A Comparative Study of Desktop, Fishtank, and Cave Systems for the Exploration of Volume Rendered Confocal Data Sets

Prabhat, Andrew Forsberg, Michael Katzourin, Kristi Wharton, and Mel Slater

Abstract—We present a participant study that compares biological data exploration tasks using volume renderings of laser confocal microscopy data across three environments that vary in level of immersion: a Desktop, Fishtank, and Cave system. For the tasks, data, and visualization approach used in our study, we found that subjects qualitatively preferred and quantitatively performed better in the Cave compared with the Fishtank and Desktop. Subjects performed real-world biological data analysis tasks that emphasized understanding spatial relationships including characterizing the general features in a volume, identifying colocated features, and reporting geometric relationships such as whether clusters of cells were coplanar. After analyzing data in each environment, subjects were asked to choose which environment they wanted to analyze additional data sets in—subjects uniformly selected the Cave environment.

Index Terms—Participant studies, participant evaluation, scientific visualization, virtual reality, volume rendering.

✦

1 INTRODUCTION

T HIS paper reports on an experiment that compared the performance and subjective evaluation of biologists who used three different types of virtual environment (VE) systems to explore their confocal microscope data. The three systems were a standard Desktop display, a Fishtank display, and a Cave-like system. (In this paper, we use the term "Cave" to refer to the generic type of VE system described in [6]). We selected these particular platforms because our science collaborators have access to them for their visualization needs—thus, the study results would help them make an informed decision as to the relative value of each. This was a within-group study with 12 participants. The outcome demonstrates overwhelmingly that superior results were obtained with the Cave system with respect to both performance and various evaluations of subjective preference. Our main contribution is identifying significant differences between relevant systems for an important scientific task. We also speculate on some possible reasons for the results.

Systems such as the Cave offer unique advantages in terms of the multiple large display surfaces, large field of view, and body-centered interaction. Such properties facilitate presence, the extent to which participants respond realistically to VEs that depict scenes from everyday life [19]. Moreover, experimental evidence suggests that, for memory and manipulation tasks, experimental participants using a Cave system perform better than in a less immersive interactive workbench [24]. However, it is not clear that the high degree of immersion provided by a Cave-like system benefits scientific understanding of three-dimensional data compared to the use of other types of display systems. In particular, in this context, is the possible benefit of the use of a Cave outweighed by economic considerations such as cost, space needed, lack of transportability, and so on? In a previous experimental study, Demiralp et al. [7] showed that, indeed, participants using a Fishtank display performed better than in the Cave both with respect to task performance, comfort, and subjective evaluation, where the task involved exploration of visual structures in a way such that the participant's viewpoint was always outside of the structure—in other words, where the participant could at all times see and manipulate the structure from the outside. In this context, by a Fishtank display, we mean a head-tracked active stereo display based on a standard Desktop monitor [2]—which is significantly less costly both in terms of finance and space requirements than a Cave. However, our own anecdotal experience with biologists who use the Cave routinely for the exploration of confocal data sets contradicted the notion that the Fishtank would be preferred to the Cave since, generally, when given the chance, the biologists have preferred to use the Cave. Is this just a question of the Cave being a more unusual and interesting experience (the "cool factor"), or does it reflect a real utility that was not captured in the study by Demiralp et al.?

We therefore carried out an experiment that compared both the performance and subjective evaluation of biologists when analyzing real data in the same Cave and

• Prabhat is with the Center for Computation and Visualization, Brown University, Box 1024, Providence, RI 02912. E-mail: prabhat@cs.brown.edu.
• A. Forsberg and M. Katzourin are with the Department of Computer Science, Brown University, Box 1910, Providence, RI 02912. E-mail: {asf, mkatzour}@cs.brown.edu
• K. Wharton is with the Department of Molecular Biology, Cellular Biology, and Biochemistry, Box G-J160, Brown University, Providence, RI 02912. E-mail: Kristi_Wharton@brown.edu.
• M. Slater is with ICREA-Universitat Politécnica de Catalunya, Barcelona, Spain. E-mail: melslater@lsi.upc.edu.

Fishtank systems used by Demiralp et al. and, additionally, a Desktop monoscopic system. In Section 2, we provide further background material regarding the impact of immersion on performance. In Section 3, we describe the systems and participant interfaces that were used by the biologists. In Section 4, we describe the methods and details of the experiments with results in Section 5. Discussion and conclusions follow in Sections 6 and 7.

2   BACKGROUND

There are very few participant studies on direct comparison between different virtual reality platforms for scientific exploration tasks. Most studies have focused on interaction techniques, presence, and usability. Participants are typically asked to perform relatively simple "benchmark" tasks. We are primarily focusing on the free-form exploration of scientific data without an emphasis on speed/accuracy performance trade-off. Participants are asked to perform realistic scientific tasks similar to ones they would conduct in their everyday research.

Sutcliffe et al. [24] evaluated the Cave, Interactive Workbench, and Reality Room for comparing presence, memory, and interaction. Participants were asked to perform a maintenance operation involving removing a bracket component from an aircraft fuselage. The Cave was found to have better usability, was better in terms of memory, and provided a higher sense of presence to its users.

Johansson and Ynnerman [10] presented an evaluation of Immersive Workbench, Desktop-VR, and a Desktop system. Participants were asked to find errors in mechanical products. Their results showed that participants using Desktop-VR performed best; however, there was no statistically significant difference between the different platforms. Mizell et al. [14] presented a comparative evaluation of immersive VR with conventional displays. For the task of assembling a physical replica of a complex 3D object, participants performed significantly better using an immersive virtual reality (IVR), especially when the enlarged object surrounded the participants.

Arthur et al. [2] introduced Fishtank VR and examined its usefulness by testing two conditions: whether or not a stereoscopic display was used and whether or not the perspective display was dynamically coupled to the positions of a user's eyes. We did not vary the same conditions in our study, so we do not know how turning off the stereo or head-tracking separately would have affected user performance. Instead, we were concerned with the impact of the different systems as a whole.

Gruchalla [8] presents a case study wherein users perform better in a real-world oil-well path editing task in an immersive environment as compared to a conventional Desktop. Our work is in a similar spirit; we report results from a real-world scientific exploration task, albeit from a different scientific domain. Raja et al. [18] and Arns et al. [1] present case studies wherein they report that immersive environments improve the performance of users in abstract information and statistical visualization tasks.

Demiralp et al. [7] presented work that is the most closely related to ours. They carried out a qualitative and quantitative evaluation of the Cave and Fishtank environments for polygonally rendered data sets. Participants preferred the Fishtank environment over the Cave for examining the connectivity of pathways. For an abstract task of looking at binary features on a noisy potato shape, participants performed better on the Fishtank. The data sets and tasks used in this study are very different from that in [7], and we characterize these differences and those of the experimental setup and results in Section 6.1.

Recent work by North [16] and Saraiya et al. [21], [20] has looked into an insight-based evaluation of information visualization tools. In [21], domain users utilized bioinformatics tools in a relatively open-ended protocol to make observations, new hypothesis, and gain insights. A domain expert codified the evaluation criteria and different tools were evaluated for their effectiveness in gaining insights. In [20], a few domain experts participated in a longitudinal study over a much longer period of time and commented on insights gained by using different tools. Our work is related in spirit to this: We are interested in the insights gained by using scientific visualization tools on real data sets. We also used an open-ended qualitative "think-aloud" protocol. However, we are evaluating different virtual reality platforms as a whole rather than individual tools. Also, our study was conducted over a relatively short period of time, so the nature of our insights is more "observational," rather than the kinds of complex deep insights that are referred to in [16] and [20].

