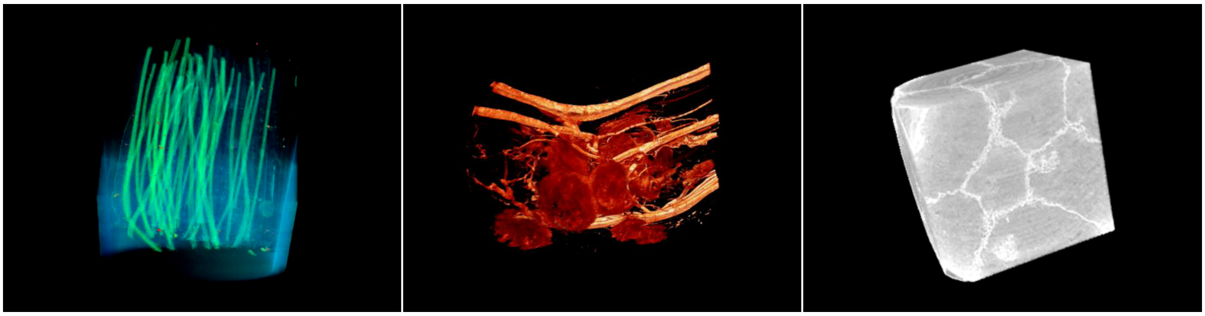
Effects of Immersion on Visual Analysis of Volume Data

Bireswar Laha, Kriti Sensharma, James D. Schiffbauer, and Doug A. Bowman



(a) 3D Scaffold dataset (b) Mouse Limb dataset (c) Fossil dataset (Parapandorina)

Fig. 1. The three volume visualized microscopic computed tomography (micro-CT) datasets used in our study.

Abstract—Volume visualization has been widely used for decades for analyzing datasets ranging from 3D medical images to seismic data to paleontological data. Many have proposed using immersive virtual reality (VR) systems to view volume visualizations, and there is anecdotal evidence of the benefits of VR for this purpose. However, there has been very little empirical research exploring the effects of higher levels of immersion for volume visualization, and it is not known how various components of immersion influence the effectiveness of visualization in VR. We conducted a controlled experiment in which we studied the independent and combined effects of three components of immersion (head tracking, field of regard, and stereoscopic rendering) on the effectiveness of visualization tasks with two x-ray microscopic computed tomography datasets. We report significant benefits of analyzing volume data in an environment involving those components of immersion. We find that the benefits do not necessarily require all three components simultaneously, and that the components have variable influence on different task categories. The results of our study improve our understanding of the effects of immersion on perceived and actual task performance, and provide guidance on the choice of display systems to designers seeking to maximize the effectiveness of volume visualization applications.

Index Terms—Immersion, micro-CT, data analysis, volume visualization, 3D visualization, CAVE, virtual environments, virtual reality.

1 Introduction

Visualization of volume data has been an important tool in a variety of application domains. Applications include medical imaging, biological visualization, geophysical exploration, solid modeling, and scientific visualization [1]. Volume data is obtained through processes such as computed tomography, magnetic resonance imaging, ultrasound, and confocal microscopy [1].

Volume data is composed of voxels on a 3D grid instead of 2D pixels. Each voxel has a numeric value based on some property of the object it is representing. These properties may include color, density, refractive index, or other material properties. There are many techniques for voxel rendering such as decomposition, isosurface rendering, maximum intensity projection, semi-transparency, and x-ray rendering [2]. For effective analysis of a 3D volume, users often need to look at the rendering from various viewpoints. Understanding the dataset requires users to mentally integrate the various views to construct a 3D mental model of the volume.

- *Bireswar Laha is with the Center for Human-Computer Interaction and the Department of Computer Science, Virginia Tech, E-mail: blaha@vt.edu.*
- *Kriti Sensharma is with the Electrical and Computer Engineering Department, Virginia Tech, E-mail: kritisen@vt.edu.*
- *James D. Schiffbauer is with the Institute for Critical Technology and Applied Science and Department of Geosciences, Virginia Tech, E-mail: jdschiff@vt.edu.*
- *Doug A. Bowman is with the Center for Human-Computer Interaction and the Department of Computer Science, Virginia Tech, E-mail: dbowman@vt.edu.*

Manuscript received 15 September 2011; accepted 3 January 2012; posted online 4 March 2012; mailed on 27 February 2012.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

A great deal of the existing research on volume visualization focuses on algorithms for offline and real-time rendering. Research on these algorithms is fast-evolving due to the advent of graphics processing units (GPUs). However, rendering performance alone is not sufficient to ensure that users understand volume visualizations.

Traditionally, researchers look at 3D volume renderings on desktop computer monitors or larger displays, which lack the more realistic visual stimuli found in more advanced displays, such as stereoscopic 3D graphics and motion parallax due to head tracking. Thus, it may be more difficult for the user to judge the size, shape, or depth of the structures in the volume data. Judgments of this sort are required not only for general understanding, but also for performing interactive tasks such as segmentation of a dataset. Conducting such tasks with traditional displays, therefore, may result in slower performance and/or erroneous interpretation.

Advanced displays with greater realism can be said to have a higher level of *immersion*, defined as the objective level of sensory fidelity provided by a system [3]. "Immersive" virtual reality (VR) systems produce sensory stimuli closer to those we experience in the real world. These systems may include fishtank VR [4], head-mounted displays (HMDs), and surround-screen projection systems like the CAVE [5]. The level of immersion is determined by a number of components. For example, components of visual immersion include field of regard (FOR; the total size of the visual field in degrees of visual angle surrounding the user), field of view (FOV; the size of the visual field in degrees of visual angle that can be viewed instantaneously), display size, display resolution, stereoscopy, head-based rendering (the display of images based on the physical position and orientation of the user's head produced by head tracking), realism of lighting, frame rate, and refresh rate [6].

We hypothesize that the effectiveness of volume visualizations could be improved by using immersive displays. But we do not know

1077-2626/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

whether it is worthwhile to use such displays given their cost and complexity. Thus, we need empirical evaluation of the effects of immersion for volume visualization. We wish to determine if the analysis of volume data could be performed faster, more accurately, with less perceived difficulty, or with more confidence with higher levels of immersion. Furthermore, it would be helpful to know which components of immersion are responsible for the effects that we find.

We conducted a controlled empirical study in which we studied the independent and combined effects of three components of immersion (head tracking, FOR, and stereoscopic rendering) on the analysis of two different volume-visualized microscopic computed tomography (micro-CT) datasets. We found significant positive effects of these components of immersion, but the results also show that the effects are dependent on how the components are combined and on the type of analysis task. Based on the results, we provide guidelines for designers of immersive systems seeking to maximize the effectiveness of volume visualization applications.

2 Background

Prior research has shown benefits of higher levels of immersion for analyzing 3D data in immersive virtual environments. A successful application of virtual reality in oil exploration and the petroleum industry is reported by Midttun [7], who claims measurable improvements to work processes and oil recovery by using an immersive VR system. Arns et al. [8] reported benefits of using an immersive CAVE display as compared to a traditional desktop display for performing structure detection tasks in a statistical visualization. Gruchalla's empirical study [9] on a well-path editing task also showed significant performance improvements in a CAVE as compared to a desktop monitor. Schuchardt et al. [10] reported improvements in performance and accuracy at a higher level of immersion when performing spatially complex and detailed search tasks in a visualization of underground cave structures.

