

# **Immersive Virtual Reality and 3D Interaction for Volume Data Analysis**

Bireswar Laha

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in  
partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Computer Science and Applications

Doug A. Bowman, Chair

David H. Laidlaw

Yong Cao

Chris L. North

Nicholas F. Polys

August 04, 2014

Blacksburg, VA

Keywords: System Fidelity, Immersion, Virtual Reality, Virtual Environments, CAVE, Head Mounted Display, HMD, 3D Interaction, 3DI, Bimanual Interaction, Two-handed Interaction, THI, Volume Visualization, Scientific Visualization, 3D Visualization, Task Taxonomy.

Copyright © 2014 Bireswar Laha

All Rights Reserved

# **Immersive Virtual Reality and 3D Interaction for Volume Data Analysis**

Bireswar Laha

## **ABSTRACT**

This dissertation provides empirical evidence for the effects of the fidelity of VR system components, and novel 3D interaction techniques for analyzing volume datasets. It provides domain-independent results based on an abstract task taxonomy for visual analysis of scientific datasets. Scientific data generated through various modalities e.g. computed tomography (CT), magnetic resonance imaging (MRI), etc. are in 3D spatial or volumetric format. Scientists from various domains e.g., geophysics, medical biology, etc. use visualizations to analyze data. This dissertation seeks to improve effectiveness of scientific visualizations.

Traditional volume data analysis is performed on desktop computers with mouse and keyboard interfaces. Previous research and anecdotal experiences indicate improvements in volume data analysis in systems with very high fidelity of display and interaction (e.g., CAVE) over desktop environments. However, prior results are not generalizable beyond specific hardware platforms, or specific scientific domains and do not look into the effectiveness of 3D interaction techniques.

We ran three controlled experiments to study the effects of a few components of VR system fidelity (field of regard, stereo and head tracking) on volume data analysis. We used volume data from paleontology, medical biology and biomechanics. Our results indicate that different components of system fidelity have different effects on the analysis of volume visualizations. One of our experiments provides evidence for validating the concept of Mixed Reality (MR) simulation.

Our approach of controlled experimentation with MR simulation provides a methodology to generalize the effects of immersive virtual reality (VR) beyond individual systems. To generalize our (and other researcher's) findings across disparate domains, we developed and evaluated a taxonomy of visual analysis tasks with volume visualizations. We report our empirical results tied to this taxonomy.

We developed the Volume Cracker (VC) technique for improving the effectiveness of volume visualizations. This is a free-hand gesture-based novel 3D interaction (3DI) technique. We describe the design decisions in the development of the Volume Cracker (with a list of usability criteria), and provide the results from an evaluation study. Based on the results, we further demonstrate the design of a bare-hand version of the VC with the Leap Motion controller device. Our evaluations of the VC show the benefits of using 3DI over standard 2DI techniques.

This body of work provides the building blocks for a three-way many-many-many mapping between the sets of VR system fidelity components, interaction techniques and visual analysis tasks with volume visualizations. Such a comprehensive mapping can inform the design of next-generation VR systems to improve the effectiveness of scientific data analysis.



## Acknowledgements

I take this opportunity to thank Poulomi, my wife and better half, for supporting me all through my career. Thanks to Prakash Chandra Banerjee (my father-in-law) for motivating me to pursue research and higher studies. Thanks to my parents, in-laws and relatives for all their support.

I express my heartfelt thanks to my PhD adviser Professor Doug Bowman for trusting in my abilities, sharpening my mind and my critical thinking, and teaching me science and research.

Special thanks to all my committee members for choosing to serve on my PhD committee and guiding me at various points of my research.

Thanks to ...

All past and present 3DI group members (with special mention to Regis Kopper, Ryan McMahan and Anamary Leal) for providing positive criticisms, and pointers for improving my research;

Patrick Shinpaugh (Virginia Tech, past) and Oliver Kreylos (University of California at Davis) for providing prompt technical support critical to my research;

Kriti Sen Sharma (GE Healthcare), James Schiffbauer (University of Missouri), John J. Socha (Virginia Tech), Josh Tan (Wake Forest School of Medicine), and Scott King (Virginia Tech) for collaborating from various scientific domains for my research;

The National Science Foundation, International Business Machines, and NGRAIN Corporation for recognizing and supporting my dissertation research through generous grant money, fellowship, and software;

All my friends and well wishers, who directly or indirectly contributed to my research.