Mania and Chalmers [12] compared the subjective presence, task performance, and cognition state employed to complete a memory task in both real worlds and simulated virtual worlds. Our study measured task performance actively (subjects answered questions in each environment) rather than through a memory recall task, and there is effectively no real-world equivalent to confocal microscopy VEs to do a complete real-versus-virtual comparison. In our study, we did not collect information related to the subjects' cognitive states (that is, episodic or semantic memory and "familiarity" of retained knowledge), but learning how knowledge was represented would be an interesting future work. Finally, we did not measure subject presence levels since this was not a depiction of an environment where one could "be." All these differences may help to account for the fact that, in general, Mania and Chalmers's study showed that *presence* (the subjective state) was not correlated with task performance but that our study showed a strong relationship between immersion and performance, where *immersion* refers to the immersive capabilities of the system rather than the subjective state of mind of the users.

Bowman and Raja [4], Ni et al. [15], McMahan et al. [13], Tan et al. [25], and Polys et al. [17] present methods for the evaluation and separating effects that may help explain the findings reported here. However, their studies differ substantially in data and subject task, and some do not include a Desktop or Fishtank environment, the standard environment for computer-aided data analysis.

3   MATERIALS

3.1   Confocal Data Sets

Laser scanning confocal microscopy has revolutionized biological science research during the past 10-15 years by providing biologists with the ability to collect high-resolution images while eliminating out-of-focus information that

is associated with image degradation. Confocal microscopy allows the user to optically section a thick specimen, as opposed to physically sectioning a specimen to reveal the cellular structure deep inside the organism. By labeling different molecules or cellular structures with fluorescent dyes or proteins, it is possible to simultaneously visualize the distribution of multiple molecules throughout a specimen. Despite the dramatic technological improvements that confocal microscopy and the development of new synthetic fluorescent molecules have brought to the biological sciences, the reconstruction and analysis of a series of optical sections still poses a profound challenge to most researchers.

Nine confocal microscopy data sets representing three different tissues were used in this study: adult Drosophila egg chambers, third instar larval Drosophila ring gland and brain, and midgut. Each data set consisted of a Z-stack of 40 sections at 512×512 pixel resolution. The specimens had been prepared by standard immunohistochemical protocols [3] and incubated with various fluorescent antibodies or nucleic acid stains to highlight the position of nuclei and the expression of specific proteins. In each case, the presence or absence of a protein, its distribution, and its localization to subcellular compartments, that is, nuclei, membranes, or its colocalization with other proteins, provides valuable information with regard to the molecular mechanisms driving different developmental processes.

In the case of the egg chambers dissected from female Drosophila ovaries, three different structures or proteins were labeled. Nuclei were labeled with a nucleic acid stain, Hoeschst, highlighting the different sizes of nuclei typical of the different tissue types making up the egg chamber. Moreover, two proteins, Broad or Hts, that show a localized expression pattern in the egg chamber were labeled with antibodies (anti-Broad-Z1 or anti-htsRC obtained from the Developmental Studies Hybridoma Bank http://dshb. biology.uiowa.edu/). Broad is expressed in a distinct set of follicle cells that exist in a layer over the future egg, oocyte. Hts is expressed in subcellular structures, ring canals that connect the nurse cells, a subset of the cells comprising the egg chamber.

The larval ring gland/brain specimens were also labeled with Hoechst to highlight nuclei. The size and position of the nuclei identify different regions of the ring gland that are typified by cells that have different physiological functions. The ring gland is innervated by cellular projections from the brain called axons, and they were labeled with an antibody (anti-Futsch/22C10—Developmental Studies Hybridoma Bank).

The nuclei of cells in the larval midgut samples were also labeled with Hoechst enabling identification of the different cellular layers of the midgut. The longitudinal and circular visceral muscles of the midgut were identified by staining for the filamentous protein actin using rhodamine-conjugated phalloidin (Sigma). Phalloidin also highlights an intracellular structure present in a group of specialized cells (copper cells) that line the lumen of a segment of the midgut. The third signal in these images arose from a fluorescently tagged fatty acid that was fed to the larvae before dissection and accumulated in the lumen of the midgut.

TABLE 1
Features of Three Environments

	Stereo	Head Tracking	Display	Screen Size
Desktop	No	No	IBM P260	16''x 12''
Fishtank	Yes	Yes	IBM P260	16''x 12''
Cave	Yes	Yes	Marquee 9500	8'x 8'

3.2 System Details

Participants were asked to examine confocal data sets on three different platforms: a monoscopic Desktop system, a Fishtank system, and a Cave-like system, whose features are described in Table 1.

Identical linux machines with Nvidia 4500G graphics cards were used to drive all platforms. The Cave system (measuring $8' \times 8' \times 8'$) used in this case had three display walls and a floor display. Each display used an Electrohome Marquee 9500LC projector, which was driven by a single linux machine. Individual screens on all platforms had a display resolution of $1,024 \times 768$ pixels. Participants were seated ~ 2 feet from the Desktop/Fishtank screen and stood ~ 4 feet from the front wall of the Cave; no restrictions were imposed on their physical motion. The display field of view (FOV) for all systems was defined by the viewing position and physical screen size. In the Cave and Fishtank, these varied as the subject moved their head. The software FOV was always set to match the display FOV.

A Polhemus Fastrak system was used to provide head/wanda tracking in all conditions. CrystalEyes stereo glasses were used in all cases. Note that, while the glasses do not have an effect in the mono Desktop case; we asked all participants to wear them nevertheless. Fig. 1 shows participants interacting on these three platforms.

On all platforms, a 3D wanda device was used to provide 6 degrees of freedom input to manipulate the data sets. The wanda joystick rotated the data set using a standard trackball implementation. In order to translate the data set, the participant pointed to the data set (aligning a rendered virtual pointing line coincident with the pointing direction of the wanda) and pressed the left button on the wanda to drag the data set. Releasing the left button stopped the translation motion of the data set.

3.3 Volume Rendering

The data sets in this study were volume rendered using the VOlume eXplorer (VOX) software [22]. The program uses hardware-accelerated texture rendering to draw volumes at interactive rates. The sizes of the data sets used in this study were $512 \times 512 \times 40$. We obtain a rendering performance of 20-25 fps on all platforms. The rendering performance and display latency was never mentioned as an issue by any of the experimental participants.

3.4 Recruitment

Twelve participants from the Division of Biology and Medicine at Brown University volunteered as a result of either an email or poster inquiry for individuals familiar with cell and developmental biology. They consisted of two faculty members, one research staff member, three graduate

Fig. 1. Mono Desktop, Fishtank, and Cave environments used in the experiment.

students, and six undergraduate students. There were three female subjects and nine male subjects. For the purpose of this study, participants had to be able to appreciate the structure of different biological tissues and the presence/ significance of florescent signals specific for the localization of a particular protein or nucleic acid. The participants needed to be able to understand the spatial layout of the tissue and the various components of cells making up test tissues. A Prequestionnaire was used to verify if the participants were qualified and had sufficient background knowledge to make such observations.

All participants had worked with confocal data sets prior to their participation, although the level of experience in dealing with these data sets varied. The faculty members and research staff member were intimately familiar with the confocal imaging process, limitations of confocal image quality, 2D/3D reconstructions of confocal Z-stacks, and their analysis. Graduate and undergraduate students were generally familiar with collecting confocal data sets and analysis with 2D and 3D imaging software.

4   METHODS

4.1   Experimental Design

This was a within groups design with 12 participants. Each participant viewed the data in the Desktop, Fishtank, and Cave systems. Hence, there were six possible orderings: DFC, DCF, FDC, FCD, CDF, and CFD. There were two participants randomly assigned to each of these six conditions. In each system, they viewed three different types of Drosophila data: the egg chamber, the brain, and the gut—always in this same order. Hence, there were nine real confocal data sets chosen with the help of an expert faculty member in the Molecular Biology, Cell Biology, and Biochemistry (MCB) Department. The data sets used are shown in Fig. 2. The egg chamber data sets 1) tend to be relatively sparse—with the nucleus of each cell labeled with the localization of specific proteins highlighted by a different fluorophore. The brain data sets 2) have a ring gland that is innervated by axonal pathways. The axons weave their way from the larval brain into the ring gland. The gut data sets 3) are relatively dense and have multiple layers of tissue.