The primary drawback of such studies is they analyze the overall effects of immersion by comparing immersive vs. non-immersive systems, but do not evaluate the effects of individual components of immersion or control for other factors. So they fail to determine which components of immersion resulted in the benefits, and do not evaluate whether a system with a moderate level of immersion might be sufficient to obtain the benefits. To address such questions, researchers use controlled experiments. Displays with a high level of immersion, such as a CAVE, can be used to produce the conditions needed for such controlled studies [11]. Using experiments of this sort, researchers have shown effects of specific components of immersion for search and comparison tasks [12, 13], single-user object manipulation [14], path tracing tasks [15], and graph visualization [16].

Immersive VR has been used successfully for visualizing volume datasets [17]. Kitamura et al. [18] reports on design of a real-time stereoscopic display system for collaborative work with volume rendered MRI and CT scan images. Zhang et al. [19] performed some of the first experiments exploring the benefits of volume visualization in a CAVE. Based on studies of visualizations of diffusion tensor magnetic resonance imaging (DT-MRI) datasets for brain tumor surgery they reported that users could interpret the data better in a CAVE than with a desktop display. In another experiment with DT-MRI datasets using streamtubes and streamsurfaces [20], Zhang et al. reported that the stereoscopy and interactivity of the VR system aided in understanding complex geometric models.

More recently, Demiralp et al. [21] reported the benefits of fishtank VR over CAVEs for certain kinds of tasks. Prabhat et al. [22] compared three display systems (desktop, fishtank VR, and CAVE) and reported significant benefits of the more immersive environments for analyzing volume data. However, we are not aware of any controlled experiments evaluating the effects of individual components of immersion on the effectiveness of volume visualizations. Our study aims to fill that gap.

3 Experiment

We designed a controlled experiment to evaluate the independent and combined effects of three components of immersion on user performance and preference with volume visualizations.

3.1 Goals and Hypotheses

Broadly, our goal is to understand the influence of the level of immersion on the effectiveness of visual analysis of volume visualizations. This leads us to our first research question:

1. *Are there any benefits of higher levels of immersion for understanding visualizations of volume data?*

Besides the general benefits of higher levels of immersion, we want to find the effects of individual components of immersion [6] on visual analysis of volume datasets, which gives us our next question:

2. *What are the effects of components of immersion on understanding visualizations of volume data?* In this study, we chose to explore the effects of three specific components of immersion on analyzing volume data.

Our third research question acknowledges that not all tasks are likely to be affected in the same way by the level of immersion, as prior studies have shown [10]:

3. *How are different visual analysis tasks affected by components of immersion?* To address this question, we designed a range of both quantitative and qualitative tasks for our study.

Finally, we were interested not only in task performance but also user preference for different levels of immersion when performing visual analysis tasks with volume datasets:

4. *When performing visual analysis of volume data, what level of immersion do users prefer?* To answer this question, we recorded and evaluated subjective ratings of the users' perceived difficulty and comfort levels at different levels of immersion.

In response to these research questions, we hypothesized the following:

1. *There are significant positive effects of higher levels of immersion for analyzing volume data*. Prior research has shown these benefits partially already [19, 20, 22].
2. *There will be significant benefits of field of regard, stereo, and head tracking*. We expect that the combination of high levels of all three will lead to the best understanding. High FOR, stereo, and head tracking together allow the user to perceive a dataset as floating in physical space, and to obtain different views of the dataset by physically walking around it.
3. *Higher levels of immersion will have significant benefits for complex spatial tasks, but not for others.* The extra spatial and proprioceptive cues offered by higher levels of immersion could benefit spatial analysis tasks involving detailed spatial judgments. However, lower levels of immersion are likely to be sufficient for simpler tasks.
4. *Users will prefer higher levels of immersion*. Higher immersion levels are similar to the real world, and we believe users will perceive lower difficulty and have more confidence in those conditions because of their familiarity.

3.2 Datasets

Micro computed tomography (micro-CT) allows 3D internal imaging of objects at the microscopic (10^{-6}) level of scale, and is an invaluable tool in various disciplines, such as tissue engineering, oncology, and geology. Micro-CT researchers visualize reconstructed volume datasets on their desktop displays, which are non-stereo, non-head-tracked and in most cases about 19" in diameter. Good visualization is essential for an accurate analysis of the dataset; consequently, there is an interest in the micro-CT community to visualize volume datasets in immersive VR systems like the CAVE.

We identified three micro-CT datasets for our study. The first was a 3D Scaffold dataset (Fig 1-a) used in bone regeneration studies

[23]. The scaffold mimics the structure of a cortical bone and contains bundles of poly-L-lactide fibers on polyglycolide cores. The individual bundles mimic the osteon, a structural unit of the bone. We used this dataset for training participants.

The second dataset was a mouse limb [24], imaged at the major knee joint of the mouse (Fig 1-b). The visualization also showed the major blood vessels, the soft tissues, and the surrounding musculature in that part of the mouse.

The third dataset was a fossil (Fig 1-c), dated to 600 million years ago, known as *Parapandorina raphospissa*. This fossil has been interpreted as a potential early animal embryo from the Doushantuo phosphorites of South China [25]. The visualization that participants viewed was of an incomplete fractured specimen.

3.3 Apparatus

We relied on existing open-source software for volume rendering of the micro-CT datasets. Our choice of hardware platform was guided by the ability to perform controlled experimentation [11].

3.3.1 Hardware

We used a Viscube display featuring three 10′ by 10′ rear-projected stereo walls and a top-projected stereo floor each running at 1920×1920 resolution. A wireless Intersense IS-900 tracking system tracked the user's head (in the conditions with head tracking on) and a wand device with a joystick and five buttons. An Infitec system provided passive stereoscopic 3D graphics in the conditions with stereo on.

3.3.2 Software framework

Our software framework for distributed volume rendering consisted of DIVERSE [26] at the bottom layer for clustering and distributed rendering, 3D Visualizer [27] for volume rendering of the micro-CT datasets, and VRUI [28] for 3D interaction with the datasets. We slightly modified the interaction techniques in VRUI to make them more suitable for our study, as described below.

3.3.3 User interface

To translate the viewpoint, the user could press the joystick forward to travel in the direction the wand was pointing, or press the joystick backward to travel in the opposite direction. Pressing the joystick to the left or right would cause the dataset to rotate about an axis perpendicular to the plane of the wand.

The user could also grab the dataset by holding down a button on the wand, after which the user's hand could be used to directly manipulate the position and orientation of the dataset. Another button press activated a cutting plane feature, which allowed the user to use hand movements to slice the dataset along any arbitrary 3D plane, revealing inner features of the volume data.

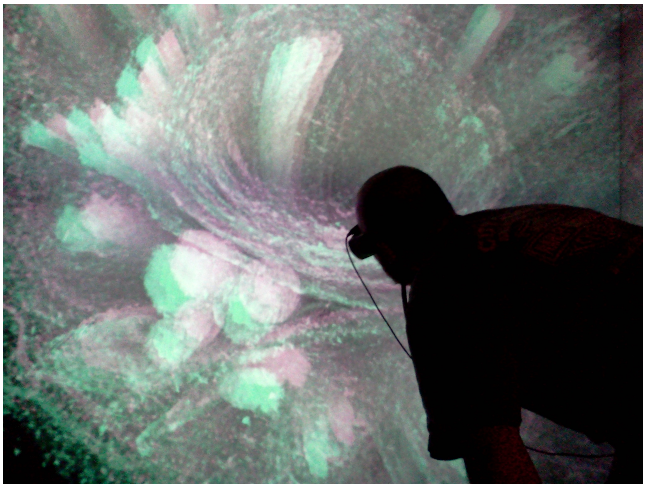


Fig. 2. A participant experiencing the training dataset in the Viscube.