## Attributions

Doug A. Bowman is a Professor of Computer Science and the Director for the Center for Human-Computer Interaction at Virginia Tech. He shaped my vision and research directions, and is a significant contributor to this whole dissertation. He has been my PhD adviser and has co-authored all my published papers (Chapters 3, 4, 5 and 6).

Kriti Sen Sharma is currently a Lead Engineer at GE Healthcare. He graduated from Virginia Tech with a PhD in Electrical and Computer Engineering. He contributed as a medical biology domain scientist to my research and co-authored the paper forming the Chapter 3 in this dissertation.

James D. Schiffbauer is an Assistant Professor in the Department of Geological Sciences at the University of Missouri. He graduated from Virginia Tech with a PhD in Geosciences. He contributed as a paleontology domain scientist to my research and co-authored two papers forming Chapters 3 and 4 in this dissertation.

John J. Socha is an Associate Professor in the Department of Engineering Science and Mechanics at Virginia Tech. He contributed as a biomechanics domain scientist to my research and co-authored the paper forming the Chapter 5 in this dissertation.

## Table of Contents

<b>Acknowledgements.....</b>	<b>iii</b>
<b>Attributions.....</b>	<b>iv</b>
<b>Table of Contents .....</b>	<b>v</b>
<b>List of Figures.....</b>	<b>x</b>
<b>List of Tables .....</b>	<b>xii</b>
<b>Definition of Key Terms .....</b>	<b>xiv</b>
<b>Chapter 1. Introduction .....</b>	<b>1</b>
1.1 Virtual Reality Background.....	2
1.2 3D Interaction Background.....	2
1.3 Problem Statement.....	3
1.4 Research Questions.....	4
1.4.1 Research Question 1 .....	4
1.4.2 Research Question 2 .....	4
1.4.3 Research Question 3 .....	4
1.5 Approach.....	5
1.5.1 For RQ1 .....	5
1.5.2 For RQ2 .....	5
1.5.3 For RQ3 .....	5
1.6 Contributions .....	6
<b>Chapter 2. Related Work.....</b>	<b>7</b>
2.1 Immersive VR in Scientific Visualization.....	7
2.2 General Studies on the evaluation of the effects of VR system fidelity .....	8
2.3 Controlled Experiments for studying effects of Components of Display Fidelity .....	10
2.4 The MR Simulation Approach and Studies of its Validity .....	12
2.5 Visualization and Interaction Techniques for Volume Data analysis .....	13
2.6 Brief survey of Bimanual Interaction Techniques, and its use for Volume Data Analysis.....	14
2.7 Generic and High Level Visualization Task Taxonomies.....	15
2.8 Conclusion .....	16
<b>Chapter 3. VR Study-I: Effects of Display Fidelity on Visual Analysis of Volume Data.....</b>	<b>17</b>

3.1	Experiment.....	17
3.1.1	Goals and Hypotheses.....	17
3.1.2	Datasets.....	18
3.1.3	Apparatus.....	19
3.1.4	Tasks.....	20
3.1.5	Exploratory Study.....	21
3.1.6	Design.....	21
3.1.7	Participants.....	22
3.1.8	Procedure.....	23
3.2	Results.....	24
3.2.1	Grades (accuracy in task performance).....	25
3.2.2	Task completion time.....	28
3.2.3	Perceived difficulty (subjective rating).....	29
3.2.4	Confidence in response (subjective rating).....	31
3.2.5	Post-questionnaire results.....	32
3.2.6	Effects of spatial abilities of the participants.....	32
3.2.7	Summary of important results.....	32
3.3	Discussion.....	32
3.3.1	Effects of display fidelity on task categories.....	33
3.3.2	Generalizability of the results.....	34
3.3.3	Implications for design.....	34
3.4	Conclusion.....	35
<b>Chapter 4. VR Study-II: Validation of the MR Simulation Approach for Evaluating the Effects of Display Fidelity on Visual Analysis of Volume Data.....</b>		<b>36</b>
4.1	Experiment.....	36
4.1.1	Goals and Hypotheses.....	36
4.1.2	Datasets.....	37
4.1.3	Apparatus.....	38
4.1.4	Tasks.....	39
4.1.5	Design.....	40
4.1.6	Participants.....	41
4.1.7	Procedure.....	41