4.2   Procedures

4.2.1   Initial Setup

Prior to the beginning of the study, participants were given a standard consent form. This informed them of their rights,

risks, and right to stop at any time without giving reasons. Participants were asked to fill out a prequestionnaire form that collected data on their background, experience with confocal data sets, and 2D/3D imaging programs. Participants were then given written instructions on the general tasks in the study; they were asked to examine the confocal data sets and "*talk out loud about the features that you observe as if you were trying to teach us about the features.*" They were told that they would be asked specific questions about each data set, and they were free to continue exploration of the data sets as they deemed fit. They were also instructed that there was no fixed time-limit on these tasks and that they were free to explore the data sets as long as they wanted. We did, however, ask them to finish the entire study in a reasonable amount of time, which was noted as 2-3 hours. All participants were videotaped during the study. They were paid $10 per hour for their participation.

One egg chamber data set was used in a training phase. Each of the three egg chamber data sets, brain data sets, and gut data sets were used in the experimental phase. For the final choice phase (described below), one egg chamber data set and one brain data set were used. These two additional data sets were different from the nine chosen for the testing phase. It is not possible to objectively quantify the complexity of realistic data sets, but the biology expert ranked all egg chamber data sets to be roughly of the same level of difficulty and similarly for the brain and gut data sets.

4.2.2   Training Phase

Participants were first shown an example egg chamber data set on each platform. The goal of the training data set was to familiarize them with the user interface and display characteristics. On each of the three platforms, they were asked to drag the data set around, bring it close to them, and move their head. In the Cave, participants were asked to drag the data set to the multiple walls of the Cave and to move around physically. This was done to ensure that they understood the capabilities of each platform. Toward the end of the exploration, they were asked to count the number of nuclei present in a certain chamber and identify the sizes of nuclei present. Once they indicated that they were comfortable with the interactions, the test data sets were brought up for examination.

4.2.3   Tasks

During the testing phase, participants examined an egg chamber, brain, and gut data set in that order. These data
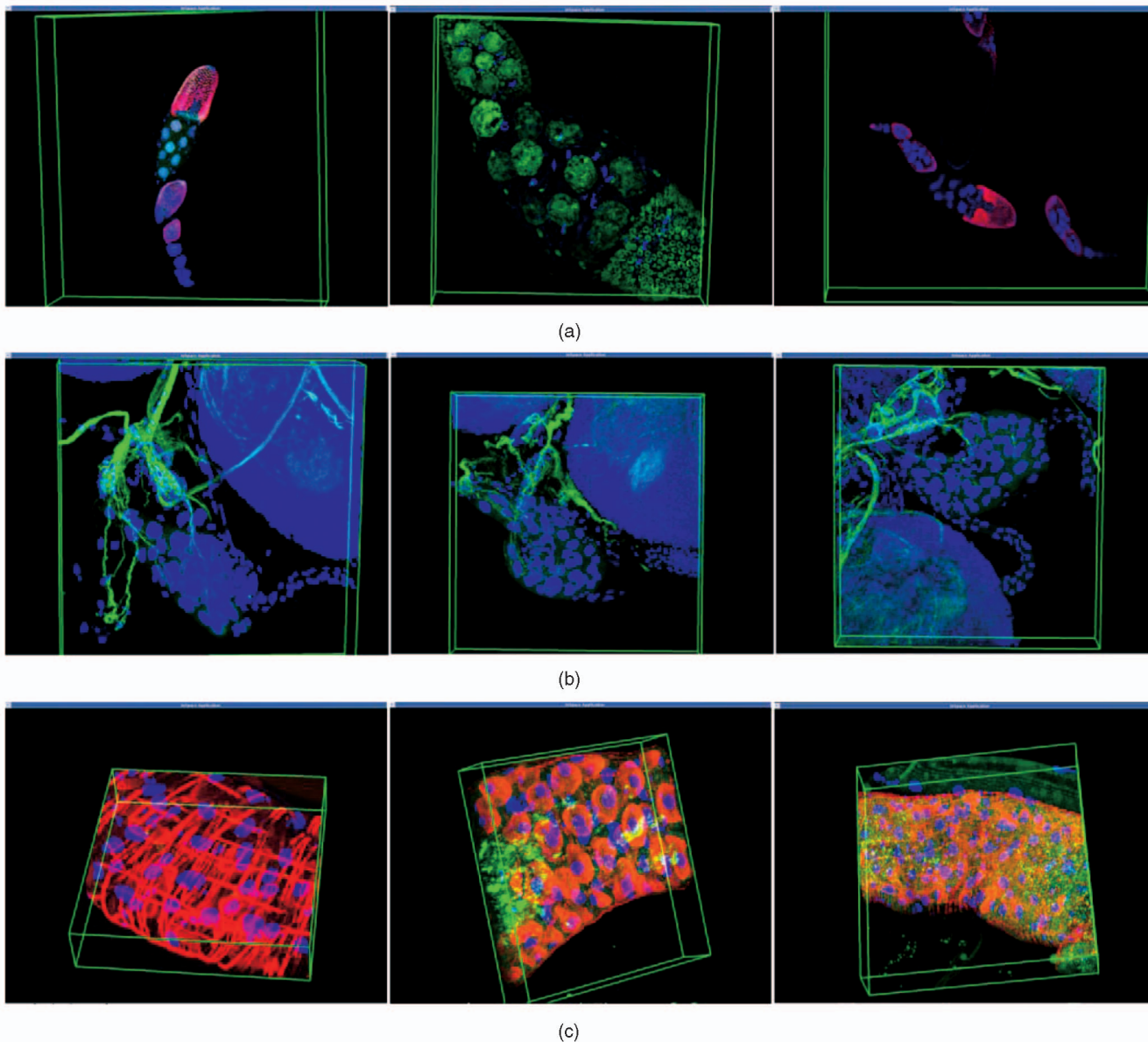
Fig. 2. Views of the data sets used in this experiment. (a) Egg chamber data sets. (b) Brain data sets. (c) Gut data sets.

sets were randomly assigned to the participant to minimize the chance of platform-data set interaction. Also, the participant was assigned in advance to one of the six possible orderings of Desktop, Fishtank, and Cave.

The participant was initially asked to talk out loud about features that they might see in the data set. The expert had determined an exhaustive list of features that were present in each of the data sets. This information was used by the experimenter to ask specific questions related to the data set being examined. The participant typically pointed to features in the data set in response to the question, but no feedback as to the correctness of their answers was provided. Through this interactive process of questioning, the experimenter made sure that the participant explored the data set sufficiently, that is, without trivially answering the questions and making sure that they had paid attention to features. For example, if in response to a colocalization question, a participant might respond "I think there is red-green colocalization" without being more specific, they

were further asked to point out where the colocalization was and, if possible, to count the number of such regions/points. Similarly, in response to determining contact between axons and nuclei, if a participant generally responded "I don't think axons are in contact with the nuclei," they were asked to follow the path of a particular axon and count how many nuclei the axon was close to. In practice, we observed that all of our participants were thoroughly engaged in the assigned tasks.

For the Egg chamber data sets, participants were informed of the labeling in each case (blue is typically DNA, red may be Broad Complex, etc.). The following typical questions were asked:

- How many kinds of nuclei do you see (small, medium, or large)?
- Do you see colocalization of pairs of colors (red and green, green and blue, or red and blue)?
- How many nuclei are present in this chamber?