In addition to these interactions, in the conditions with head tracking on, the display of volume data in the Viscube was based on the physical position and orientation of the user's head. Fig 2 shows a head-tracked user examining the training dataset in the Viscube.

3.4 Tasks

Tasks for each dataset were either quantitative or qualitative in nature. Quantitative tasks required the participants to count particular features of the dataset. Qualitative tasks required the user to describe particular characteristics of the dataset using their own words. In both task types, participants gave their answers verbally, and the experimenter recorded the responses of the participants on paper.

The primary requirement was that the tasks under study should be of actual importance to the researchers working with the dataset, in order to ensure that any benefits of immersion we found would have real-world relevance. We thus collaborated with two researchers who have had several years working experience with the two micro-CT datasets chosen for our study. Together, we came up with a list of preliminary tasks, and refined the lists after performing the tasks ourselves and running pilot studies with five participants.

During the design of the tasks, we kept in mind that our participants might be novices (see section 3.7). We developed tasks that were important to experts but at the same time were not highly technical, ensuring that they could be performed without prior knowledge of micro-CT. The only background assumed was a very basic understanding of what blood vessels, cells, bones, and other such simple biological structures look like.

For the training dataset, we designed three quantitative and three qualitative tasks. We also ended up with six tasks for the mouse limb dataset, of which three were quantitative. The fossil dataset had 12 tasks, of which three were quantitative. Based on the results of an exploratory study (see section 3.5), we narrowed our list to four tasks for the mouse limb (three quantitative) and six for the fossil (three quantitative), which we studied in our main controlled experiment. The final task lists can be seen in the appendix.

The specific tasks designed for each dataset in our study were different, because the tasks for one dataset wouldn't make any sense for the other dataset. However, the tasks came from the same abstract task categories for both datasets, and were aimed at creating generalizable results (see Table 2, section 5.1 and 5.2).

The experts on the mouse limb and the fossil datasets also led the creation of a standardized scoring rubric for grading the various tasks, ensuring uniformity in grading. The rubric laid down the criteria for specific levels of correctness of the responses, and specified grades ranging from 0.0 to 1.0.

3.5 Exploratory Study

We conducted an exploratory study for the purpose of evaluating whether the level of immersion affected task performance with our two micro-CT datasets, and to refine our task lists, procedures, and user interface. The study had 12 volunteer participants, half of whom had considerable experience analyzing micro-CT datasets. The two datasets ([23-25]) and their tasks, always presented in the same order, were evaluated with two levels of immersion, varied within subjects. The two conditions we tested were: (1) the standard CAVE with four screens, stereo, and head tracking as the "high immersion" condition, and (2) a single wall of the CAVE with no stereo and no head tracking as the "low immersion" condition.

Aside from recording the responses of the participants to each task, we also measured the time taken by the participants for each task, and recorded their responses on a 1-7 Likert scale for the task difficulty level and their subjective level of confidence in their responses for every task. Responses to qualitative and quantitative tasks were graded offline by the experts on our team.

This short study identified positive trends of benefits of immersion on performance for certain tasks we picked for closer study in our main controlled study, explained in the following sections. The tasks that were left out are given in the appendix. We left out these tasks as the results showed almost no difference between the high and low levels of immersion condition for these tasks.

3.6 Design

In this controlled experiment, we chose to study the effects of three components of immersion, keeping other factors constant. We picked stereo (ST), field of regard (FOR), and head-based rendering or head tracking (HT) as the three independent variables in our study. ST had levels of 'on' (stereoscopic) and 'off' (monoscopic). FOR had levels 'high' (with all the four screens of Viscube displaying content) and 'low' (with only the front screen displaying content). HT also had levels 'on' and 'off', which are self-explanatory. The hardware platform we chose helped us to eliminate several confounds in our study. For all conditions, the system produced the same frame rate, refresh rate, resolution, and realism of lighting. Participants wore the passive stereo goggles in all conditions that they experienced, which kept the amount of light entering their eyes the same, and also kept the field of view constant.

Table 1. Conditions Experienced by the Eight Groups in the Experiment

Group No.	First Condition (Mouse-Limb)			Second Condition (Fossil)		
	FOR	ST	HT	FOR	ST	HT
1	High	On	On	High	On	Off
2	High	On	Off	High	On	On
3	High	Off	On	High	Off	Off
4	High	Off	Off	High	Off	On
5	Low	On	On	Low	On	Off
6	Low	On	Off	Low	On	On
7	Low	Off	On	Low	Off	Off
8	Low	Off	Off	Low	Off	On

With two levels of each of the three factors, we had eight different conditions in our study. HT was a within-subjects variable while FOR and ST were between-subjects variables; thus, each participant experienced two of the eight conditions, one with HT on and one with HT off. It was not practical to vary all three components within subjects in our study, because eight conditions per subject would have required eight datasets and would have been overly long. We chose to vary HT within subjects because we wanted to study whether individuals used different strategies to explore the datasets with and without HT (see section 4). All participants first performed tasks with the Mouse Limb dataset. Half of the participants had HT on for the Mouse Limb tasks, while the other half had HT off (Table 1). HT was not confounded with dataset; rather, each participant simply experienced one level of HT with each dataset. The effects of HT were determined by comparing the results of different participants (those with HT on and those with HT off) within the same dataset.

The dependent variables in our study were the amount of time taken for each task and the responses of the participants to each task (that were recorded and graded offline by two experts using the grading rubric). We also recorded participants' responses for the difficulty levels of each task, and their subjective levels of confidence in their answers for each task, both on 1-7 Likert scales.

3.7 Participants

We recruited 47 voluntary participants for our study, all of whom self reported to have no prior experience in analyzing micro-CT datasets. Five of them were pilot subjects. Based on the results of a spatial ability test [29], we dismissed two participants who had low negative scores, resulting in a final list of forty participants for our main study. There were 19 females and 21 males. The average age was 21.45 years, ranging from 18 to 30 years.

3.8 Procedure

Prior to beginning the study, participants were given a standard informed consent form (the study was approved by the university's Institutional Review Board). Participants then filled out a background questionnaire that captured information on their demographics and experience with using computers, watching 3D stereo movies, 3D virtual environments, and analyzing CT and micro-CT datasets. Then they took a spatial ability test [29]. Based on the results of the test, they were assigned to one of eight groups, such that the average spatial abilities of the eight groups were approximately the same. We then introduced the participants to the hardware and explained the interactions that they could perform with the system.

Domain scientists or experts will typically have some specialized strategies for interaction with volume datasets, which are informed by their experience. Keeping that in mind, we devised the various tasks with the training dataset (Fig 1-a) such that the novice participants were trained in the different strategies used by experts for analyzing the two main micro-CT datasets (Fig 1-b and c) that we were studying. The participants performed five tasks with the training dataset using the same condition with which they would then experience the mouse limb dataset. The training lasted around 15–20 minutes. The training tasks were both quantitative and qualitative, and they trained the participants to understand a micro-CT dataset in 3D and also answer questions related to counting a particular feature, searching for hidden structures in 3D, comparing and commenting on similar structures in 3D, analyzing a micro-CT dataset as a whole and in parts, and slicing the dataset to comment on structures inside a 3D volume. In addition, participants learned how to use the various interactions. The participants were also given three rotation tasks about three orthogonal axes, to make sure they could rotate the volume data in any direction they wanted without any assistance.

