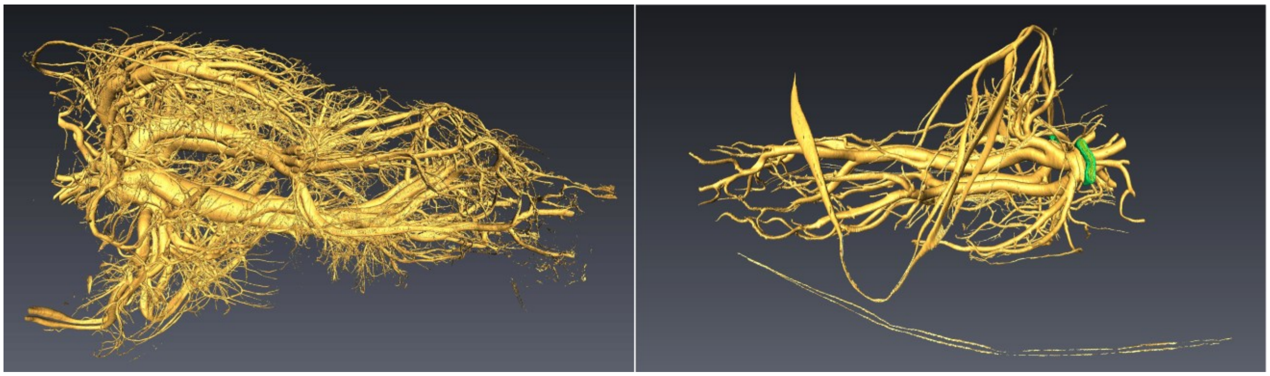
Effects of VR System Fidelity on Analyzing Isosurface Visualization of Volume Datasets

Bireswar Laha, Doug A. Bowman, and John J. Socha



(a) *Pterostichus* dataset (used for training) (b) *Platynus* dataset (used in the main study)

Fig. 1. Isosurfaces of tracheal systems generated from micro-CT scans of beetles used in our VR system fidelity evaluation study.

Abstract—Volume visualization is an important technique for analyzing datasets from a variety of different scientific domains. Volume data analysis is inherently difficult because volumes are three-dimensional, dense, and unfamiliar, requiring scientists to precisely control the viewpoint and to make precise spatial judgments. Researchers have proposed that more immersive (higher fidelity) VR systems might improve task performance with volume datasets, and significant results tied to different components of display fidelity have been reported. However, more information is needed to generalize these results to different task types, domains, and rendering styles. We visualized isosurfaces extracted from synchrotron microscopic computed tomography (SR-μCT) scans of beetles, in a CAVE-like display. We ran a controlled experiment evaluating the effects of three components of system fidelity (field of regard, stereoscopy, and head tracking) on a variety of abstract task categories that are applicable to various scientific domains, and also compared our results with those from our prior experiment using 3D texture-based rendering. We report many significant findings. For example, for search and spatial judgment tasks with isosurface visualization, a stereoscopic display provides better performance, but for tasks with 3D texture-based rendering, displays with higher field of regard were more effective, independent of the levels of the other display components. We also found that systems with high field of regard and head tracking improve performance in spatial judgment tasks. Our results extend existing knowledge and produce new guidelines for designing VR systems to improve the effectiveness of volume data analysis.

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

1 Introduction

Volume visualization offers 3D spatial representations of data generated from various technologies such as computed tomography (CT), magnetic resonance imaging (MRI), confocal microscopy, and ultrasound, and is used extensively to analyze scientific data in various domains including medicine, biology, paleontology, archaeology, engineering, and astronomy [12]. Typically, scientists and researchers use desktop systems, either custom-made or provided by commercial manufacturers of imaging and scanning systems (e.g., Xradia[1] and GE healthcare). These systems usually offer a non-immersive environment to perform the various visual analysis tasks that scientists perform in their research, which often involve analyzing complex structures in 3D volumes.

Virtual Reality (VR) offers an immersive medium for scientific visualization. Such higher-fidelity rendering systems may reveal spatially complex structures in ways easier to analyze, explore, and understand over traditional non-immersive systems [5]. VR researchers investigating the effects of the fidelity of immersive VR systems have run empirical studies showing significant benefits of more immersive systems [21, 32, 33].

VR researchers have broken down immersive VR systems into specific components with objective and measureable levels of fidelity [24], and are running controlled studies reporting effects of individual and combined components of VR system fidelity [3] and on analyzing volume visualization [7, 16, 17].

As the field progresses with gathering empirical results, it is important to generalize our findings across different scientific domains. Previously, we have attempted to tie results to abstract task types [17, 22, 23], but we reported our findings tied only to a few abstract task categories, and mostly involved search tasks.

In addition, there are a number of techniques to visualize a volume, such as decomposition, isosurface rendering, maximum intensity projection, semi-transparency, and x-ray rendering [19]. Each of these techniques offers unique ways of analyzing a volume. Each of the components of VR system fidelity offer different cues to

- *Bireswar Laha is with the Center for Human-Computer Interaction and the Department of Computer Science, Virginia Tech. E-mail: blaha@vt.edu.*
- *Doug A. Bowman is with the Center for Human-Computer Interaction and the Department of Computer Science, Virginia Tech, Blacksburg, VA. E-mail: bowman@vt.edu.*
- *John J. Socha is with the Department of Engineering Science and Mechanics, Virginia Tech, Blacksburg, VA. E-mail: jjsocha@vt.edu.*

Manuscript received 12 September 2013; accepted 10 January 2014; posted online 29 March 2014; mailed on 1 May 2014.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

[1] http://www.xradia.com/solutions/index.php

the user; field of regard (FOR) provides extra virtual space for exploration of structures (thus reducing clutter), stereoscopy (ST) provides better depth cues, and head tracking (HT) provides better motion parallax [3]. It is important to understand whether the effects of these components are consistent for each volume rendering style of a volume, or if their effects depend on the rendering style.

To address these questions, we designed a controlled experiment to evaluate the effects of three components of VR system fidelity on a wide variety of abstract task types. We chose to run a study analyzing isosurface visualization of volumes, in contrast to previous studies that used 3D texture rendering of the volumes [17, 21]. We chose to study synchrotron microscopic CT datasets from the domain of biomechanics to find out if the effects of the components of VR system fidelity are realizable with datasets from a domain different than medical biology, and paleontology, datasets from which researchers have evaluated previously [17, 21].

We report significant improvements in performance for analyzing isosurface visualization of volume datasets, tied to individual and combined effects of the three components of VR system fidelity that we studied (FOR, ST, HT). We also compare the results with those from our previous empirical studies reporting effects of components of VR system fidelity for volume visualization. Our results indicate that the effects of the components of VR system fidelity may depend on the style of rendering a volume.