- Would you describe the shape of a certain feature as convex or concave?
- Can you locate the oocyte?
- How many ring canals do you see in this data set?

For the Brain data sets, participants were informed of the labeling (blue is DNA and green is a protein present in axons). They were asked to comment on the ring gland in the center and how the axons project onto the ring gland. Typical questions were as follows:

- How many different size populations of nuclei do you see in the ring gland (one, two, three, etc.)?
- Do the axons touch any of the nuclei in the ring gland?
- Are the axons present near the surface of the ring gland or do they go deep inside?
- Do the axons appear to touch any of the cells? How many cells do they touch?
- How many times does an axon branch?

For the gut data sets, they were informed of the labeling (blue is DNA, red is filamentous actin, and green is fatty acids). They were asked the following questions:

- How many differently sized populations of nuclei do you see?
- Are these populations located near the outside, inside, or uniformly dispersed throughout (small nuclei might be present only on the outside)?
- Comment on the distribution of the fatty acids. Are they clustered or randomly dispersed throughout? If they are sparse, do they lie in a plane?
- Does the actin colocalize with the nuclei?
- Are the nuclei always encapsulated in cup-shaped actin enriched structures?
- How many layers of actin do you see? How many layers of nuclei are present?

In order to successfully answer these questions, the participants had to look at the data sets from multiple perspectives. For example, in order to answer whether red and green were colocalized, it was imperative that they would examine the voxels from multiple points of view. If a red voxel was in front of a green voxel, it may have appeared yellow (colocalized) from a front view, but a side view would dispel that idea. In order to determine whether there was contact between the axons and nuclei, the participant would have needed to either drag the data set or move his/her head to follow the path of the axon between the nuclei. In order to count the number of layers in a gut data set, they would have had to consider different views of the gut and infer that there were concentric layers of different types of tissue.

All the questions asked of the participants had a biological significance. Although the process of answering a question can be thought of as finding/resolving a geometrical feature in the data set, the questions were motivated solely by biological considerations.

4.2.4 Platform Selection

After the testing phase was complete, the participant was asked to move to a neutral position away from the systems. The participant was informed that two more data sets would be examined, one being an egg chamber and the other a brain data set, and that the tasks would be the same

as before. The participant was then asked to select which system he or she would use for these tasks. These tasks were then performed on the chosen system.

4.2.5 Postquestionnaire and Debriefing

After the participant had completed all the tasks of examining data sets, they were asked to complete a further questionnaire that assessed their subjective comparisons between the systems on a number of criteria. Following this, they were debriefed and answered a number of questions in an interview carried out by the experimenters.

5 RESULTS

5.1 Task Performance Response Variables

An exhaustive list of features present in each data set had been prepared in advance by the biology expert. Each feature was weighted according to importance with weights determined in advance by the expert. After the experiments were completed, the video recording for each participant was reviewed by one of the experimenters. Scores were assigned according to the extent to which the participant found and discussed each particular feature. An example is shown in Table 2. The first column of the table shows the particular feature to be identified, whereas the second column shows the importance attributed to those features as a score out of 100.

The final column shows the score between 0 and 1 being the experimenter's assessment of how well the participant did on that particular feature. The overall score is then the weighted average over all features. A similar analysis was carried out for each participant on each data set across every platform. This results in three response variables that provide a measure of task performance. We refer to these as "egg," "brain," and "gut," respectively. Each is a value between 0 and 100, where 100 is an equivalent of what the expert would have scored. Note that, in exceptional cases, participants were awarded 10 extra points if they did additional sophisticated (regular or comparative) analysis beyond what was enumerated by the expert. Out of a total of 108 trials (12 users $\times$ 9 data sets), extra points were awarded in only five instances.

Table 3 shows the means and standard deviations of these response scores over all subjects for each type of display. It can be seen that, for each category of data, the performance in the Cave was superior to that using the Fishtank, which in turn was superior to that using the Desktop system. This is shown again in Fig. 3. One-way analysis of variance (ANOVA) was used to test the null hypotheses for the egg, brain, and gut data sets, which is that the results from the three displays were from the same underlying distribution, against the alternative hypotheses that the means of the distributions are different. For Fig. 3a, the null hypothesis is rejected ($p = 0.0002$). Multiple comparison tests (with an overall significance level of 5 percent) show that the mean Cave score is significantly higher than the mean scores of each of the other two, but there is no significant difference between the other two. For Fig. 3b, the null hypothesis is rejected ($p = 0.0003$) and the multiple comparison tests give the same result as for Fig. 3a.

TABLE 2
Scoring the Performance of a Participant on an Egg Data Set

Feature	Prior Weight	Score
General description	20	0.5
Find the ring canals	10	1.0
How many ring canals are present (16)	10	0.75
How many canals touch the oocyte (4)	10	0.5
How many kinds of nuclei are present (2)	15	0.0
Follicles are uniformly distributed in the top-left younger chamber	10	0.0
Large green mass in present in the lower right (oocyte)	5	0.5
Faint blue is dispersed near top of chamber	10	1.0
DNA density is differentiated	10	0.0
Total	100	45

For Fig. 3c, the null hypothesis is rejected ($p = 0.0001$), and the results of the multiple comparison tests are again the same. The residual errors of each of the three ANOVAs were tested for normality with a Jarque-Bera test [9], and the hypothesis of normality was not rejected in any case, the significant levels being 0.95, 0.60, and 0.45, respectively.

However, there were six different orders of presentation of the three systems, Desktop (D), Fishtank (F), and Cave (C). It is possible that order may be a confounding factor, and hence, we consider whether there was any order effect.

Fig. 4 shows plots of the mean and range of the performance by type of data and order (these are drawn as line graphs solely for reasons of clarity). It can be seen throughout that the performance on the Cave is at least as good as that of the other two for every data set and for every order except one (order DFC on brain). In order to understand this better, we carried out a two-way analysis of variance for each data type with factors Display (three levels) and Order (six levels) and two observations per cell. The two-way ANOVA tests the hypotheses that the Display means are all equal and also that the Order means are all equal. Then, multiple comparison tests (all simultaneously at the 5 percent level) can examine the relationships among the means in detail.

For the egg data, the hypothesis that the display means are equal can be rejected ($p = 0.001$), the hypothesis that the means of all order effects are equal cannot be rejected ($p = 0.31$), and there is no interaction effect ($p = 0.83$). The multiple comparison test suggests that the Cave mean is significantly greater than the Desktop mean, but there is no significant difference between the mean Desktop and Fishtank and between the mean Fishtank and Cave.

For the brain data, the significance level for Display is $p = 0.0001$, for Order is $p = 0.71$, and here, there is an interaction effect $p = 0.03$. Hence, we would conclude that there is a significant difference between the Display means, with the multiple comparison tests revealing that the Cave mean is significantly higher than the means of Desktop and Fishtank, but there is no significant difference between these two. The interaction effect can be accounted for by inspection in Fig. 4b and order DCF, where there is a change in ranking between the means. There is no particular explanation about why this result is different from all of the rest.

For the gut data, the significance level for Display is $p = 0.0016$, for Order is $p = 0.96$, and for Interaction is $p = 0.69$. Hence, in this case, there is a difference between the Display means, with the Cave mean again being significantly higher than the other two but with no significant difference between Desktop and Fishtank.

The residual errors of each of the above ANOVAs have been tested for Normality using a Jarque-Bera test, and in each case, the hypothesis of Normality cannot be rejected ($p = 0.72, 0.79$, and 0.96, respectively, for egg, brain, and gut residual errors).