After the training, participants took a short break, after which they started working with the mouse limb dataset (Fig 1-b). We asked participants to be as accurate as possible in their responses, and told them that there was a time limit for each task. They completed four tasks with the mouse limb dataset (see Appendix) at the selected level of immersion. The experimenter recorded responses, task times, difficulty level, and confidence level for each task.

After another short break, participants performed the same training as before, but in the second assigned condition, in which they would experience the fossil dataset (Fig 1-c). They performed seven tasks with the fossil dataset (see Appendix); the experimenter recorded their responses.

During task performance with the main datasets, if the participants digressed too much from the expected strategy for that particular task, we guided them towards the correct expert strategies (for a list of the main strategies for each task identified by the domain experts, please see the appendix). In this way, we tried to emulate expert strategies as closely as possible.

Finally, we asked participants to complete a post-questionnaire. It captured their opinions for both the head-tracked and non-head-tracked conditions on 1-7 Likert scales for: comfort level, ease of getting the desired view and exploring the dataset in general, and ease of understanding the features of a dataset and doing different tasks with the dataset. For both levels of HT, participants also rated the effectiveness of three visual analysis strategies: changing the viewpoint by rotating or grabbing the dataset with the wand, slicing the dataset with the wand, and physically walking around the dataset to look from different viewpoint.

The datasets in each condition were rendered at the same initial position and orientation in front of the participants, but they then were allowed to interact with the datasets based on their own strategies. Each question was read out loud to the participants, using consistent wording. After the maximum time limit for a task was reached, the participant was stopped and asked for their final response for the task.

4 RESULTS

In this section, we present the statistically significant results of our study. Except for the measured time, which was a numeric continuous variable, all the other dependent variables in our study were numeric ordinal values. To understand the main effects and the two and three-factor interaction effects of our three independent variables (FOR, ST, HT), we ran a three-way analysis of variance (ANOVA) on the values of the time metric, and an Ordinal Logistic Regression based on a Chi-square statistic on all other metrics.

For significant interactions between the components, we evaluated which combinations were significantly different from which of the others using Student's t test for the time metric. For all other metrics, which had numeric ordinal responses, we employed the two-sided Wilcoxon Signed-Rank Tests for post-hoc analyses.

Table 2. Relative Weights of the Tasks in each Dataset

Mouse limb task#	Task Type	Weights	Fossil Task#	Task Type	Weights
M1	Simple search	0.25	F1	General description	0.15
M2	General description	0.15	F2	Internal feature search	0.25
M3	Visually complex search	0.3	F3	General description	0.05
M4	Spatially complex search	0.3	F4	Visually complex search	0.25
			F5	Visually complex search	0.1
			F6	General description	0.15
			F7	Simple search	0.05

To understand the overall effects on the tasks with a particular dataset (mouse limb and fossil), the dataset experts on our team assigned weights to the tasks, based on the perceived relative importance of the tasks to their research group (see Table 2). The assigned weights for each dataset added up to one. We represent these weighted totals as ∑M and ∑F for the mouse limb and fossil datasets respectively, for every metric in our study. These are the weighted averages of the scores or values in each metric. We will also use the task numbers as defined in the appendix for explaining the results here (e.g., 'M1' denotes the first task with the mouse limb dataset, and 'F4' denotes the fourth task with the fossil dataset). We also classified the tasks into several abstract task categories, which are shown in table 2 and discussed in section 5.1.

4.1 Grades (accuracy in task performance)

We found several significant main effects of the different components of immersion on the grades received by the participants (see Table 3). All of the main effects indicated better grades for higher levels of immersion, except for two tasks in which ST off was significantly better than ST on.

Table 3. Significant Main Effects on Grades

Task: source	χ^2	p-value	Note (higher grade is better)
M1: FOR	4.936	0.026	High FOR better
M3: FOR	9.069	0.003	High FOR better
∑M: FOR	7.493	0.006	High FOR better
F1: FOR	8.422	0.004	High FOR better
F4: FOR	5.217	0.022	High FOR better
∑F: FOR	9.295	0.002	High FOR better
M1: HT	5.348	0.021	HT on better
M4: HT	6.215	0.013	HT on better
∑M: HT	4.792	0.029	HT on better
M4: ST	22.746	< 0.0001	ST on better
∑M: ST	4.594	0.032	ST on better
F1: ST	7.271	0.007	ST off better
F2: ST	5.900	0.015	ST off better

We also found several significant interaction effects of the different components of immersion on the grades received by the participants (see Table 4). These interactions are plotted in Figures 3–6).

Table 4. Significant Interaction Effects on Grades

Task: source	χ^2	p-value	Mean values in descending order (higher is better)	
M2: FOR & ST	4.444	0.035	FOR high ST off	0.74
			FOR low ST on	0.74
			FOR low ST off	0.67
			FOR high ST on	0.56
M4: FOR & HT	7.234	0.007	FOR high HT on	0.96
			FOR low HT on	0.89
			FOR low HT off	0.88
			FOR high HT off	0.67
F5: FOR & HT	6.422	0.011	FOR low HT off	0.675
			FOR high HT off	0.67
			FOR high HT on	0.4
			FOR low HT on	0.375
F1: ST & HT	4.920	0.027	ST off HT on	0.75
			ST off HT off	0.59
			ST on HT off	0.58
			ST on HT on	0.42

Post-hoc tests indicate that for M4 grade, FOR high HT on (p = 0.031) is significantly better than all other conditions and FOR high HT off (p = 0.016) is significantly worse than all other conditions. For F5 grade, both FOR low HT off (p = 0.031) and FOR high HT off (p = 0.025) are significantly better than both FOR high HT on (p = 0.047) and FOR low HT on (p = 0.025). Also for F1 grade, ST off HT on is significantly better (p = 0.012) than ST on HT on.

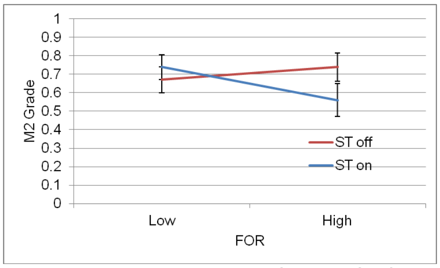


Fig. 3. Interaction between FOR and ST for M2 grade.

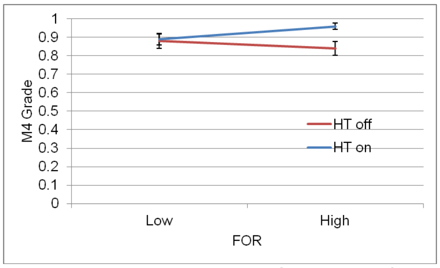


Fig. 4. Interaction between FOR and HT for M4 grade.

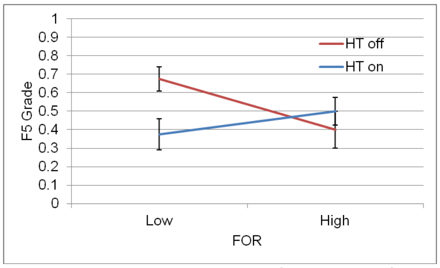


Fig. 5. Interaction between FOR and HT for F5 grade.

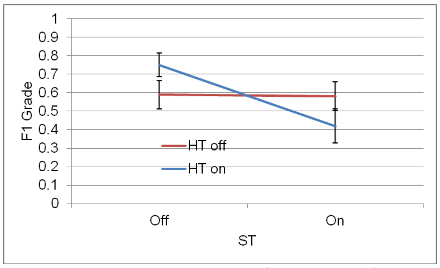


Fig. 6. Interaction between ST and HT for F1 grades.