2 Related Work

Previous researchers seeking to find the effects of immersive VR for scientific visualization initially ran empirical studies comparing more immersive systems directly to less-immersive ones. Some of the earlier studies comparing whole systems against each other reported benefits of CAVE-like systems over desktop systems for interpreting volume visualized diffusion tensor magnetic resonance imaging (DT-MRI) datasets from brain tumor surgery [32], and for analyzing confocal microscopy images of biomedical datasets [21]. Demiralp et al. showed benefits of fishtank VR over CAVE for shape perception tasks [9]. Such empirical results, although very important, failed to tie their results to individual components of the VR system. Their results were thus not generalizable to VR systems beyond the ones directly compared in their studies.

As researchers defined the components of immersive VR system more formally [24], others started running controlled experiments evaluating the individual and combined effects of the components of VR system fidelity on task performance [3]. As Bowman and colleagues showed, displays with higher levels of system fidelity can be used to recreate VR systems with lower levels of fidelity of the same components, using the concept of VR simulation [2, 4]. Such controlled simulations may be used to create generalizable results [15].

Our prior research reported significant effects on various tasks with biomedical (mouse limb), and paleontological (fossil) datasets tied to FOR, ST, and HT [17]. In a follow-up study, we replicated some, but not all, of the significant effects with comparable levels of FOR and HT created with a head mounted display (HMD) system [16], seeking to generalize the results across VR platforms using the concept of VR simulation [4]. In another closely related study, Ragan et al. reported the effects of FOR, HT, and ST on small-scale spatial judgment tasks, while analyzing underground cave systems [22]. In yet another controlled study, Chen et al. reported significant effects of stereo and display size on task performance with DT-MRI datasets [7].

All these empirical studies, although reporting significant findings generalizable based on the components of VR system fidelity studied as independent variables, evaluated performance on tasks specific to the dataset and the domain (identified as a limitation in [15]). Some prior work attempted to identify more general task categories [17, 23]; we build on this work by explicitly evaluating tasks from a systematic list of task categories.

Further, if we attempt to assemble all the significant findings from different empirical studies together, to form generalizable results for volume data analysis, we notice that the studies were run with different volume rendering techniques. As these techniques differ fundamentally in their visual representation of data, the set of visual analysis tasks may also differ significantly among the visualization techniques. For example, isosurface rendering of tubes will more likely require tasks of the type *spatial judgment* (as the user will need to understand the gaps between the vessels), while 3D texture rendering of volumes, as used in our prior work [17], will have more cloudy data, and might involve more *search* based tasks.

In this paper, we tie the effects of VR system fidelity to a wide range of abstract visual analysis tasks for volume visualization (with an effort to generalize results across scientific domains [15]), and to compare the effects between volume rendering techniques.

3 Experiment

We designed a controlled experiment to evaluate the single factor and multi-factor effects of three components of VR system fidelity on task performance in a wide variety of generic task types for analysis of isosurface visualization of volume datasets.

3.1 Goals and Hypotheses

Our main objective in this study is to find out whether different levels of VR system fidelity affect task performance with volume datasets, when the mode of rendering is isosurface visualization. Thus, our first research question is:

1. *Are there any effects of VR system fidelity for analyzing isosurface visualization of volume datasets?*

If we find that the level of VR system fidelity affects task performance, we are further interested in knowing the individual and combined effects of individual components of VR system fidelity for analyzing volumes using isosurface visualization [3, 24]. As prior studies evaluating the effects of VR system fidelity have reported significant effects of field of regard (FOR), stereoscopy (ST), and head tracking (HT) [7, 17, 22], we choose to look at three components of system fidelity for analyzing isosurface rendering of volume datasets. This gives us our next research question:

2. *What are the individual and combined effects of FOR, ST, and HT on analyzing isosurface rendering of volumes?*

In this study, we chose to have two levels each of FOR (90º and 270º), ST (on and off), and HT (on and off).

We are interested in evaluating performance on a wide variety of abstract task categories (see section 3.4) mapped to volume datasets from various domains. This leads us to our third research question:

3. *Are specific tasks of the same abstract type affected similarly by the components of VR system fidelity?*

We are interested in knowing if the effects of VR system fidelity on visual analysis tasks vary with the rendering technique used. This gives us our final research question:

4. *Are the effects of the components of VR system fidelity similar when analyzing isosurface rendering of volumes vs. 3D texture rendering of volumes?*

To the best of our knowledge, we are unaware of any empirical study evaluating the effects of VR system fidelity on different volume rendering techniques (e.g., isosurface vs. semi-transparent). To gather some preliminary findings, we planned to compare the results of this study with those from our previous study [17].

Tied to each of our research questions, we had the following hypotheses:

1. *Higher levels of VR system fidelity will produce better task performance with isosurface visualization of volumes.* Results of previous empirical studies, although reported with different styles of rendering, support this hypothesis in general [21, 32, 33]; but there are some results against this as well [9].
2. *Higher levels of different components of VR system fidelity will improve task performance both individually and when two of them are combined (e.g., FOR and HT both at higher levels).*

Again, some prior results support this claim [17, 22], while others challenge it partially [7].

3. *The different components of VR system fidelity will affect the different abstract task types to different degrees, but there will be noticeable trends tied to individual or combined components of VR system fidelity in each abstract task category.* Prior studies have tried to categorize their significant findings to generalizable task categories [17, 23]. The components of VR system fidelity we evaluated differ fundamentally in their affordances. Thus, intuitively, these would affect the different task types to different degrees as the task types also differ fundamentally (see section 3.4).
4. *The components of VR system fidelity will have different sets of significant effects based on the rendering style used to visualize the volumes.* Our hypothesis stems from our observation that rendering techniques differ fundamentally in the visual representation of data, and the fact that the components of VR system fidelity offer varying affordances for visual analysis [3].

3.2 Datasets

Micro-CT (µCT) is a form of computed tomography that uses x-rays to produce 3D imagery of small, centimeter-scale objects with micrometer-scale resolution. Although widely used, benchtop µCT devices are not as powerful as µCT conducted at 3rd-generation synchrotron light sources, which yields the highest quality µCT data currently available (known as SR-µCT, [25, 31]). Synchrotron simply refers to the way that the x-rays are produced. Data from synchrotron beamlines (place where x-ray experiments are done using synchrotron x-rays) are typically processed in the lab using desktop computers with high-end graphics cards, commercial software, and large flat-screen monitors. Depending on the quality of data, identification of features of interest is done by automated or manual segmentation (a way of visually highlighting or distinguishing features of interest in a volumetric data set), which can be the most time-intensive step in data processing.