Although it is acceptable to carry out an ANOVA with $n = 2$ replications per cell, obviously, more would be preferable. Fig. 5 shows the situation when results from all three data types are combined, so that there are $n = 6$ observations per cell. Again, it is clear that the mean Cave results are, as a whole, higher than the others. An ANOVA with six replications per cell reveals a highly significant difference between the Display means ($p < 0.0001$) but no difference in Order ($p = 0.48$) and no interaction effect ($p = 0.14$). The multiple comparison tests shows that the mean Cave result is significantly higher than the mean Fishtank result, which in turn is significantly higher than the mean Desktop result. The distribution of the residual errors of the model is consistent with a Normal distribution ($p = 0.30$).

TABLE 3
Means and Standard Deviations
of the Task Performance Responses

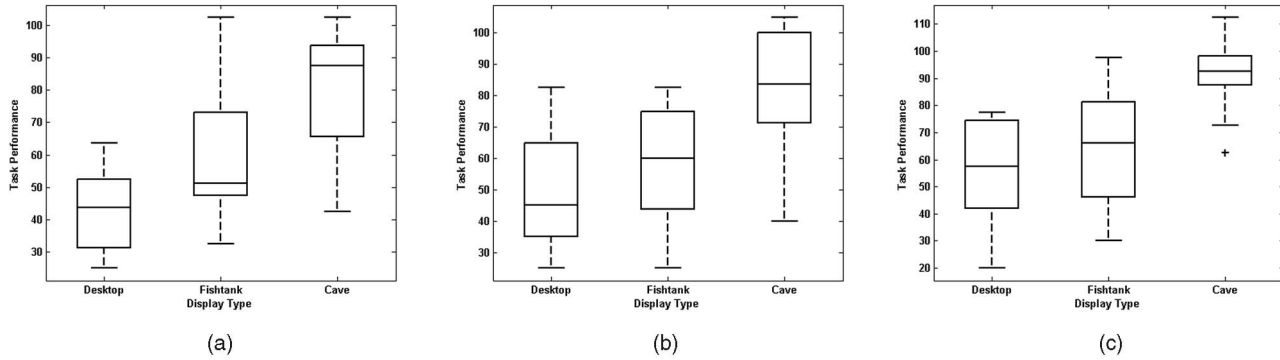|	Desktop	Fishtank	Cave
Egg	43.44 (12.77)	59.27 (20.86)	79.27 (20.22)
Brain	48.54 (18.29)	58.96 (18.78)	82.50 (19.13)
Gut	56.88 (18.66)	64.17 (21.72)	91.35 (13.44)

Fig. 3. Box plots of task performance for (a) egg, (b) brain, and (c) gut for display types. The boxes show the interquartile range with the median as the horizontal bar. The whiskers extend to the minimum and maximum of the data or, if these are outside the range of 1.5 times the interquartile range, then the outliers are shown as crosses.
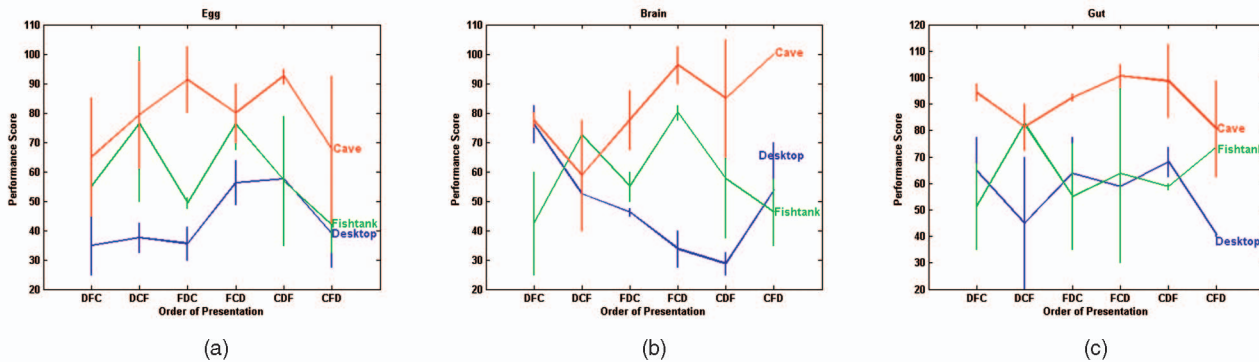

Fig. 4. Plot of means and range of performance data by Order of Presentation ($n = 2$ per entry).

5.2 Regression Analysis on Task Performance Responses

The analysis in the previous section does not take into account the influence of confounding variables such as the experience of the participants in dealing with confocal data sets and other possible factors.

The independent factor (main effects) of the design was of course the type of display (1 = Desktop, 2 = Fishtank, 3 = Cave). Order was taken into account by the use of two factors "order1" (= 1, 2, or 3) representing the type of display used first and "order2" (the type of display used second).
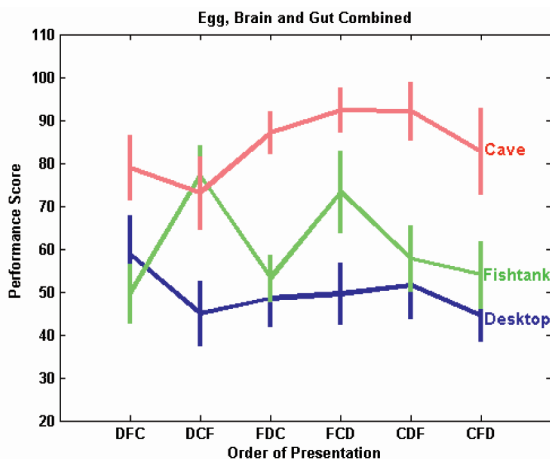

Fig. 5. Combined performance results for egg, brain, and gut data. Each graph is the mean and standard error of all $n = 6$ performance scores by order and type of data.

Additional explanatory variables available from the prequestionnaire were

- Gender (1 = Male, 2 = Female);
- Occupation (1 = undergrad, 2 = grad, 3 = research staff, 4 = faculty);
- Previous experience with the use of Desktop, Fishtank, or Cave (1 = no, 2 = yes);
- Experience with confocal data (1 = Never looked at them before, 2 = Novice, just starting to look at these data sets, 3 = Familiar with these data sets, look at them occasionally, 4 = Experienced in looking at these data sets);
- Experience with imaging software (how often used imaging software: 1 = Never, 2 = Occasionally, once a year, 3 = Once a month, 4 = Once a week, 5 = Almost every day); and
- Experience with 3D imaging software (how often imaging data is looked at with 3D software: 1 = Never, 2 = Occasionally, once a year, 3 = Once a month, 4 = Once a week, 5 = Almost every day).

These types of explanatory variables are important since these may influence the performance of participants, yet we are not really interested in this. We use the regression analysis to factor out such explanatory variables where they significantly affect the response variable.

For the egg performance data, when these additional variables are included in a normal regression analysis, none of them are significant, so the Display remains the only significant variable. For the brain performance data, the basic model including display, order, and the interaction

TABLE 4
Analysis of Questionnaire Responses on Usability for Desktop (D), Fishtank (F), and Cave (C)

Question		Median	IQR	Analysis of Variance
How would you rate each platform in terms of learning about the data-sets?	D	4.5	1	$p = 4.7 \times 10^{-7}$. All displays are significantly different from one another.
	F	6	0.5	
	C	7	0	
How would you rate each platform in terms of answering the specific questions about the datasets?	D	4	1	$p = 4.05 \times 10^{-6}$. D is significantly lower than both the other two.
	F	6	1	
	C	7	1	
How would you rate each platform in terms of ease of use in exploring the datasets?	D	3.5	4	$p = 0.0257$. D and C are significantly different, but there is no difference between D and F or F and C.
	F	6	1.5	
	C	6	1	
How confident are you about finding all common/regular features on the environments?	D	4	3.5	$p = 0.0031$. D and C are significantly different, but there is no difference between D and F or F and C.
	F	6	1.5	
	C	6.25	1	
How confident are you about finding all unusual/anomalous features on the environments?	D	2	1	$p = 5.66 \times 10^{-5}$. D is significantly lower than F and C.
	F	5.5	0.5	
	C	6	0	

The median and interquartile ranges are shown for each of the questions. A Kruskal Wallis one-way analysis of variance was used to produce the last column. The P values are for rejecting the hypothesis of equal medians. The subsequent comments are based on multiple comparison tests all at the 5 percent level of significance.

between them was used. On top of this, having used a Fishtank system before was significant and positively associated with the brain score ($t = 2.85$ on 18 d.f., $p = 0.011$). No other variables were significant. For the gut data, none of the additional variables were significant.