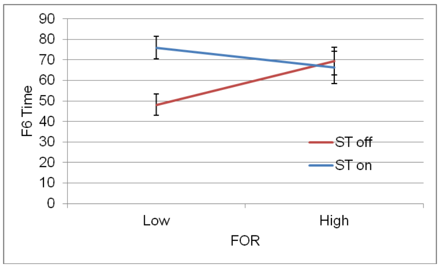


Fig. 8. Interaction between FOR and ST for F6 time (in sec).

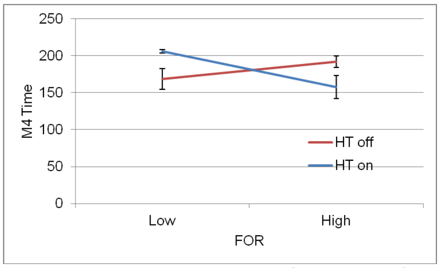


Fig. 10. Interaction between FOR and HT for M4 time (in sec).

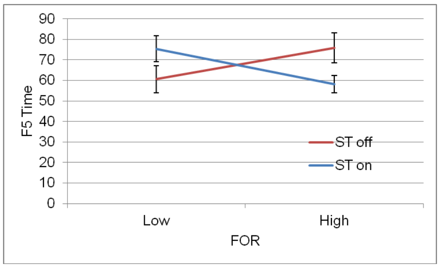


Fig. 7. Interaction between FOR and ST for F5 time (in sec).

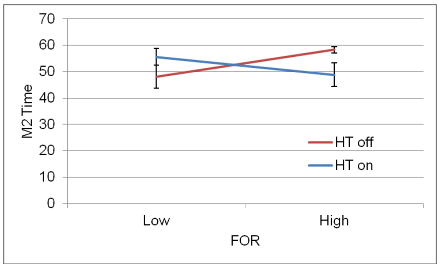


Fig. 9. Interaction between FOR and HT for M2 time (in sec).

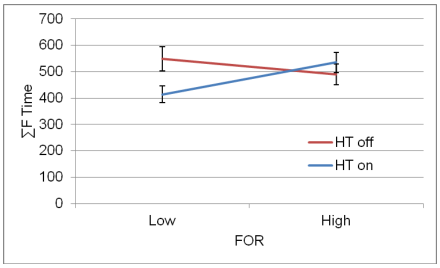


Fig. 11. Interaction between FOR and HT for ∑F time (in sec).

4.2 Task completion time

We found a significant main effect of ST on time for M4 (F Ratio = 4.8280, p = 0.0354), with ST on significantly slower.

We found several significant interaction effects (see Table 5) of the different components of immersion on the time taken by the participants in F5 (Fig 7), F6 (Fig 8), M2 (Fig 9), M4 (Fig 10), ∑M, and ∑F (Fig 11).

Table 5. Significant Interaction Effects on Time

Task: source	F-Ratio	p-value	Least square means in ascending order (lower is better)	
F5: FOR & ST	6.592	0.015	FOR high ST on	58.12
			FOR low ST off	60.62
			FOR low ST on	75.4
			FOR high ST off	75.93
F6: FOR & ST	6.019	0.02	FOR low ST off	48.15
			FOR high ST on	66.35
			FOR high ST off	69.42
			FOR low ST on	75.92
M2: FOR & HT	5.735	0.023	FOR low HT off	48.15
			FOR high HT on	48.85
			FOR high HT off	55.6
			FOR low HT on	58.29
M4: FOR & HT	11.651	0.002	FOR high HT on	157.76
			FOR low HT off	168.77
			FOR high HT off	192.05
			FOR low HT on	206.28
∑M: FOR & HT	8.793	0.006	FOR high HT on	333.62
			FOR low HT off	339.34
			FOR high HT off	386.93
			FOR low HT on	402.32
∑F: FOR & HT	5.487	0.026	FOR low HT on	414.34
			FOR high HT off	490.06
			FOR high HT on	535.43
			FOR low HT off	548.47

Post hoc tests indicate that for F6 the FOR low ST off condition was significantly better than the FOR low ST on condition; for M4 the FOR high HT on condition was significantly better than the FOR high HT off and FOR low HT on conditions, and the FOR low HT on condition was significantly worse than the FOR low HT off and FOR high HT on conditions. For ∑M, the FOR high HT on and FOR low HT off were significantly better than the FOR low HT on condition. Interestingly in ∑F, the FOR low HT on condition was significantly better than the FOR high HT on and FOR low HT off conditions.

4.3 Perceived difficulty (subjective rating)

We found significant main effects of FOR and ST on difficulty perceived by the participants for different tasks (see Table 6).

Table 6. Significant Main Effects on Difficulty

Task: source	χ^2 statistic	p-value	Note (lower difficulty is better)
F3: ST	6.895	0.009	ST on better
F4: ST	5.903	0.015	ST on better
F5: ST	8.034	0.005	ST on better
F7: ST	5.945	0.015	ST on better
M1: FOR	4.733	0.03	High FOR better
F5: FOR	4.175	0.041	Low FOR better

We also found several significant interaction effects of the different components of immersion on the difficulty levels perceived by the participants (see Table 7).

Table 7. Significant Interaction Effects on Difficulty

Task: source	χ^2	p-value	Mean values in ascending order (lower is better)	
F6: FOR & ST	6.335	0.012	FOR high ST on	2.5
			FOR low ST off	2.8
			FOR low ST on	3.2
			FOR high ST off	3.8
M4: FOR & HT	8.607	0.003	FOR high HT on	4.7
			FOR low HT off	4.8
			FOR high HT off	5.6
			FOR low HT on	6
F4: FOR & HT	7.088	0.008	FOR low HT on	3.4
			FOR high HT off	4.5
			FOR low HT off	4.9
			FOR high HT on	5.1
F1: ST & HT	5.946	0.015	ST off HT off	1.2
			ST on HT on	1.5
			ST on HT off	2
			ST off HT on	2.2
M2: FOR, ST &	5.663	0.017	FOR high ST on HT on	1.2
			FOR low ST on HT off	2
			FOR high ST off HT on	2.4

HT			FOR high ST off HT off	2.4
			FOR high ST on HT off	3.2
			FOR low ST off HT on	3.2
			FOR low ST on HT on	3.4
			FOR low ST off HT off	3.6
F6: FOR, ST & HT	5.845	0.016	FOR high ST on HT off	1.6
			FOR low ST on HT on	2.6
			FOR low ST off HT on	2.8
			FOR low ST off HT off	2.8
			FOR high ST on HT on	3.4
			FOR high ST on HT off	3.4
			FOR low ST on HT off	3.8
			FOR high ST off HT off	4.2

Post-hoc tests indicate that for M4, the FOR low HT off condition was significantly better than FOR low HT on condition ($p = 0.039$). For F4, the FOR low HT on condition was significantly better than the FOR low HT off condition ($p = 0.039$), and the FOR low HT on condition was significantly better than FOR high HT on ($p = 0.031$). Also for F1 the ST off HT off was significantly better than the ST off HT on condition ($p = 0.031$).

4.4 Confidence in response (subjective rating)

We found significant main effects of FOR and HT on confidence levels of the participants in different tasks (see Table 8).