Here, we used two SR-µCT datasets (see Fig. 1) collected from the 2-BM beamline at the Advanced Photon Source, Argonne National Laboratory, for our testing. The first dataset was used for training, and consisted of the tracheal system of a carabid beetle (commonly known as ground beetles and belonging to the family Carabidae, there are tens of thousands of species) from the genus *Pterostichus*. The second dataset was used for testing, and consisted of a different carabid beetle species from the genus *Platynus*. These carabid beetles are of scientific interest owing to the species' dynamic tracheal behaviors [27, 30]; both exhibit a rhythmic compressing and reinflation of parts of the tracheal system, with a compression event occurring on the scale of seconds and repeating cyclically on the order of ten times per minute. These compression cycles are thought to produce air movement and so to augment diffusive gas exchange [26].

Although SR-µCT produces high-quality 3D data, the spatial resolution is on the order of a micron, and there exist parts of the tracheal system with tubes of smaller diameter (called 'tracheoles'). Because these were not resolved by the x-rays, they are not included in our 3D rendering. In addition to the tracheal tubes that were visualized, the datasets also include spiracles, which are valve-like elements that serve as the environmental entrance to the system [6].

We used Avizo[2] to generate the isosurfaces using manual and auto segmentation. In all cases, the voxels for inclusion were chosen for best matching the outline of the tracheal tubes. We used open source software to render these isosurfaces in our VR system (see 3.3.1).

Fig. 2. A participant in the FOR_ST_HT condition inside the CAVE.

3.3 Apparatus

3.3.1 Hardware and Software

We used a four-screen CAVE-like system (Fig. 2) [8] with three rear-projected 10' by 10' walls, and a top-projected floor, each with passive Infitec[3] stereo (used in conditions with ST on), and running at 1920×1920 resolution. The head tracking (in the HT on conditions) was provided by an Intersense IS-900 wireless tracking system[4], which also tracked a wireless wand with five buttons and a joystick.

We used open source software to interface with the hardware. DIVERSE [13] provided support for distributed rendering on our cluster of computers running the CAVE system. VRUI [14] provided support for interaction using the wand and the head tracker, through a plugin written to interface with the DIVERSE software. We used an isosurface renderer called meshviewer from the KeckCaves[5] lab for rendering the isosurfaces of the volumes in the CAVE.

3.3.2 User Interactions

We provided users a grab interaction with six degrees of freedom about the absolute position of the grab. This could be activated by pressing the trigger button at the bottom of the wand with the index finger. In addition to the grab action, the users in the head tracked conditions (HT; see Table 1) could also use positional head tracking to get different viewpoints around the datasets based on their absolute head movements inside the CAVE system, which gave them an added mode of interaction.

3.4 Tasks

One of the main objectives of this study was to evaluate the effects of VR system fidelity over a wide variety of abstract task types, so that the significant findings from this study could be generalized to multiple scientific domains. We thus leveraged a list of abstract task types we developed by interviewing domain scientists from medical biology, paleontology, geophysics, and biomechanics over the last few years. The task categories include the following:

1. *Search*—searching for a feature in the dataset or counting the number of a particular type of feature
2. *Pattern recognition*—recognizing repeated characteristics or a trend through the dataset

[2] http://www.vsg3d.com/avizo/overview

[3] http://www.infitec.net/

[4] http://www.intersense.com/

[5] http://idav.ucdavis.edu/~okreylos/ResDev/KeckCAVES/Applications.html

3. *Spatial judgment*—judging the position and/or the orientation of a feature in a 3D spatial context, on an absolute or relative basis, including whether two features are intersecting or not
4. *Quantitative estimation*—estimating the numeric value of some property (e.g., density, size) of the dataset, a region, or a feature
5. *Shape description*—describing qualitatively the shape of either the whole or some part of the dataset)

With these in mind, we developed a set of tasks of real research interest to a researcher in biomechanics, so that the significant findings from this study would come from realistic and relevant tasks. Each of these tasks was assigned to one of the abstract task types. The final set of 15 tasks designed for this study is in the appendix, with the task types noted next to each. The tasks included five search tasks, six spatial judgment tasks, two quantitative estimation tasks, one shape description task, and one pattern recognition task. All the tasks were open-ended but had objective answers, except for task T3, for which we gave the participants five answer options to choose from, to reduce the chances of large variations in their responses.

It is important to note here that we chose to run our evaluation study with novice participants instead of experts in the domain, similar to prior studies [11, 16, 17]. The arguments supporting this choice include expert participants self-reporting as novices in prior studies [17], and the fact that volume data analysis requires significant training [18], indicating that many of these domains have scientists who are similar to novices [28]. Having novice participants also allows us to avoid any confounding effects based on prior knowledge level.

Since the participants were novices, we removed all technical terms from the tasks, but kept the essence of the task the same as designed by the domain scientist. This reduces the potential risk of using novices in the study, because the tasks were easily understandable without domain knowledge, while still representing tasks performed by real-world domain scientists. For example in T6 we said 'top half' instead of 'dorsal side.' Wherever necessary, we included clear explanations. In T8, for example, we defined 'spiracles' for the participants, and also showed them examples before they began.

We also included a 20-minute training session for participants consisting of five tasks spanning the various task types, and teaching them appropriate strategies for completing each task.

3.5 Design

We designed a controlled experiment to study the effects of three components of VR system fidelity as independent variables—field of regard (FOR), stereoscopic rendering (ST), and head tracking (HT). FOR had levels 270º (all four walls of the CAVE system used to render the isosurfaces) and 90º (only the front wall of the CAVE displaying the isosurfaces). ST had levels 'on' (stereoscopic), and 'off' (monoscopic). HT had levels 'on' (head position tracked), and 'off' (the virtual camera was fixed in the center of the CAVE). This gave us eight between-subjects conditions for our study. Table 1 provides the case-sensitive labels for each of these eight independent conditions, which we shall use consistently in this paper.

Table 1. Conditions experienced by the eight groups in the experiment, and their case-sensitive labels used in this paper

Group#	FOR	ST	HT	Label
1	270	On	On	FOR_ST_HT
2	270	On	Off	FOR_ST_ht
3	270	Off	On	FOR_st_HT
4	270	Off	Off	FOR_st_ht
5	90	On	On	for_ST_HT
6	90	On	Off	for_ST_ht
7	90	Off	On	for_st_HT
8	90	Off	Off	for_st_ht

Using the same software and hardware to replicate the different conditions allowed us to keep the other components of VR system fidelity, which include display size, screen resolution, refresh rate, frame rate, and latency [3], at the same level [2, 24]. Participants in all conditions (even the conditions with monoscopic rendering) wore the stereo goggles, which ensured that they experienced the same field of view and the same brightness levels in all conditions.

We had four dependent variables in our study (the study metrics). Two of these were quantitative and objective: the accuracy of the responses of the participants (evaluated offline based on a rubric created by our domain scientist), and the time taken to complete each task. The other two metrics were quantitative and subjective, and included the participants' ratings on seven-point scales for the perceived difficulty of each task, and the level of confidence in each of their answers.