5.3 Analysis of the Postquestionnaire Data

In this section, we consider the several questions where participants were asked to compare between Desktop, Fishtank, and Cave. All responses in the postquestionnaire were measured on a Likert scale of 1-7, where 1 meant the worst response (unproductive, the platform hindered my exploration, very difficult and probably missed features), and 7 was the best response (productive, I could not have answered the questions otherwise, very easy and confident of observing all features). We use the nonparametric Kruskal-Wallis one way ANOVA to test between the responses on the three systems [11]. We use nonparametric statistics since we are dealing with ordinal data, where the absolute values have no intrinsic meaning. The results are shown in Table 4.

5.3.1 Other Questions

All participants but one preferred the Cave (Overall, what environment would you prefer for looking at your data sets?), and all least preferred the Desktop (Overall, what environment would you least prefer for looking at your data sets?).

Regarding the "perceived value" of features "for exploring and conducting research" on the data sets, the median scores for stereo, head-tracking, large display, and large field of view are 7, 5, 7, and 7, respectively, actually with head tracking significantly lower than the other two (5 percent level).

There are no differences between the three systems with respect to eye strain or any of the other comfort variables, which, as a whole, have low scores indicating comfort.

In general, we observe that the Cave is rated best for all of these questions followed by the Fishtank and the Desktop. The Cave is rated to be significantly better than the Desktop and Fishtank for learning about the data sets, answering specific questions and ease of use. It is important to note that, although the participants were mixed in their rating for confidence level in finding all common features (as compared to Fishtank), they feel significantly more confident about finding unusual/uncommon features in the Cave.

5.3.2 Open Questions

Participants were also asked to comment on the advantages and disadvantages of the Desktop, Fishtank, and Cave environments and discussed specific features that they liked or disliked in the environment. Table 5 lists the number of times that the features were mentioned in favor or against each platform. For example, three participants reported negative statements about the Cave regarding its lack of vibrant and clear colors, whereas four people made positive statements about the image clarity and resolution of the Fishtank. Note that these categories were not recommended or suggested to the participants but were spontaneously mentioned.

Individual comments from participants indicate that, with the Cave, "It was easier to follow an axon," "I could quickly get a sense of spatial relationships," "I could rotate the data set quickly without losing detail," "It was easier to see/count individual structures," "It was easier to verify colocalization," and "It was easier to know when all details have been spotted and when to stop looking."

For the Fishtank, individual comments read "It was easy to get a quick sense of spatial relationships," "I did not get an immersive feeling," "The data set kept getting cutoff by the screen," and "It was difficult to find an appropriate viewing size, I kept losing focus."

For the Desktop, individual comments read "It was difficult to explore images," "It was difficult to verify colocalization," "It was difficult to follow the path of an axon," "When I turned the image, the side view became

TABLE 5
Comments on Specific Features of the Three Systems

Feature	Comments on the Cave		Comments on the Fishtank		Comments on the Desktop	
	Positive	Negative	Positive	Negative	Positive	Negative
Vibrant and clear colors	-	3	2	-	2	-
Image clarity and resolution	-	2	4	-	1	-
Large image area	5	-	-	3	-	1
Data looks good close up	4	-	-	4	-	2
Can walk into and around the data	6	-	-	3	-	-
Ease of manipulation	5	-	-	4	2	4

For each system, the number of times that a specific feature was mentioned positively or negatively by the participants is shown.

unclear," and "I had to make assumptions while analyzing the data."

5.3.3 Postquestionnaire Conclusions

Participant ratings from the postquestionnaire indicate that they felt most at ease and most confident of their observations in the Cave, followed by the Fishtank. The Desktop, in the words of one participant, "was a distant third." Participants seemed to indicate that image quality, size of the image or the ability to view data close up, and the ability to move into and around the data are factors in their exploration. The Fishtank is mentioned for its image quality and resolution, whereas the Cave is mentioned for a general ease of exploration and spatial understanding. The overall preference in favor of the Cave seems to indicate to us that, for these tasks and these data sets, the advantages offered by the Cave seem to outweigh those provided by the Fishtank and Desktop environments in the minds of the participants.

5.4 Behavior of Participants

Table 6 shows a classification of the proportion of time that the participants spent interacting with the data sets. It can be seen that they tended to rotate the data set longer in the Desktop than in the Fishtank and Cave systems. The participants typically dragged the data sets to a comfortable position and exclusively used rotation thereafter.

On the Fishtank system, participants tended to drag the data set in and out of the screen to a distance where they could fuse the image comfortably. Near the beginning of the training phase, they quickly realized that bringing the data

TABLE 6
Behavior of Participants on the Different Displays—The Percentage of Time They Engaged in Dragging, Rotating, or Not Moving the Data Set

	Dragging dataset	Rotating dataset	Neither	Total
Desktop	4%	64%	32%	100%
Fishtank	4%	53%	43%	100%
Cave	18%	43%	39%	100%

set very close to themselves would cause the images to separate (that is, they could not fuse the images any longer), and they pushed it back. Unlike the Desktop, participants selectively used rotation when it was necessary.

On the Cave system, participants dragged the data set onto one, two, or three walls. Typically, participants kept the data set in front of them (front wall + floor) and dragged it closer if needed. Rotations were used when needed, but certainly not as much as on the Desktop system. For the brain and gut data sets, some participants immersed themselves in the data set and followed the axon pathways. For future studies, it would be interesting to further classify the amount of time that users spent in head rotation, leaning, crouching, or walking; we did not instrument our logs with that information for the present study.

5.5 Platform Choice

Although questionnaires are a useful method to assess the experiences of the participants, it is important to present a situation to them in which their professed preference has practical consequences for their future actions. In our study, after the participants had examined the data sets on all three platforms, they were asked to perform a similar task on two further data sets on any of the three platforms as they wished.

All 12 of them opted to use the Cave for examining the last two data sets. When posed with the question, 10 participants said they would "definitely use the Cave." Two participants said that they would either use the Fishtank or the Cave system. To help them decide, we asked them "if you were paying $100 to use the Fishtank or the Cave to explore the next two data sets, what would you choose?" Both participants immediately chose the Cave. All 12 users examined the data sets following their choice of platform.

5.6 Debriefing

There was a debriefing interview in which the participants were told the purpose of the study and asked a number of subsidiary questions.

In response to a question on ease of interaction and comfort level, all participants responded that they felt that the interaction was intuitive and were generally comfortable on all systems. Two participants commented that they felt mildly nauseous in the Cave, but soon overcame this. Five participants commented that they were uncomfortable with dragging the data sets close to them on the Fishtank, whereas it was very easy to do so in the Cave. All participants were of the opinion that they were most at ease with the interaction in the Cave environment. The interaction felt the most natural and intuitive.