Table 8. Significant Main Effects on Confidence

Task: source	χ^2	p-value	Note (higher confidence is better)
M1: FOR	5.768	0.016	High FOR better
F3: FOR	4.945	0.026	High FOR better
F3: HT	7.230	0.007	HT on better
F5: HT	6.454	0.011	HT on better

We found some significant interaction effects of the different components of immersion on the confidence levels (see Table 9).

Table 9. Significant Interaction Effects on Confidence

Task: source	χ^2	p-value	Mean values in descending order (higher is better)	
F3: FOR & ST	6.029	0.014	FOR high ST off	5.2
			FOR low ST on	5.1
			FOR high ST on	5
			FOR low ST off	4.2
F4: FOR & HT	4.965	0.026	FOR low HT on	5.2
			FOR high HT off	2
			FOR high HT on	3.9
			FOR low HT off	3.3

Post-hoc tests indicate that for F4, the FOR low HT on condition was significantly better than the FOR low HT off condition ($p = 0.004$).

4.5 Post-questionnaire results

The subjective ratings of the participants in the post-questionnaire also produced several interesting significant results.

We found some significant interaction effects of the different components of immersion on the comfort levels of the participants. Post-hoc tests indicated that for comfort, both the FOR high ST off condition ($p = 0.031$) and the FOR low ST on condition ($p = 0.025$) were significantly better than the FOR low ST off condition.

We found a significant main effect of HT ($\chi^2 = 3.854$, $p = 0.0496$) on the participant's ease of obtaining the desired view and exploring the datasets in general, with ease higher when HT was on. There were also significant interactions of different components of immersion on this metric. Post-hoc tests indicated that the ST off HT on condition was significantly better than the ST off HT off condition ($p = 0.002$).

Finally, we observed a significant main effect of ST for the participants' ease of understanding the features of a dataset ($\chi^2 = 4.405$, $p = 0.036$), with higher ease when ST was on.

4.6 Effects of spatial abilities of the participants

We ran pair-wise correlation analyses between spatial abilities of the participants and the different metrics in our study. We found no significant correlations.

4.7 Summary of important results

Looking across the different metrics from above, we summarize the main effects of the three components of immersion (FOR, HT, and ST) here.

High FOR improved grades for a variety of tasks. There was no significant effect on speed of performance due to FOR.

With HT on, grades improved in three cases, and also confidence levels were higher in two tasks. Participants felt that the ease of getting the required view and exploring the dataset in general was higher with head-based rendering. HT did not seem to affect the speed of performance.

With ST on, perceived difficulty was reduced for several tasks. Participants also felt that the ease of understanding the features of a dataset was higher with stereoscopic vision. Stereo produced mixed results for accuracy, with grades improving in two cases but decreasing in two others. Stereo also caused slower performance in one task.

There were also some interesting interaction effects. We observed significant interactions between FOR and HT in several cases, with FOR high / HT on and FOR low / HT off proving to be better than the other two combinations. There was also a significant interaction between FOR and ST for several tasks.

5 Discussion

We found that most of the effects of high immersion were positive. Grades for $\sum M$ (overall performance measure for the mouse limb dataset) improved significantly with high levels of FOR, ST and HT. Grades for $\sum F$ (overall performance measure for the fossil dataset) improved with high FOR.

Of the three components of immersion, FOR had the most positive effects on the widest range of tasks. FOR improved grades and boosted confidence in a variety of quantitative tasks (M1, M3, F4).

FOR also improved grades and confidence in certain qualitative tasks (F1, F3) where the participant needed an overall understanding of the entire dataset to give the correct response. In F1, the response depended on how the participant understood the 3D volume as a whole, and in F3 the participant had to examine the entire dataset to understand how the different parts compared to each other. The primary advantage of high FOR is that the user can move very close to the dataset yet still have the entire dataset visible in peripheral vision. With a single screen (low FOR), the dataset is clipped by the edges of the screen when the viewpoint is close.

Components of immersion had differential effects on different tasks. For example, in M4, in which the participant had to search for blood vessels from a clouded maze of floating materials of different densities, ST proved to have very significant benefits. The additional depth cue provided by ST aids in visually understanding convoluted structures, the boundaries of which otherwise would be very hard to determine from the vague outlines with a monoscopic display. But stereo also worsened performance in F1 and F2. Reasons for this may include eye strain, since F1 and F2 were tasks near the end of the experiment. F2 also required slicing to understand the internal structures of the fossil, and stereo rendering may have been confusing during this process as parts of the internal structures were appearing and disappearing.

We also observed interactions of stereo with FOR in several cases (M2, F5, F6, and F3) with the higher levels of the components improving grades, and reducing time and difficulty in most cases.

The similarity of stereoscopic vision to the real world might have contributed to the higher user ratings in post-questionnaire for ease of getting the view they wanted.

The interaction of HT and FOR is also very interesting. For tasks M2, M4, and F5, the combination of FOR and HT resulted in better grades or faster performance, if both were at higher levels (FOR high and HT on, which is most similar to the real world) or both were at lower levels (FOR low and HT off, which is closest to a traditional desktop display). This might be due to participants' level of familiarity with these two conditions.

5.1 Effects of immersion on task categories

Similar to the Schuchardt et al. study [10], we roughly categorized the different tasks in our experiment (see Table 2). The details of each task are given in the appendix.

The different task types are defined as follows:

- Simple search—quantitative search for features which are easy to understand spatially and visually distinct
- Visually complex search—quantitative search involving analysis of features that are visually indistinct or vague
- Spatially complex search—quantitative search involving analysis of features that are spatially crowded or dense
- Internal feature search—quantitative search for features inside a dataset, by slicing through the data
- General description—qualitative description of features in a dataset or the dataset as a whole

For the simple search tasks that we studied (M1 and F7), all significant effects of immersion were positive. High FOR improved grade, reduced difficulty and improved confidence on M1. ST on reduced the difficulty of F7, and HT on improved performance in M1. Thus, high immersion may not be required for simple search, but it can be beneficial.

For the visually complex search tasks, we found that high FOR improved grades in two out of three (M3 and F4), and ST on reduced the difficulty levels in two out of three (F4 and F5). HT on also improved confidence in F5. The interaction of FOR and HT had significant effects on F4, for difficulty and confidence levels, with participants perceiving the combination of low FOR and HT on as less difficult and producing more confidence. This task required users to look closely at very small intracellular bodies in the fossil dataset, so HT was helpful in obtaining the correct view, but high FOR may have been distracting. Overall, it appears that higher immersion can benefit visually complex search, but not in the same way for each task.

The only spatially complex search task (M4) involved counting the number of hidden blood vessels in a dense, twisting mass of vessels. ST on improved grades but reduced speed, probably because participants took more time to count once ST allowed them to spatially distinguish one blood vessel from the next. HT on also improved the grades in M4. The combination of high FOR and HT on positively affected grades, time and difficulty level for this task. As prior studies showed [10], spatially complex search tasks seem to benefit strongly from higher levels of immersion.

In the only internal feature search task that we had (F2), ST on actually degraded performance, perhaps because stereoscopy is difficult to understand during slicing. We will need to study more tasks like this one to understand the effects more clearly.

High FOR had positive effects on two out of four general description tasks—it improved grades in F1 and improved confidence in F3. Also, ST on reduced difficulty and HT on improved confidence in F3. The interaction effects of various components of immersion were mixed in this task category, however, which probably indicates the need for further study of more definitive and fine-grained groupings of such tasks.

5.2 Generalizability of the results

We ran our main experiment with novice participants (people with little or no experience with micro-CT datasets), but believe that the objective results we report can be generalized to the community of domain scientists who work with volume data at different scales, based on the generic task categories (see **Error! Reference source not found.**), although subjective measures may vary.