3.6 Participants

We recruited 72 voluntary participants for our study, four of whom were pilot participants. We dismissed 12 participants (who scored less than 3 out of 20 on a spatial ability test [10]) giving us a total of 56 participants, distributed in the eight study groups (seven participants per group), with closely comparable average spatial ability in each group (overall average of 12.1 out of a maximum 20). The participants were all undergraduate or graduate students ranging from 18 to 38 years of age, with an average age of 21.8 years. There were 26 males and 30 female participants. All of them self-reported no prior experience in analyzing volume datasets in general, or isosurface visualization of volumes.

3.7 Procedure

The Institutional Review Board at our university approved our study. After arrival, participants signed an informed consent form, informing them of their rights to withdraw at any point from the experiment. They then filled out a background questionnaire capturing information related to their demographics, and their experience with VR systems and analyzing volume visualization. Then they were asked to take a spatial ability test [10]. Following the test, they were introduced to the CAVE system. The participants were then given an introduction to the background of our experiment, facilities to be used, and study procedures.

The participants then performed five training tasks with a training dataset (Fig. 1-a). During the training, the participants were introduced to the 3D interface and trained on the different strategies and interactions for performing the tasks (very similar to those they would face during the main part of the study). The training lasted for around 20 minutes, after which the participants were given a short break, during which time the experimenter loaded the main dataset (Fig. 1-b).

After the break, the participants performed 15 tasks in a consistent order (see appendix) with the main dataset, taking a short break after the seventh or eighth task. To maintain consistency between the participants and the condition, the datasets were rendered at the same initial position before each task, and we used consistent phrasing for every question, which was read aloud to the participants. Each task consisted of listening to a question, analyzing the dataset for the answer, and reporting the answer back to the experimenter. The experimenter recorded the responses to the questions, along with the time taken to carry out each task. Finally, the participant reported a subjective rating of the task difficulty and a subjective level of confidence in their answer, on two seven-point scales.

After completing the tasks, participants filled out a post questionnaire, capturing on seven-point scales their ease of getting viewpoints, ease of analyzing the dataset, frequency of using the grab action and walking around the dataset, and their levels of fatigue, eye strain, and dizziness. The experimenter then conducted a final free-form interview to answer any additional questions from the participants.

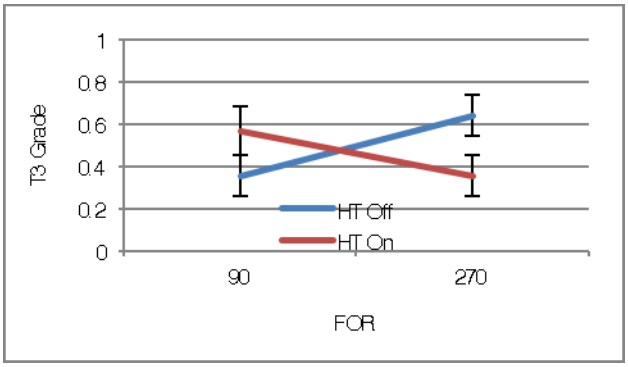


Fig. 3. Interaction between FOR and HT for Grade in T3, a Quantitative Estimation Task.

4 Results

Here we report the statistically significant results in our study. All dependent variables in our study were of numeric ordinal type, except for the time metric, which was numeric continuous. Thus, to know the main and interaction effects on the independent variables (FOR, ST, HT), we ran an Ordinal Logistic Regression based on a Chi-square statistic on all metrics, except for the time metric, for which we ran a three-way analysis of variance (ANOVA).

When we found a significant two-way or three-way interaction between the independent variables, to know which combinations were significantly different, we used a two-sided Wilcoxon Signed-Rank Tests for post-hoc analyses for all metrics, except for the time metric, for which we ran the Student's t test.

We decided against running a multivariate analysis of variance as the tasks in our study are intentionally and fundamentally different, and the metrics are of different data types.

Unlike previous studies [17, 21], we decided not to base our analysis primarily on a cumulative score (weighted average of scores obtained in each task) to compare the independent conditions on an overall basis, as we consciously tried to group the tasks in fundamentally distinct categories. We report the results tied to the abstract task groups in our study (see section 5.2).

The significant main and interaction effects of the display components on the various task types are summarized in Table 9.

4.1 Grades (task performance accuracy)

We observed five significant main effects of FOR and ST on the grades obtained by the participants, shown in Table 2 below; higher levels of these improved accuracy of task performance in all cases.

Table 2. Significant Main Effects on Grades

Task: source	χ^2	DF	p-value	Comparison between levels of components
T11: FOR	5.267	1	0.0217	FOR 270 more accurate
T4: ST	8.369	1	0.0038	ST on more accurate
T6: ST	7.936	1	0.0048	ST on more accurate
T9: ST	5.068	1	0.0244	ST on more accurate
T14: ST	4.557	1	0.0328	ST on more accurate

We also observed three significant interaction effects on accuracy of task performance. These are in Table 3, and shown in Fig. 3 and Fig. 4, with standard error bars. Post-hoc tests show that for T3, the condition FOR_ht produced significantly more accurate task performance than the conditions FOR_HT and for_ht (p=0.0469).

4.2 Completion time (speed of task performance)

There were several significant main effects of FOR and ST on the task completion time, shown in Table 4. Higher levels of VR system fidelity components improved speed of task completion in each case.

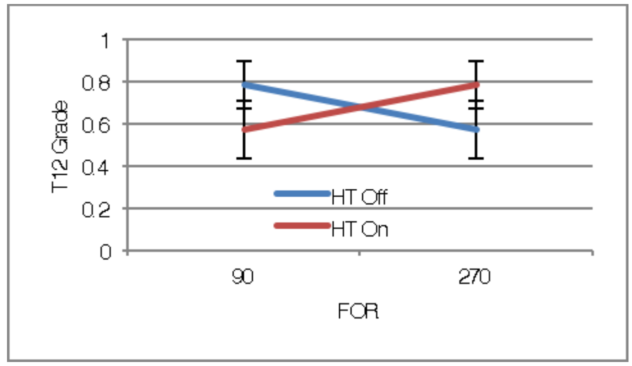


Fig. 4. Interaction between FOR and HT for Grade in T12, a Pattern Recognition Task.

We observed a significant three-way interaction effect of FOR, ST and HT on the speed of completion of task T5 (see Table 5). Post-hoc tests indicate that all conditions with *stereo on* were faster than others, and the performance in the highest fidelity condition was significantly faster than that in the lowest fidelity condition.

4.3 Perceived levels of difficulty (subjective metric)

We observed a significant main effect of ST on perceived levels of difficulty reported by the participants. Participants felt that stereo reduced the difficulty level of task T14 ($\chi^2_{df=1}$=5.479; p=0.0192).

