons described above for finding no contextual differences was that users did not "see" the contexts; users tended to put the potato objects right in front of them to search it efficiently for the target feature, and this necessarily limited their view of the contexts. Conversely, users must view the scene as a whole in order to perform a *looking-out* task, since this type of task necessitates seeing the objects as well as the context. It is thus possible that differences in user performance between contexts might become apparent for a *looking-in* task using an object with a semi-transparent or sparse (e.g., lattice) surface.

4.3 Task, Stimuli, and Interaction Choice

We chose the visual search task of finding a protruded feature on a noisy surface for two reasons. First, the task seemed an appropriate abstraction of the exploratory data visualization process. In our observation, an important part of the work of scientists doing exploratory involved the examining geometric structures in order to locate hard-to-find features. In partially reconstructing this process, we decided to use rectangular and triangular protrusions from potato-like forms as the hard-to-find features of these forms. We adjusted the difficulty of the task by changing the amount of the noise added to the surface and the extent of the perturbations to the underlying shape. We believed that there would be a general correspondence between performance on our task and performance in scientific explorations. Second, the suggestion in the literature for 2D visual search that context would make a difference supported the choice of a visual search task.

We chose interaction methods in the different environments using a "best-of-breed" approach: We tried to pick interaction methods that would be as effective as possible for each environment. This approach was intended to avoid biasing results by using identical methods of interaction when those methods would be inappropriate in one of the environments.

In retrospect, given that we did not measure a difference due to visual context, the advantages of the immersive CAVE environment were probably underutilized. As we discussed above, using *looking-out* tasks is one approach to utilize these advantages better. There are also other task choices that might make better use of the CAVE environment including relationship identification and navigation.

4.4 Should Display Differences Be Normalized?

CAVE and fishtank VR systems differ not only in intrinsic design such as size and field of view, but also in extrinsic technology-related limitations such as brightness and resolution. We chose to compare the displays as they were, not normalizing for the extrinsic differences; subsequent studies might be able to attribute some of the differences we find to specific extrinsic factors, such as angular resolution. Kasik et al. have already showed the positive effect of a crisp display on user performance [15]. Quantifying the effects of by-design and by-technology differences is important in steering the future efforts of both application developers and display designers.

4.5 Abstract versus Application-Specific User Studies

We can categorize user studies as abstract or application-specific according to the tasks and performance criteria used. The two schemes are complementary, and each has its own uses. Abstract user studies evaluate user performance on abstract tasks and contexts using generic performance criteria. The quantitative study presented here was an abstract user study. It is inherently easier to reproduce, quantify, and generalize abstract user studies than application-specific ones. However, how to transfer results from abstract user studies to real application domains is not obvious without a formalism that can relate the abstract tasks evaluated to a specific real-world application.

On the other hand, application-specific studies use tasks, contexts and performance criteria that are directly related to an application domain, so that their results can potentially find immediate implications in that domain. Recent work by Laidlaw et al. reported an application-specific quantitative user study that compared different 2D field visualization methods using expert (and nonexpert) users for a set of domain-specific objective tasks such as locating and identifying critical points and advecting a particle in the field [16]. The present anecdotal work using the DT-MRI brain visualization application was a qualitative, application-specific study.

4.6 Some Observations Concerning Head-Tracking and Stereo

We observed that users generally did not move their heads during the quantitative user study. We tested this observation by repeating the study with the higher-resolution fishtank VR without head tracking, using three subjects. While the number of subjects is suitable only for a pilot study, the results, consistent with our observations, suggest that turning off head-tracking may not make a significant change the users' speed or accuracy. This may be because of the nature of our *looking-in* task: Viewing and manipulating a relatively small object from the outside in reduces the need for head motion. Also, the small size of the virtual object facilitates faster rotation with the hand-held device, which is more natural in viewing an object of that size than moving the head.

Similarly, in order to test the effect of stereo display in the quantitative study, we repeated the study on the higher-resolution fishtank VR without stereo using six new subjects. We found that stereo helped the users locate the feature faster, but had no significant effect on the identification of the feature type. Users were statistically significantly faster with stereo display than with mono ($F_{time}(1,17) = 7.877, p_{time} = 0.012$). This finding is in line with earlier work: Ware et al. showed that stereo with hand-coupled motion was the most effective combination in 3D graph visualization [6]. There was not, however, a significant difference in task accuracy between stereo and mono cases in the present study ($F_{error\,rate}(1,17) = 1.219, p_{error\,rate} = 0.285$). This is probably because of the "binary" accuracy measure used. We anticipate a significant effect if a more sensitive measure is used.

5 Conclusion

For our visual search task, and possibly for similar *looking-in* tasks:

- Fishtank VR displays are more accurate and faster than CAVEs.
- Visual context did not demonstrate a statistically significant effect on user performance.
- When the user manipulates a virtual object using a six-DOF hand-held device coupled with the object, stereo display contributes significantly to user performance; head-tracking may not.

Acknowledgments

The authors thank John F. Hughes for sharing insights on spherical harmonics and Daniel F. Keefe for his help in software development. They also thank Katrina Avery for her excellent editorial help. This work has been supported partially by US National Science Foundation (CCR-0086065, CCR-0093238, and CNS-0427374).