The field of micro-CT imaging is relatively new and most domain scientists have little experience in analyzing micro-CT datasets [30, 31]. In the field of palaeontology, for example, very few researchers presently use micro-CT techniques for analyzing fossils. This not only means that there are very few experts in this domain, but also that most of the domain scientists are novices or close to novices. Presently, a great deal of training and teaching activity is going on in this field for new domain scientists [32].

We have also attempted to classify our tasks into abstract categories (see Table 2), so that the results tied to those groups could be generalized further to other domains of volume data. We are currently refining this task taxonomy by interviewing domain scientists, in order to have more generalizable results for our future studies.

5.3 Implications for design

Our findings in this study were with volume datasets used in active research, and in a controlled environment. We claim that the results of our study can be generalized to cases where users are performing similar visual analysis tasks with volume data. Based on the findings in this study, we have the following recommendations for designers of volume visualization systems:

- We recommend systems with higher FOR, like the CAVE, for a wide variety of visual analysis tasks with volume data.
- Having head tracking along with high FOR might also benefit spatial search tasks, as this combination is familiar and allows users to physically walk around the dataset.
- Finally, for systems designed for tasks that involve analyzing spatially complex structures in a 3D volume, use stereoscopic rendering and/or head tracking.

6 Conclusions and Future work

Revisiting the research questions (section 3.1), we found that most of the benefits of high immersion for analyzing volume data are positive. Of the three components of immersion that we evaluated in our controlled study, FOR had the most positive effects on the widest range of tasks.

However, levels of immersion have differential effects on different tasks. We observed the most positive effects of immersion for tasks involving *visually and spatially complex search* for features in the volume datasets. General descriptive tasks showed mixed effects of immersion.

Users showed mixed preferences for the different levels of immersion. They felt that stereoscopic vision made it easier to understand features of a volume dataset, and head-based rendering provided greater ease of obtaining the desired view they wanted and of exploring the datasets in general.

Although we found evidence of high FOR improving task performance for analyzing volume data, we do not yet know what level of FOR is needed to provide these benefits. An interesting follow up study could vary FOR with three or more levels to examine these effects more closely. Future studies should also evaluate the effects of other components of immersion, like field of view, display size, and display resolution on similar tasks.

Finally, although we found that different search tasks benefit from higher levels of immersion, we do not know exactly what type of search tasks benefit the most, and why the other types of tasks (general descriptive tasks) show mixed results from immersion. We need further refinement of our task classifications and verification of our conclusions about the effects of immersion on these task types.

APPENDIX

Tasks with the mouse limb dataset:

M1. *You are allowed a maximum time of 1 minute for this task. How many soft tissues like these could you count in the dataset?* (**Simple search**) – A soft fluffy structure was shown near the surface, which was a soft tissue. The task was to visually search for similar structures in the dataset.
Strategy: Rotate the dataset completely about any axis.

M2. *For your next task, you are allowed a maximum time of 1 minute. This is a bone of the mouse limb. In your own words, please describe the inner core of the bone.* (**General description**) – A particular bone was shown. The task was to describe the structure of the bone marrow.
Strategy: Look at the bone marrow from different angles. Might also use the slice tool to look at various cross-sections.

M3. *For your next task, you are allowed a maximum time of 2 minutes. Here is an example of a distinct bone segment. Count the other distinct bone segments in the sample. If a bone branches into two or more parts, only count it once. Also describe the overall structure formed by the bone segments – what letter of the alphabet does the structure resemble?* (**Visually complex search**) – In the previous task, the participant worked with a bone. This task was to visually search for bone structures present in the entire dataset, and connect them together to form a letter of the alphabet.
Strategy: Rotate and also slice and look from various angles.

M4. *For your last task, you are allowed a maximum time of 3 minutes and 30 seconds. Now let's say, this is the top and this is the bottom of the structure. Please find and count the number of distinct blood vessels which are visible from the top and bottom, but cannot be seen from any side-view.* (**Spatially complex search**) – The task was to look separately from the top and bottom of the dataset and search for blood vessels present at each end. There were two intermingled blood vessels visible from the top and inside the dataset, and one visible at the bottom.
Strategy: Rotate to the top and bottom. Slice and look from different angles, and also look inside the dataset.

Tasks with the fossil dataset:

F1. *For your first task, you are allowed a maximum time of 1 minute. Describe in your own words the 3D structure that you see in front of you.* (**General description**) – The task was to describe the structure of the entire dataset.
Strategy: Look from different angles and describe.

F2. *For your next task, you are allowed a maximum time of 3 minutes. As you can see, the Parapandorina specimen consists of multiple bounded compartments, each of which is potentially a cell. You can see these compartments forming in 3D as you use the cutting plane to slice through the dataset in a single direction. Please count the number of compartments that you can identify.* (**Internal feature search**) – A cell structure was shown in 3D inside the fossil volume. The task was to count all the cells that the participant can identify in the entire volume.
Strategy: Slice through the dataset, spatially constructing the cells in different layers of the volume; look from different angles.

F3. *For your next task, you are allowed a maximum time of 1 minute 30 seconds. Describe the shape of the borders and the joints between the cells. Are the borders and joints between different cells comparable or different in shape?* (**General description**) – A border is the line separating two cells, and a joint is where more than two cells join. This task is about describing the shape of the borders and joints in the volume, comparing the borders to each other and joints to each other.
Strategy: Slice, rotate and look from different angles.

F4. *For your next task, you are allowed a maximum time of 3 minutes. Within each cell there may or may not be intracellular bodies. In how many of the cells can you identify intracellular bodies? How many per cell?* (**Visually complex search**) – An intracellular body was shown in 3D. The task was to scan through the entire volume and search for intracellular bodies in all the cells that the participant had identified.
Strategy: Slice through the entire volume, look from different angles, identify structures in cells and reconstruct them in 3D.

F5. *For your next task, you are allowed a maximum time of 1 minute and 30 seconds. Please identify the position of the intracellular bodies within the cells (e.g. – near the boundary or centrally located).* (**Visually complex search**) – The location of intracellular bodies inside a cell in 3D could be close to a boundary or more towards the center. The task was to search for gaps between the intracellular bodies and the boundaries in all directions in every cell and report.
Strategy: Slice; look from different angles through the volume.

F6. *For your next task, you are allowed a maximum time of 1 minute and 30 seconds. Describe the shape of the intracellular bodies. Are they of comparable or differing shape?* (**General description**) – The task was about describing the structures of the intracellular bodies in the cells.
Strategy: Slice; look from different angles through the volume.

F7. *For your last task, you are allowed a maximum time of 1 minute and 30 seconds. This crack is called a fracture. How many cells does this fracture cut through?* (**Simple search**) – The fracture near the bottom right corner (looking from the cut surface of the fossil) was shown. The task was to search for cells through which the fracture cut.
Strategy: Look from different angles; slice if necessary.

Tasks that were dropped after the exploratory study:

With the Mouse Limb dataset:

1. Describe the shape/texture of the tumors.
2. These are the blood vessels. How many blood vessels do you see in this dataset? Do they continue from end to end? Please count the ones you might find on the other side of the dataset as well.

With the fossil dataset:

1. Are the cells of similar or identical volume?
2. Describe the shape of each cell. Are the cells of comparable or differing shapes?
3. Are the intracellular bodies of similar size and volume? If not, then please describe how many of those you see are different and how?
4. Can you identify any fractures or breakages in the dataset?
5. Please describe any other features that you observed, or anything else of interest that you noticed in this fossil. You can also comment about the texture of the individual cells and of the nuclei/intracellular bodies.