We also observed four cases of significant interaction between FOR, ST and HT on the difficulty of tasks, as shown in Table 6.

Post-hoc tests indicate that for task T11, the condition ST_ht was significantly less difficult (p=0.0425) than both st_HT and st_ht, the condition for_ST_ht was significantly less difficult than FOR_st_HT (p=0.0156) and for_st_ht (p=0.0313), and the condition FOR_st_ht was significantly less difficult than for_st_ht (p=0.0313).

4.4 Confidence levels in response (subjective metric)

We observed a significant main effect of HT on the perceived levels of confidence of the participants in their answers. For T4, the participant's confidence was significantly improved by head tracking ($\chi^2_{df=1}$ = 6.104, p=0.0135).

We observed four significant interaction effects of FOR, ST, and HT on the perceived confidence levels, shown in Table 7. Post-hoc tests indicate that for T3, the participants had significantly higher confidence in the condition for_ht than in for_HT (p=0.0449), and for T11, they had significantly higher confidence in the conditions st_ht (p=0.0054) and ST_HT (p=0.0156) than in ST_ht. For T7, the participants had significantly higher confidence in the condition FOR_ST_ht than both the conditions for_ST_ht (p=0.0313) and for_st_HT (p=0.0469).

Table 3. Significant Interaction Effects on Grades

Task: source	χ^2	DF	p-value	Mean values in descending order (higher is better)	
T3: FOR & HT	6.3325	1	0.0119	FOR_ht	0.64
				for_HT	0.57
				FOR_HT	0.36
				for_ht	0.36
T12: FOR & HT	4.9773	1	0.0257	FOR_HT	0.79
				for_ht	0.79
				FOR_ht	0.57
				for_HT	0.57
T8: FOR, ST & HT	5.4470	1	0.0196	for_ST_HT	0.97
				FOR_st_HT	0.94
				FOR_ST_ht	0.94
				for_ST_ht	0.92
				FOR_ST_HT	0.89
				for_st_HT	0.83
				for_st_ht	0.81
				FOR_st_ht	0.71

Table 4. Significant Main Effects on Time

Task: source	F-Ratio	DF	p-value	Comparison between levels of components
T2: FOR	4.2137	1	0.0456	FOR 270 faster
T5: ST	5.1657	1	0.0276	ST on faster
T7: ST	8.9653	1	0.0043	ST on faster
T8: ST	11.6616	1	0.0013	ST on faster
T9: ST	7.6505	1	0.008	ST on faster
T10: ST	5.3591	1	0.0249	ST on faster

Table 5. Significant Interaction Effects on Time

Task: source	F-Ratio	DF	p-value	Mean values (lower is better) - pairs not connected by the same letter are significantly different		
T5: FOR, ST & HT	4.5743	1	0.0376	FOR_ST_HT	A	42.51
				for_ST_ht	A	51.43
				for_ST_HT	ABC	69.04
				FOR_ST_ht	ABC	80.36
				FOR_st_ht	ABC	81.79
				for_st_HT	BC	83.17
				FOR_st_HT	BC	92.13
				for_st_ht	C	97.54

Table 6. Significant Interaction Effects on Perceived Difficulty of Tasks

Task: source	χ^2	DF	p-value	Mean values in ascending order (lower is better)	
T5: FOR & HT	4.2870	1	0.0384	for_HT	5.0
				FOR_ht	5.1
				for_ht	5.7
				FOR_HT	5.9
T11: ST & HT	5.6840	1	0.0171	ST_ht	4.6
				ST_HT	5.4
				st_ht	5.6
				st_HT	5.6
T9: FOR, ST & HT	4.4970	1	0.0339	for_st_HT	5.4
				FOR_st_ht	5.4
				FOR_ST_HT	5.9
				for_st_ht	6.1
				for_ST_ht	6.1
				for_ST_HT	6.4
				FOR_st_HT	6.4
				FOR_ST_ht	6.7
T11: FOR, ST & HT	5.211	1	0.0224	for_ST_ht	4.4
				FOR_st_ht	4.9
				FOR_ST_ht	4.9
				for_st_HT	5.0
				for_ST_HT	5.4
				FOR_ST_HT	5.4
				FOR_st_HT	6.1
				for_st_ht	6.3

4.5 Post questionnaire results (subjective ratings)

In the post questionnaire, participants reported that they grabbed the dataset significantly less ($\chi^2_{df=1}$ = 4.0835, p=0.0433) in the higher FOR conditions. They also reported to have walked significantly more frequently around the dataset to look from various viewpoints ($\chi^2_{df=1}$ = 4.9529, p=0.0260), and also felt less dizzy ($\chi^2_{df=1}$ = 4.0332, p=0.0446) when head tracking was working.

There were two significant interaction effects of FOR and HT on the ease of obtaining desired viewpoints, and dizziness, shown in Table 8. Post-hoc tests indicate that the participants felt significantly less dizzy in the condition for_ht than in the conditions for_HT (p=0.0332) and FOR_ht (p=0.002), and also significantly less dizzy in the condition FOR_HT than in the conditions for_HT (p=0.0293) and FOR_ht (p=0.002).

We also found that majority of the participants in the FOR_HT condition felt the need for a fourth wall of the CAVE for many of the tasks, indicating the need for a VR system with 360º field of regard. Also, few of the participants thought zooming was helpful in certain tasks, but more usable with surrounding visuals (higher FOR).

Table 7. Significant Interaction Effects on Perceived Confidence

Task: source	χ^2	DF	p-value	Mean values in descending order (higher is better)	
T3: FOR & HT	5.9430	1	0.0148	for_ht	4.9
				FOR_HT	4.8
				FOR_ht	4.2
				for_HT	4.1
T7: ST & HT	6.2699	1	0.0123	ST_HT	5.0
				ST_ht	4.8
				st_ht	4.7
				st_HT	4.4
T11: ST& HT	6.2346	1	0.0125	st_ht	5.6
				ST_HT	5.4
				st_HT	5.3
				ST_ht	4.4
T7: FOR, ST & HT	7.2561	1	0.0071	FOR_ST_ht	5.6
				for_ST_HT	5.3
				for_st_ht	5
				FOR_st_HT	5
				FOR_ST_HT	4.7
				FOR_st_ht	4.4
				for_ST_ht	4
				for_st_HT	3.7

Table 8. Significant Interaction Effects on Post Questionnaire Ratings

Effect: source	χ^2	DF	p-value	Label and Mean values	
Ease of getting viewpoint: FOR & HT	4.7709	1	0.0289	FOR_HT	5.9
				for_ht	5.8
				FOR_ht	5.2
				for_HT	5.1
Dizziness: FOR & HT	12.3238	1	0.0004	for_ht	1.4
				FOR_HT	1.5
				for_HT	2.4
				FOR_ht	3.0

5 Discussion

Addressing our first research question (regarding the effects of the components of fidelity), we found significant main effects as well as multi-factor interaction effects of FOR, ST, and HT on the visual analysis of isosurface visualization of volume datasets. All the significant main effects of FOR, ST and HT on the principal metrics in our study (grade, time, difficulty and confidence) showed improved task performance with higher levels of fidelity, which strongly supports our first hypothesis. To illustrate this at a high level, Fig. 5 and Fig. 6 show the average time and grade across the different conditions in our study. These figures reflect the overall trend that time decreases and grade increases (in general) with increasing fidelity levels of the display components.