Participants were told to imagine a situation where, over the next year, they had $1,000 to spend on carrying out

similar imaging analyses. They were told that they should abstract away from all concerns such as travel and suppose that all systems were available to them in their own laboratory. The situation would be that they could use the Desktop system freely, but for the Fishtank or Cave, they would be required to pay $100 for each use. Which system would they choose on the whole? All participants preferred to use the Fishtank or Cave environment rather than the Desktop. Some commented that they would use the Desktop for quickly scanning/verifying the data sets and then proceed onto the Fishtank/Cave for detailed analysis. When given a forced choice between Fishtank and Cave, everybody chose the Cave. The subsequent questions tried to ascertain how much more they would pay for the Cave over the Fishtank given their finite budget. The median level that participants would pay for the Cave over the Fishtank was $200, with a range from $150 to $400.

We would not like to associate any significance to the exact cutoff limits determined by the participants in response to this hypothetical situation. It would be a mistake to interpret these numbers as "on the average, the Cave is perceived to be twice as useful as the Fishtank." The response depends a lot on the background of the participant, their prior research experience, etc. For example, one participant replied "Data is important, if I need to understand an important data set I'll pay anything I can." One participant commented "I won't pay more than $200 for the Cave; it's unlikely I'll see twice as much."

In spite of the objections raised above, we think that it is instructive to ask this question. Often, people are concerned about the "cool factor" of the Cave or the Fishtank, the fact that people might be encountering them for the first time and, hence, rating them higher. Six participants had been to the Cave before, and the other six were using it for the first time. None of them had seen a Fishtank system before. By asking the participants to assess the value of the system in a practical situation, we can somewhat force them to look beyond the "wow" factor and assess if this would be practically useful for them. In addition, we feel that the process of questioning the participant during the testing phase ensured that they explore the data set thoroughly and scientifically, thus leaving little room for the "cool factor" to influence their decisions.

6 DISCUSSION

6.1 Comparison with Other Work

It is interesting to note that our results are different from those reported in the work by Demiralp et al. [7]. Although the 12 participants uniformly chose the Cave environment over Fishtank/Desktop in this present study, Demiralp et al. report that five (out of six) participants preferred the Fishtank system for exploring a brain data set in a qualitative evaluation. The primary reasons for preferring the Fishtank were stated as better display characteristics (brightness and angular resolution). Our experiment, which used an identical setup to their study, had similar display characteristics, and yet, the results are quite different. Although it is hard to compare two participant studies with different participant populations and tasks directly, we highlight a few key differences in the two experiments.

Participants in the qualitative study described in [7] performed a limited range of spatial tasks—they were primarily looking for fibers connecting one region of the brain to another. Although we do not have access to their participant behavior data, we hypothesize that, in order to perform connectivity tasks in a dense, opaque, occluded data set, a participant can sit back, rotate the data set, and try to resolve the features. Our tasks (such as colocalization, spatial distribution of features, and searching for contact points) required the participant to consider the data sets from many more vantage points.

Participants in our study were looking at volume-rendered data sets, which, in general, tend to consist of clouds of points without objective boundaries or edges. The data sets in [7] were polygon rendered and primarily consisted of nonintersecting tubular structures. We hypothesize that participants were following sharp features and structures in their tasks, whereas in our tasks, participants were primarily concerned with estimating sizes, checking for colocalizations, checking for intersections between features, and generally assessing overall structural components. Our data sets (at least the egg chamber and brain data sets) are relatively sparse as compared to the representation used in [7]. Our representation shows fuzzy blobs, whereas the other representation is analogous to a dense group of mangled wires. It is likely that a participant does not feel comfortable about placing their head in this type of dense structure due to the degree of presence induced by the stereo and head-tracked displays.

In our task, participants are allowed more freedom in moving the data sets (translation and rotation), whereas Demiralp et al.'s experimental setup fixed the data set in a virtual position in the world. Participants could move their head or walk around (in the Cave), but they could not drag/rotate the data set. In practice, participants do not take many steps in the Cave—they tend to drag data sets onto different screens or rotate the data set instead. By restricting the movement of the data, the Cave is effectively reduced to a one wall display that did not take full advantage of its capabilities.

Participants were primarily asked for their subjective feedback in [7]. The exact nature of the operations that each participant might have performed could have been different across participants since they were only instructed to evaluate the quality of the display as a part of the instructions before the study started. Although this protocol presents some interesting insights, a more controlled study where all participants need to conduct identical well-defined tasks is preferable. Instructing them to evaluate the quality of the display will likely distract them from their task and, especially if the quality of the display is a factor, it will affect their subsequent comments/feedback. Moreover, although subjective feedback is useful, it is instructive to present a situation to the participant wherein he or she has to consider all the advantages and disadvantages of the platform(s) and decide to act upon a final choice based on their practical experience with specific tasks.

6.2 Speculations on the Results

The conclusion from this study is that, both with respect to performance measures and subjective evaluation, the Cave was generally preferred to the Fishtank, which was in turn preferred to the Desktop. The most interesting comparison is between Cave and Fishtank, since these each have characteristics that may have been thought suitable for this

type of visualization, and in some respects, the Fishtank has advantages with respect to the Cave (such as resolution and color), and it appeared to outperform the Cave in the earlier study discussed in Section 6.1.

What can account for the differences that we observed? It cannot be due to stereo or head tracking, since the Cave and Fishtank both have these features. Moreover, for this experiment, color and resolution of the display were not overriding factors since these are better for the Fishtank than the Cave. We suggest that there are two interlocked factors that led to the superior results for the Cave: size and engagement of the whole body in the data exploration. The Cave displays the structures on the same scale as the size of the human body, and this, together with tracking, enables participants to use their body as part of the very process of data exploration; they can move around, bend down, and look around in addition to moving the structure itself by grabbing and rotating and translating it. In other words, acquiring understanding of a visual structure is not simply a function of using the eyes, but additional information arises from proprioceptive and vestibular feedback from body movements.

There is an analogous situation in VEs that provides some support for this hypothesis. When people use virtual reality in order to explore a space, the evidence suggests that they typically do not learn to navigate this space well, and a reason put forward for this is that, usually, navigation in a VE is carried out while participants are in a static position but using a wand or joystick button in order to move virtually. A recent review of this literature can be found in [5], where it is suggested that there needs to be body engagement in navigation and that the lack of self-movement associated with navigation leads participants to build incomplete cognitive maps. Indeed, one study in a head-mounted display-delivered VE provides some evidence that navigation needs to be accompanied by appropriate self-motion in order to achieve strong presence. In that between-groups study [26], participants moved through a VE by the traditional point-and-click joystick, by walking in place, or by really walking (making use of a wide area ceiling tracker). It was found that the level of presence was greatest for those who really walked, was less for those who simulated walking by walking in place (though there was no significant difference between these two), and was significantly lower than both of these for those who used the joystick. Although data exploration is, of course, somewhat different from navigation, these share the similar characteristic that, in order to understand the data (as with a space), one has to build the equivalent of a spatial cognitive map of it. If body self motion is a necessary part of this process, then it is less likely to be achieved while sitting in front of a Fishtank display since, although it has head tracking, the limited size of the display and, therefore, the working space, does not support such body engagement.

6.3   Future Work

Our future work will therefore address the issue of what specifically is it about the Cave that makes it a preferred platform to the Fishtank and Desktop in the context of the type of realistic data that we have described in this paper. Given the discussion above, our starting point is the idea that body engagement is a critical feature that enables successful data exploration (for this type of data).

However, there are many factors that vary between a Cave and a Fishtank system: larger display area, multiple display surfaces, larger field of view, and different display resolution and brightness characteristics. Ideally, we would vary these factors one at a time and see if the participant preference/performance changes. Clearly, this requires a much larger number of participants. Given the specialized nature of these data sets and qualifications required to appreciate the features in these data sets, it is challenging to recruit a sufficient number of qualified participants.

The question of measuring performance is even more challenging in some sense. Typical participant studies simplify the task so that a performance metric can be formulated and measured. Unfortunately, the simplification process leaves the task bearing no resemblance to the actual complex scientific tasks that researchers perform in real life. It is very hard to generalize results from such studies and see how they would apply to a realistic scenario.