ACKNOWLEDGMENTS

The authors wish to thank Patrick Shinpaugh (Virginia Tech) and Oliver Kreylos (University of California, Davis) for their prompt technical support, and Virginia Tech Research Computing's Visionarium Lab for the hardware and facilities used to run this study. The CT studies were performed in the SBES Advanced Multiscale Computed Tomography Facility (SAM-CT), supported by an NIH SIG grant (RR025667), an NSF MRI grant (CMMI0923297) as well as ICTAS and SBES internal funding.

REFERENCES

[1] A. Kaufman, "Volume visualization," *The Visual Computer,* vol. 6, pp. 1-1, 1990.

[2] G. Marmitt, H. Friedrich, and P. Slusallek, "Interactive volume rendering with ray tracing," *Eurographics State of the Art Reports,* pp. 115–136, 2006.

[3] M. Slater, "A note on presence terminology," *Presence connect,* vol. 3, 2003.

[4] K. W. Arthur, K. S. Booth, and C. Ware, "Evaluating 3D task performance for fish tank virtual worlds," *ACM Transactions on Information Systems,* vol. 11, pp. 239–265, 1993.

[5] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, Anaheim, CA, 1993, pp. 135-142.

[6] D. A. Bowman and R. P. McMahan, "Virtual Reality: How Much Immersion Is Enough?," *Computer,* vol. 40, pp. 36-43, 2007.

[7] M. Midttun, R. Helland, and E. Finnstrom, "Virtual reality--adding value to exploration and production," *The Leading Edge,* vol. 19, pp. 538-544, 2000.

[8] L. Arns, C. Cruz-Neira, and D. Cook, "The Benefits of Statistical Visualization in an Immersive Environment," in *Proceedings of the IEEE Virtual Reality*, 1999, pp. 88-95.

[9] K. Gruchalla, "Immersive well-path editing: investigating the added value of immersion," in *Proceedings of the IEEE Virtual Reality*, 2004, pp. 157–164.

[10] P. Schuchardt and D. A. Bowman, "The benefits of immersion for spatial understanding of complex underground cave systems," in *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, 2007, pp. 121–124.

[11] D. Bowman and D. Raja, "A method for quantifying the benefits of immersion using the cave," *Presence-Connect,* vol. 4, 2004.

[12] T. Ni, D. A. Bowman, and J. Chen, "Increased display size and resolution improve task performance in Information-Rich Virtual Environments," in *Proceedings of Graphics Interface*, 2006, pp. 139-146.

[13] 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," *Computer Animation and Virtual Worlds,* vol. 18, pp. 19-38, 2007.

[14] R. P. McMahan, D. Gorton, J. Gresock, W. McConnell, and D.A. Bowman, "Separating the effects of level of immersion and 3D interaction techniques," in *Proceedings of the ACM symposium on Virtual reality software and technology*, 2006, pp. 108–111.

[15] W. Barfield, C. Hendrix, and K. Bystrom, "Visualizing the structure of virtual objects using head tracked stereoscopic displays," in *IEEE Virtual Reality Annual International Symposium*, 1997, pp. 114-120.

[16] C. Ware, and G. Franck, "Evaluating stereo and motion cues for visualizing information nets in three dimensions," *ACM Transactions on Graphics,* vol. 15, pp. 121–140, 1996.

[17] N. Ohno and A. Kageyama, "Scientific visualization of geophysical simulation data by the CAVE VR system with volume rendering," *Physics of The Earth and Planetary Interiors,* vol. 163, pp. 305-311, 2007.

[18] Y. Kitamura, T. Nakashima, K. Tanaka, and T. Johkoh, "The IllusionHole for Medical Applications," in *Proceedings of IEEE Virtual Reality*, 2007, pp. 231-234.

[19] S. Zhang, C. Demiralp, D. F. Keefe, M. DaSilva, D. H. Laidlaw, B. D. Greenberg, P. J. Basser, C. Pierpaoli, E. A. Chiocca, and T. S. Deisboeck, "An Immersive Virtual Environment for DT-MRI Volume Visualization Applications: A Case Study," in *Proceedings of IEEE Visualization*, 2001, pp. 437-584.

[20] S. Zhang, C. Demiralp, and D. H. Laidlaw, "Visualizing Diffusion Tensor MR Images Using Streamtubes and Streamsurfaces," *IEEE Transactions on Visualization and Computer Graphics,* vol. 9, pp. 454-462, 2003.

[21] 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 Transactions on Visualization and Computer Graphics,* vol. 12, pp. 323-330, 2006.

[22] Prabhat, A. Forsberg, M. Katzourin, K. Wharton, and M. Slater, "A Comparative Study of Desktop, Fishtank, and Cave Systems for the Exploration of Volume Rendered Confocal Data Sets," *IEEE Transactions on Visualization and Computer Graphics,* vol. 14, pp. 551-563, 2008.

[23] K. Sensharma, T. Andric, W.J. Freeman, C.L. Wyatt, and G. Wang, "Micro-CT for osteon-like scaffolds," presented at the 11th Annual Conference of the North Carolina Tissue Engineering and Regenerative Medicine Society, 2009.

[24] K. Sensharma, D.M. Vasilescu, A.S.K. Puliyakote, E.A. Hoffman, T. Andric, W.J. Freeman, C. Markert, J.D. Schiffbauer, S. Xiao, H. Yu, and G. Wang, "Novel Biomedical and Biological Applications using Lab-based Multi-scale CT System," presented at the Annual Meeting of the Biomedical Engineering Society, Oct 2011.

[25] J. D. Schiffbauer, S. Xiao, K. Sen Sharma, and G. Wang, "The origin of intracellular structures in Ediacaran metazoan embryos," *Geology (in press),* 2012.

[26] J. Kelso, S.G. Satterfield, L.E. Arsenault, P.M. Ketchan, and R.D. Kriz, "DIVERSE: A Framework for Building Extensible and Reconfigurable Device-Independent Virtual Environments and Distributed Asynchronous Simulations," *Presence: Teleoperators and Virtual Environments,* vol. 12, pp. 19-36, 2003.

[27] M. I. Billen, O. Kreylos, B. Hamann, M.A. Jadamec, L.H. Kellogg, O. Staadt, and D.Y. Sumner, "A geoscience perspective on immersive 3D gridded data visualization," *Computers & Geosciences,* vol. 34, pp. 1056-1072, 2008.

[28] O. Kreylos, "Environment-Independent VR Development," in *Advances in Visual Computing*. vol. 5358, G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. K. Chun, T. Rhyne, and L. Monroe, Ed., ed: Springer Berlin / Heidelberg, 2008, pp. 901-912.

[29] R. B. Ekstrom, J.W. French, and H.H, Harman, "Cognitive factors: Their identification and replication," *Multivariate Behavioral Research Monographs,* 1979.

[30] E. L. Ritman, "Micro-computed tomography-current status and developments," *Annual Review of Biomedical Engineering,* vol. 6, pp. 185-208, 2004.

[31] S. R. Stock, *Microcomputed tomography: Methodology and applications*: CRC Press, 2008.

[32] S. J. Lee, K. Sensharma, E. A. Fox, and G. Wang, "Micro-CT scanner training in a 3D virtual world: Second Life aided training and education (SLATE)," presented at the BMES Annual Meeting, Austin, TX, 2010.