Table 9 gives an overview of how many significant results were observed compared to the total number of significance tests we performed. As the table and the results in section 4 show, we found significant effects for 12 of the 15 tasks in our study, with all of these effects favoring the higher-fidelity conditions. We consider this to be strong evidence of the benefits of higher fidelity VR systems for isosurface visual analysis.

Table 9. Distribution of significant effects across task types; a cross (X) denotes a significant main effect of the variable at the top of the column for the task in the row, for the metric in the column header. Similarly, connected circles (O) denote significant interaction effects

	Grade			Time			Difficulty			Confidence		
	FOR	ST	HT	FOR	ST	HT	FOR	ST	HT	FOR	ST	HT
T1 (search, counting)												
T4 (search, counting)		X										X
T5 (search)				O	O	O	O		O			
T8 (search)	O	O	O		X							
T9 (search)		X			X		O	O	O			
T2 (spatial judgment)				X								
T6 (spatial judgment)		X										
T10 (spatial judgment)					X							
T11 (spatial judgment)	X						O	O	O		O	O
T13 (spatial judgment)												
T14 (spatial judgment)		X						X				
T3 (quantitative estimation)	O		O							O		O
T15 (quantitative estimation)												
T7 (shape description)					X					O	O	O
T12 (pattern recognition)	O		O									

Referring to our second research question, on the individual and combined effects of FOR, ST, and HT, we found that adding stereo alone significantly improved the task performance in many cases, showing that stereo strongly supports visual analysis of isosurface visualization. Higher FOR alone significantly improved task performance in a few cases, while head tracking alone significantly improved task performance for only one task.

5.1 Interaction effects of the components of fidelity

Field of regard and head tracking produced several significant interaction effects. The conditions FOR_HT (most similar to real world), and for_ht (most similar to a desktop) produced significantly higher grades in T12, and higher confidence levels in T3. In these conditions (FOR_HT, and for_ht), the participants had significantly higher ease of getting the viewpoints they wanted around the dataset, and felt significantly less dizzy than in the other two conditions. These observations could be attributable to the familiarity of the participants to these conditions, as previous studies have also found [17, 20]. But these two conditions (FOR_HT and for_ht) also produced significantly lower grades in T3, and higher difficulty levels in T5, suggesting a possible interaction with the task types, which we look at more closely in the next section.

We also observed several significant interactions between ST and HT on the subjective metrics (difficulty and confidence levels). Again, the lack of generality in these findings suggests possible interaction with the task types, as discussed in the next section.

We observed many significant three-way interaction effects between FOR, ST, and HT. The first one on the grade metric of T8 indicated higher accuracy in conditions with any two of the system fidelity components at the higher level. The next one suggested faster completion rate for T5 in the conditions with stereo on.

The significant three-way interactions on the difficulty metric were not directly comparable, but we did notice a trend tied to a task type, which we discuss in the next section. Another significant three-way interaction suggested lower confidence levels in the conditions with just one component of system fidelity at the higher level.

5.2 Effects of VR system fidelity in different task categories

For our third research question, on the influence of task type on the effects of fidelity, we found that different task types had different sets of single and multi-factor significant effects. Table 9 allows us to examine the consistency of the significant results within each task type. We expected that tasks of the same type would result in similar effects, but Table 9 makes it clear that this was not the case in general. The lack of consistency within the task categories implies that not all tasks in a given category are created equal—the specifics of the particular task have an important effect on performance. Still,

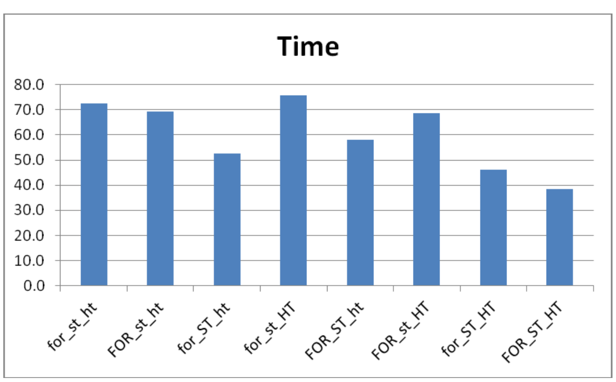


Fig. 5. Average time in different conditions.

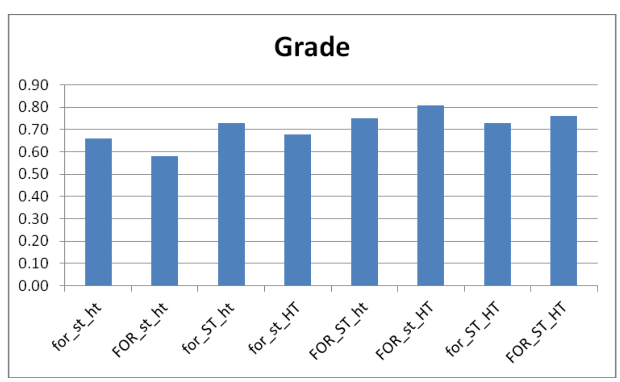


Fig. 6. Average grade in different conditions.

we can draw some more general conclusions from our findings.

Stereo significantly improved task performance for quite a few search tasks (T4, T5, T8 and T9). For task T9, both the accuracy and the speed of task completion improved significantly with stereo. Also, in the significant three-way interaction between FOR, ST and HT on task T5 (search task), all conditions with stereo were faster than others. These results together present strong evidence of stereo alone improving performance for search tasks in isosurface visualization. We believe that the better depth perception provided by stereo might have allowed faster identification of occluded structures and features in the mesh of isosurfaces.

Stereo also significantly improved task performance for a few spatial judgment tasks (T6, T10, T14), and significantly reduced the perceived difficulty for T14. Better depth perception provided by stereo might have improved the analysis of gaps and connections needed in spatial judgment tasks.

Higher FOR significantly improved task performance in two spatial judgment tasks (T2, T11). We suggest that the extra real estate of visual imagery might have made it easier to judge the spatial gaps or connections and lowered the time needed to re-contextualize the detailed judgment within the whole dataset.