References

[1] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti, "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the Cave," *Proc. 20th Ann. Conf. Computer Graphics and Interactive Techniques,* pp. 135-142, 1993.
[2] C. Ware, K. Arthur, and K.S. Booth, "Fish Tank Virtual Reality," *Proc. Conf. Human Factors in Computing Systems,* pp. 37-42 1993.
[3] I. Biederman, A.L. Glass, and E.W. Stacy Jr., "Searching for Objects in Real-World Scenes," *J. Experimental Psychology,* vol. 97, no. 1, pp. 22-27, 1973.
[4] J.M. Wolfe, "What Can 1 Million Trials Tell Us about Visual Search?" *Psychological Science,* vol. 9, pp. 33-39, 1998.
[5] D. Mumford, S.M. Kosslyn, L.A. Hillger, and R.J. Herrnstein, "Discriminating Figure from Ground: The Role of Edge Detection and Region Growing," *Proc. Nat'l Academy of Science,* vol. 84, pp. 7354-7358, 1987.
[6] C. Ware and G. Franck, "Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions," *ACM Trans. Graphics,* vol. 15, no. 2, pp. 121-140, Apr. 2000.

[7] K. Arthur, K. Booth, and C. Ware, "Evaluating 3D Task-Performance for Fish Tank Virtual Worlds," *ACM Trans. Information Systems,* vol. 11, pp. 239-265, 1993.
[8] R. Pausch, D. Proffitt, and G. Williams, "Quantifying Immersion in Virtual Reality," *Proc. 24th Ann. Conf. Computer Graphics and Interactive Techniques,* pp. 13-18, 1997.
[9] G. Robertson, M. Czerwinski, and M. van Dantzich, "Immersion in Desktop Virtual Reality," *Proc. 10th Ann. ACM Symp. User Interface Software and Technology,* pp. 11-19, 1997.
[10] D.A. Bowman, A. Datey, U. Farooq, Y.S. Ryu, and O. Vasnaik, "Empirical Comparisons of Virtual Environment Displays," Technical Report TR-01-19, Dept. of Computer Science, Virginia Tech., 2001.
[11] S. Zhang, Ç. Demiralp, D. Keefe, M. DaSilva, D.H. Laidlaw, B.D. Greenberg, P. Basser, C. Pierpaoli, E. Chiocca, and T. Deisboeck, "An Immersive Virtual Environment for DT-MRI Volume Visualization Applications: A Case Study," *Proc. Conf. Visualization '01,* pp. 437-440, 2001.
[12] W.W. Tryon, "Evaluating Statistical Difference, Equivalence, and Indeterminacy Using Inferential Confidence Intervals: An Integrated Alternative Method of Conducting Null Hypothesis Statistical Tests," *Psychological Methods,* vol. 6, no. 4, pp. 371-386, 2001.
[13] M.M. Chun, "Contextual Cueing of Visual Attention," *Trends in Cognitive Science,* vol. 4, pp. 170-178, 2000.
[14] J.M. Wolfe, A. Oliva, T.S. Horowitz, S.J. Butcher, and A. Bompas, "Segmentation of Objects from Backgrounds in Visual Search Tasks," *Vision Research,* vol. 42, pp. 2985-3004, 2002.
[15] D.J. Kasik, J.J. Troy, S.R. Amorosi, M.O. Murray, and S.N. Swamy, "Evaluating Graphics Displays for Complex 3D Models," *IEEE Computer Graphics and Applications,* vol. 22, pp. 56-64, 2002.
[16] D.H. Laidlaw, M. Kirby, C. Jackson, J.S. Davidson, T. Miller, M. DaSilva, W. Warren, and M. Tarr, "Comparing 2D Vector Field Visualization Methods: A User Study," *IEEE Trans. Visualization and Computer Graphics,* vol. 11, pp. 59-70, 2005.

Çağatay Demiralp received the BS degree from Ege University, Turkey, in 2000 and the master's degree from Brown University in 2004, both in computer science. His research interests include computer graphics, scientific visualization, and human interaction in VR environments.

Cullen D. Jackson received the PhD degree in experimental psychology from Brown University and the BS degree in computer science from Trinity University. He is a postdoctoral research associate in the Computer Science Department at Brown University. His research interests include scientific visualization, human object and space perception, 3D user interactions, and intelligent interfaces. He is a member of the IEEE.

David B. Karelitz received the master's degree in computer science from Brown University in 2003. His current research interests include high-resolution visualizations and making visualization easier for end users.

Song Zhang received the BS degree in computer science from Nankai University, China, in 1996 and worked on optical character recognition algorithms at the Nankai Institute of Machine Intelligence from 1996 to 1998. Since 1998, he has been a computer science doctoral student at Brown University. His research interests include scientific visualization, medical imaging techniques, and computer graphics.

David H. Laidlaw received the PhD degree in computer science from the California Institute of Technology, where he also did postdoctoral work in the Division of Biology. He is an associate professor in the Computer Science Department at Brown University. His research centers on applications of visualization, modeling, computer graphics, and computer science to other scientific disciplines.