Although we have selected a wide range of real-life confocal data sets in this study, we have not quantified the "difficulty" metric of a data set. Arguably, if such a metric could be formulated, results from studies similar to ours could be generalized to comparable data sets. It is likely that a range of different metrics need to be considered depending on the task, for example, depth/occlusion complexity could be considered for counting features and their intersections.

In a more recent study, also with real biological data and participants with a range of expertise, we have used specific quantitative tasks (counting the branching structures of neurons, counting the number of cells in a chamber, etc.) and have measured performance on such tasks, and this will be reported in a later paper.

7   CONCLUSION

We have presented an in-depth participant study that compares participant experiences on three different platforms (Desktop, Fishtank, and Cave) for exploring volume-rendered confocal data sets. The selected tasks were motivated by actual biological explorations and were performed on real data sets. Participants quantitatively performed better in the Cave environment. Participants indicated that they preferred the Cave most for these tasks, followed by the Fishtank and Desktop platforms. This preference is expressed through consistent responses in their choice of environment to explore in postquestionnaires and a debriefing session.

REFERENCES

[1] L. Arns, C. Cruz-Neira, and D. Cook, "The Benefits of Statistical Visualization in an Immersive Environment," *Proc. IEEE Virtual Reality (VR '99)*, vol. 88, 1999.

[2] K.W. Arthur, K.S. Booth, and C. Ware, "Evaluating 3D Task-Performance for Fish Tank Virtual Worlds," *ACM Trans. Information Systems,* vol. 11, no. 3, pp. 239-265, 1993.

[3] E. Bangi and K.A. Wharton, "Dpp and Gbb Exhibit Different Effective Ranges in the Establishment of the BMP Activity Gradient Critical for Drosophila Wing Patterning," *Developmental Biology,* vol. 295, pp. 178-193, 2006.

[4] D. Bowman and D. Raja, "A Method for Quantifying the Benefits of Immersion Using the CAVE," *Presence-Connect,* vol. 4, no. 2, 2004.

[5] J.R. Brotons-Mas, S. O'Mara, and M.V. Sanchez-Vives, "Neural Processing of Spatial Information: What We Know about Place Cells and What They Can Tell Us about Presence," *Presence-Teleoperators and Virtual Environments,* vol. 15, no. 5, pp. 485-499, 2006.

[6] 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.

[7] C. Demiralp, C.D. Jackson, D.B. Karelitz, S. Zhang, and D.H. Laidlaw, "CAVE and Fishtank Virtual-Reality Displays: A Qualitative and Quantitative Comparison," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 3, pp. 323-330, May/June 2006.

[8] K. Gruchalla, "Immersive Well-Path Editing: Investigating the Added Value of Immersion," *Proc. IEEE Virtual Reality (VR '04),* Mar. 2004.

[9] C.M. Jarque and A.K. Bera, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economics Letters,* vol. 6, no. 3, pp. 255-259, 1980.

[10] P. Johansson and A. Ynnerman, "Immersive Visual Interfaces-Assessing Usability by the Effects of Learning/Results from an Empirical Study," *J. Computing and Information Science in Eng.,* vol. 4, no. 2, pp. 124-131, 2004.

[11] W.H. Kruskal and W. Allen Wallis, "Use of Ranks in One-Criterion Variance Analysis," *J. Am. Statistical Assoc.,* vol. 47, no. 260, pp. 583-621, Dec. 1952.

[12] K. Mania and A. Chalmers, "The Effects of Levels of Immersion on Memory and Presence in Virtual Environments: A Reality Centered Approach," *Cyberpsychology and Behavior,* vol. 4, no. 2, pp. 247-264, 2001.

[13] R.P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D.A. Bowman, "Separating the Effects of Level of Immersion and 3D Interaction Techniques," *Proc. ACM Symp. Virtual Reality Software and Technology,* 2006.

[14] D.W. Mizell, S.P. Jones, M. Slater, and B. Spanlang, *Comparing Immersive Virtual Reality with Other Display Modes for Visualizing Complex 3D Geometry,* http://www.cs.ucl.ac.uk/research/vr/Projects/Immersion/Experiment2/paper.pdf, 2002.

[15] T. Ni, D.A. Bowman, and J. Chen, "Increased Display Size and Resolution Improve Task Performance in Information-Rich Virtual Environments," *Proc. Conf. Graphics Interface,* 2006.

[16] C. North, "Visualization Viewpoints: Toward Measuring Visualization Insight," *IEEE Computer Graphics and Applications,* vol. 26, no. 3, pp. 6-9, May/June 2006.

[17] N.F. Polys, S. Kim, and D.A. Bowman, "Effects of Information Layout, Screen Size, and Field of View on User Performance in Information-Rich Virtual Environments," *Proc. ACM Symp. Virtual Reality Software and Technology,* 2005.

[18] D. Raja, D. Bowman, J. Lucas, and C. North, "Exploring the Benefits of Immersion in Abstract Information Visualization," *Proc. Immersive Projection Technology Workshop,* 2004.

[19] M.V. Sanchez-Vives and M. Slater, "From Presence to Consciousness through Virtual Reality," *Nature Revs. Neuroscience,* vol. 6, no. 4, pp. 332-339, 2005.

[20] P. Saraiya, C. North, and K. Duca, "An Insight-Based Longitudinal Study of Visual Analytics," *IEEE Trans. Visualization and Computer Graphics,* vol. 12, no. 6, Nov./Dec. 2006.

[21] P. Saraiya, C. North, and K. Duca, "An Insight-Based Methodology for Evaluating Bioinformatics Visualizations," *IEEE Trans. Visualization and Computer Graphics,* vol. 11, no. 4, pp. 443-456, July/Aug. 2005.

[22] J.P. Schulze, *DeskVOX,* http://deskvox.sourceforget.net, 2007.

[23] J.P. Schulze, A.S. Forsberg, and M. Slater, "Analysis of Subject Behavior in a Virtual Reality User Study," *Proc. Eighth Ann. Int'l Workshop Presence (Presence '05),* Slater M, ed., pp. 255-260, 2005.

[24] A. Sutcliffe, B. Gault, and J.E. Shin, "Presence, Memory and Interaction in Virtual Environments," *Int'l J. Human-Computer Studies,* vol. 62, no. 3, pp. 307-327, 2005.

[25] D.S. Tan, D. Gergle, P. Scupelli, and R. Pausch, "Physically Large Displays Improve Performance on Spatial Tasks," *ACM Trans. Computer-Human Interaction,* vol. 13, no. 1, pp. 71-99, Mar. 2006.

[26] M. Usoh et al., "Walking > Walking-in-Place > Flying, in Virtual Environments," *Proc. SIGGRAPH,* pp. 359-364, 1999.
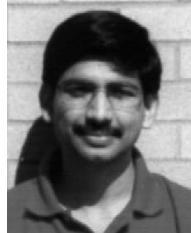
Prabhat received the BTech degree in computer science and engineering from IIT, Delhi, and the ScM degree in computer science from Brown University. He is a research staff member at the Center for Computation and Visualization, Brown University. His research interests include virtual reality, scientific visualization, parallel rendering systems, and machine learning.

Andrew Forsberg received a master's degree in computer science from Brown University. He is a research staff member in the Graphics Group, Brown University. His research interests include 2D and 3D user interfaces, scientific visualization, and VR applications.

Michael Katzourin is an undergraduate student in the Department of Computer Science, Brown University. His research interests include virtual reality and human computer interaction techniques.

Kristi Wharton is a faculty member in the Department of Molecular Biology, Cell Biology and Biochemistry, Brown University.

Mel Slater is an ICREA research professor at the Universitat Politècnica de Catalunya Centre de Realitat Virtual, Barcelona.

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