Head tracking significantly improved confidence in identifying the beetle's legs based on the configuration of the tracheal tubes (task T4). This suggests that moving the head to easily obtain views from different angles might aid the brain in distinguishing major anatomical patterns.

From the two-way interactions between ST and HT on the difficulty and confidence metrics, we found that for a spatial judgment task (T11), the condition ST_ht (stereo on but head tracking off) improved confidence levels significantly but was also perceived significantly more difficult by the participants than the condition st_ht (both stereo and head tracking off). We surmise that while better depth cues improve confidence in task performance, the accommodation-convergence mismatch inherent to stereoscopic projection systems may also make the use of stereo for detailed spatial judgments feel more difficult to users.

The two-way interaction between FOR and HT for task T12, which required both search and spatial understanding, showed significantly higher scores in the conditions most similar to the real world (FOR_HT) and most similar to a desktop (for_ht). Like previous studies have reported [17, 20], VR systems with familiar fidelity levels (based on real world experience) might prove beneficial for search and spatial understanding tasks.

On the other hand, the two-way interaction between FOR and HT for a size estimation task (T3) had significantly higher scores for the FOR_ht and for_HT conditions. It is unclear why the higher fidelity FOR_HT conditions did not perform as well here.

A closer look at the three way interactions between FOR, ST and HT on the search tasks (T8, T5) revealed that the conditions producing better performance in the grade and time metrics had at least two of FOR, ST and HT at the higher level, or at least ST at the higher level, but the same conditions produced higher difficulty levels in T9 (another search task) as well. This observation resonates with the finding that stereo alone, as well as FOR and HT together, improved the task performance in search tasks. The perceived higher difficulty might have been due to unfamiliarity to the higher display fidelity levels.

5.3 Effects of VR system fidelity with different rendering styles

To address our fourth research question (whether the effects of fidelity are dependent on rendering style), we conducted a meta-analysis of our results, comparing them with the significant findings from our recent controlled experiment, which reported visual analysis task performance with the same three components of VR system fidelity (FOR, ST, HT) as independent variables, but on 3D texture based volume visualization [17]. That study found a number of main effects due to FOR, while in the current study, the majority of significant single-factor effects were due to ST. We believe that the better depth cues provided by stereo (through binocular disparity) aids in better judging the gaps or connections between the isosurface rendering of the tracheal tubes (see Fig. 1-b), as through a dense network of vessels [1, 29]. Stereo may not be as effective with 3D texture rendering, as the dense suspended matter occludes much of the gaps between the structures, but the extra virtual space provided by higher FOR might serve to unclutter the dense volume rendered using 3D textures [17].

Table 10. Re-defining categories of tasks from our previous study [17]

Mouse task#	Abstract Task Type	Fossil Task#	Abstract Task Type
M1	Search	F1	Shape Description
M2	Shape Description	F2	Search
M3	Search	F3	Quantitative Estimation
M4	Spatial Judgment	F4	Search
		F5	Pattern Recognition
		F6	Shape Description
		F7	Spatial Judgment

In order to compare the results with respect to task types, we re-categorized the tasks from our prior experiment based on our current definitions of abstract task categories (see section 3.4), as shown in Table 10.

The significant interaction between FOR and HT that we observed for the grade metric in task T12 (requiring both search and spatial judgment) was very similar to the significant interaction between FOR and HT that our prior study found for the grade metric in task M4. A closer look at these two tasks (T12 from our current study and M4 from our earlier study [17]) suggests a strong similarity in terms of *spatial judgment*. The comparability between the FOR and HT interaction graphs for M4 and T12 indicates that the conditions FOR high HT on and FOR low HT off are quite suitable for tasks requiring spatial judgment, as other studies have also reported recently [20, 22].

Our prior study also found significant three-way interactions between FOR, ST and HT for shape description tasks [17], as we did for one such task in the current study. Task F6 in our earlier study was very similar to task T7 in our current experiment. Looking closely at the three-way interaction effects for these two tasks, we found that the conditions FOR_ST_ht and for_ST_HT produced lower perceived difficulty levels in task F6 in our prior study, and the same two conditions produced higher levels of confidence in task T7 in the current experiment. This indicates that describing shapes in 3D volumes is affected in complex ways by VR system fidelity, independent of rendering style. It also suggests the need for further research exploring the interactions between VR system fidelity, and shape description tasks in volume visualizations.

5.4 Implications for design

Based on the significant findings from this study, we offer a few implications for designing immersive VR systems to improve task performance while analyzing volume visualizations:

1. For analysis of isosurface rendering, stereoscopic displays can be very effective (particularly for search and spatial judgment tasks). For analysis of volume visualization based on 3D texture, systems with high FOR are more effective, independent of the fidelity of other components of the VR system.
2. When analyzing isosurface rendering, higher levels of fidelity based on FOR, ST and HT can improve analysis speed in a variety of tasks.
3. We recommend VR systems with both FOR and HT at higher levels for tasks that require spatial judgment in volumes.

6 Conclusions and Future Work

We ran a controlled experiment evaluating the effects of three components of VR system fidelity (FOR, ST, HT) on visual analysis task performance with isosurface visualization of SR-μCT volume datasets. We found that higher levels of fidelity, overall, resulted in improved task performance. In particular, stereo had the strongest effects on task performance (among FOR, ST and HT), with significantly better performance on several search and spatial judgment tasks. FOR improved performance in two spatial judgment tasks, and HT improved confidence in one search task.

We compared our current findings with those from our previous experiment, indicating that the effects of VR system fidelity may vary based on the rendering technique used to visualize a volume. In particular, stereo might be useful for analyzing isosurfaces, while FOR might improve analysis of semi-transparent volume rendering. Based on our findings we provided design guidelines for VR systems, based on the fidelity of the components of display, for effective task performance with volume datasets.

This study also raised some intriguing questions: Why do we see the pair of significant interactions that are almost mirror images of each other (Fig. 3 and Fig. 4)—what differences in the tasks might have caused these interactions, and what do they mean? We also want to be able to explain more clearly why we observed so many positive effects of stereo on speed of task completion, and what characteristics of the tasks were responsible for this recurring and significant result.

Much work is needed before we can understand more clearly the effects of the different components of VR system fidelity on different abstract task types in volume visualization, necessary for recommending effective VR systems for scientists and researchers looking to optimize task performance. Also, this study is one of the first that reports an interaction between VR system fidelity and the rendering style of volume visualization. We will need further investigation to have a stronger mapping between the components of a VR system and the effectiveness of volume rendering. Finally, we as a community of VR and visualization researchers need to identify and define abstract task categories cutting across various scientific domains of volume data, so that we can leverage that framework to empirically evaluate VR system fidelity on task performance with volume visualization [15].

Appendix

Tasks with the *Platynus* dataset:

T1. Air sacs are parts of the tracheal system that are balloon-like in shape, and are distinguished from tracheal tubes, which are cylindrical. Does this specimen possess any air sacs? If yes, how many? (**Search, counting**)

T2. Look at this circular object near the head of the animal. Is this connected to the surrounding tracheal tubes? If yes, then show the connection point. (**Spatial Judgment**)

T3. Scan the entire body. Find the tracheal tubes of the largest and smallest diameters. How many times bigger is the biggest tube than the smallest one? When you are done, please let me know - I will show you five options to choose from. (**Quantitative Estimation**). Options: 5, 15, 30, 50, 60.

T4. How many legs are there? Please identify each one. (**Search, counting**)

T5. This is a leg. The leg connects to the body at the bend. How many tracheal tubes connect the body to this leg? (**Search**)

T6. Find the tracheal tubes in the abdomen. Are there any tracheal tubes in the top half of the abdomen that definitively connect the left and right portions of the system? To qualify, the tracheal tube reaching across the body must connect to the other side; it can't end blindly in the abdomen. If yes, are there multiple locations? (**Spatial Judgment**)

T7. Most tracheal tubes are circular in cross section, or nearly circular. Do any tracheal tubes exhibit a decidedly non-circular cross-section? If so, where in the body are they located? (**Shape Description**)

T8. The spiracles are the oval-shaped regions that act as valves between the tracheal system and the external air. This is an example of a spiracle inside this beetle. How many spiracles can you find in this entire sample? Search both the left and right sides of the beetle. (**Search**)

T9. Does the number of spiracles on the left side match the number of spiracles on the right side? If not, what is the difference? (**Search**)

T10. The manifold is the part just below the spiracle, where the tracheal tubes join. For this spiracle (third one on the left side), how many tracheal tubes connect to the manifold? (**Spatial Judgment**)

T11. Examine the number of tracheal tubes entering the manifold of the spiracle 5 on both the left and right sides. Are they equal? If no, by what number are they different? (**Spatial Judgment**)

T12. Is there a spiracle that is connected to only one tracheal tube? If yes, which one is it? (**Pattern Recognition**)

T13. This is the spiracle-1. Now trace this tracheal tube towards the head, and count the number of times it branches. At each branching point, always choose the larger branch. (**Spatial Judgment**)

T14. Look at this tracheal tube in the abdomen region. Please trace this tube to its closest spiracle. Which spiracle is it? (**Spatial Judgment**)

T15. What region of the body appears to have the highest density of tracheal tubes, in a one cubic foot space? These are the regions I want you to look at. I will ask you to arrange these regions in terms of decreasing density of tracheal tubes, from highest to lowest. (**Quantitative Estimation**)

Acknowledgments

Our research was supported by the National Science Foundation under Grant No. 1320046 and 0938047, an IBM PhD Fellowship (2013-14), and by the Virginia Tech Institute for Critical Technology and Applied Science (ICTAS). Use of the Advanced Photon Source, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the U.S. DOE under Contract No. DE-AC02-06CH11357. Thanks to Oliver Kreylos from University of California, Davis, for providing the Meshviewer software to render the isosurfaces on Vrui platform.

References

[1] 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.

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

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

[4] D. A. Bowman, C. Stinson, E. D. Ragan, S. Scerbo, T. Höllerer, C. Lee, R. P. McMahan, and R. Kopper, "Evaluating effectiveness in virtual environments with MR simulation," in *Interservice/Industry Training, Simulation, and Education Conference*, 2012.

[5] S. Bryson. (1996) Virtual reality in scientific visualization. *Communications of the ACM*. 62–71.

[6] R. F. Chapman, *The Insects - Structure and Function.*, 4th ed.: Cambridge University Press, 1998.

[7] J. Chen, H. Cai, A. P. Auchus, and D. H. Laidlaw, "Effects of Stereo and Screen Size on the Legibility of Three-Dimensional Streamtube Visualization," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, pp. 2130 -2139, 2012.

[8] 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.

[9] 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.

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

[11] A. Forsberg, J. Chen, and D. H. Laidlaw, "Comparing 3D vector field visualization methods: A user study.," *IEEE Transactions of Visualization and Computer Graphics,* vol. 15, pp. 1219-1226, 2009.

[12] A. E. Kaufman, "Volume visualization," *ACM Computing Surveys,* vol. 28, pp. 165-167, 1996.

[13] 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.

[14] 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.

[15] B. Laha and D. A. Bowman, "Identifying the Benefits of Immersion in Virtual Reality for Volume Data Visualization," presented at the Immersive Visualization Revisited Workshop of the IEEE VR conference, 2012.

[16] B. Laha, D. A. Bowman, and J. D. Schiffbauer, "Validation of the MR Simulation Approach for Evaluating the Effects of Immersion on Visual Analysis of Volume Data," *IEEE Transactions of Visualization and Computer Graphics,* vol. 19, pp. 529-538, 2013.

[17] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman, "Effects of Immersion on Visual Analysis of Volume Data," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, pp. 597-606, 2012.

[18] 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.

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

[20] R. P. McMahan, D. A. Bowman, D. Zielinksi, and R. Brady, "Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, pp. 626-633, 2012.

[21] 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.

[22] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman, "Studying the Effects of Stereo, Head Tracking, and Field of Regard on a Small-Scale Spatial Judgment Task," *IEEE Transactions on Visualization and Computer Graphics,* vol. 19, pp. 886-896, 2013.

[23] 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.

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

[25] J. J. Socha and F. De Carlo, "Use of synchrotron tomography to image naturalistic anatomy in insects.," in *Developments in X-Ray Tomography VI: 2008, SPIE*, San Diego, CA, 2008, pp. 70780A-70787.

[26] J. J. Socha, T. D. Förster, and K. J. Greenlee, "Issues of convection in insect respiration: Insights from synchrotron X-ray imaging and beyond.," *Respiratory Physiology & Neurobiology,* vol. 173, pp. S65-S73, 2010.

[27] J. J. Socha, W. K. Lee, J. F. Harrison, J. S. Waters, K. Fezzaa, and M. W. Westneat, "Correlated patterns of tracheal compression and convective gas exchange in a carabid beetle.," *Journal of Experimental Biology,* vol. 211, pp. 3409-3420, 2008.

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

[29] 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.

[30] M. W. Westneat, O. Betz, R. W. Blob, K. Fezzaa, W. J. Cooper, and W. K. Lee, "Tracheal respiration in insects visualized with synchrotron X-ray imaging.," *Science,* vol. 299, pp. 558-560, 2003.

[31] M. W. Westneat, J. J. Socha, and W. K. Lee, "Advances in biological structure, function, and physiology using synchrotron x-ray imaging.," in *Annual Review of Physiology*, 2008, pp. 119-142.

[32] 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.

[33] 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.