4.2	Results.....	42
4.2.1	Significant results from current experiment .....	43
4.2.2	Comparison of MR simulator platforms.....	46
4.3	Discussion.....	50
4.3.1	MR Simulation Validity .....	50
4.3.2	Effects of FOR.....	51
4.3.3	Implications for design .....	51
4.4	Conclusions.....	52
<b>Chapter 5. VR Study-III: Effects of VR System Fidelity on Analyzing Isosurface Visualization of Volume Datasets .....</b>		<b>53</b>
5.1	Experiment.....	53
5.1.1	Goals and Hypotheses.....	53
5.1.2	Datasets.....	55
5.1.3	Apparatus .....	56
5.1.4	Tasks .....	57
5.1.5	Design .....	58
5.1.6	Participants .....	58
5.1.7	Procedure .....	59
5.2	Results.....	59
5.2.1	Grades (task performance accuracy) .....	60
5.2.2	Completion time (speed of task performance).....	61
5.2.3	Perceived levels of difficulty (subjective metric).....	61
5.2.4	Confidence levels in response (subjective metric) .....	61
5.2.5	Post questionnaire results (subjective ratings).....	64
5.3	Discussion.....	66
5.3.1	Interaction effects of the components of fidelity .....	66
5.3.2	Effects of VR system fidelity in different task categories .....	68
5.3.3	Effects of VR system fidelity with different rendering styles .....	69
5.3.4	Implications for design .....	70
5.4	Conclusion .....	70
<b>Chapter 6. 3DI Design-I: The Volume Cracker Technique.....</b>		<b>72</b>
6.1	Volume Cracker Design .....	72

6.1.1	Goals .....	72
6.1.2	Design .....	73
6.2	Evaluation .....	76
6.2.1	Goals and Hypothesis .....	76
6.2.2	Preliminary Evaluation by Experts .....	76
6.2.3	Study Design .....	77
6.2.4	Implementation .....	77
6.2.5	Participants .....	78
6.2.6	Tasks .....	78
6.2.7	Procedure .....	79
6.2.8	Results .....	80
6.3	Discussion .....	81
6.3.1	Novice vs. Expert Participants .....	82
6.3.2	Simulated Datasets vs. Real Datasets .....	83
6.4	Conclusion .....	83
<b>Chapter 7. 3DI Design-II: The Bare Hand Volume Cracker Technique .....</b>		<b>84</b>
7.1	Bare-Hand Volume Cracker .....	85
7.1.1	Overcoming prior problems with the Volume Cracker .....	85
7.1.2	Design Challenges with the Leap .....	85
7.1.3	Bimanual-symmetric interface design for BHVC .....	86
7.1.4	Evaluation of the bimanual-symmetric design of BHVC .....	87
7.1.5	Bimanual-asymmetric interface design for BHVC .....	88
7.2	Evaluation .....	91
7.2.1	Study Design .....	91
7.2.2	Goals and Hypotheses .....	91
7.2.3	Hardware and Software .....	92
7.2.4	Implementation .....	93
7.2.5	Datasets .....	94
7.2.6	Tasks .....	94
7.2.7	Participants .....	95
7.2.8	Procedure .....	95

7.2.9 Results.....	96
7.3 Discussion.....	98
7.4 Conclusion and Future Work.....	100
<b>Chapter 8. Task Taxonomy .....</b>	<b>101</b>
8.1 Introduction.....	101
8.2 Designing the Visual Analysis Task Taxonomy .....	102
8.2.1 Proposed task taxonomy .....	103
8.3 Evaluating the design of the taxonomy .....	105
8.3.1 Overview of Survey .....	105
8.3.2 Results and Discussion .....	106
8.4 The Role of the Task Taxonomy in our VR and 3DI Evaluation studies.....	111
8.5 Conclusion .....	111
<b>Chapter 9. Conclusion and Future Work.....</b>	<b>113</b>
9.1 Primary Contributions .....	113
9.1.1 Effects of VR System Fidelity.....	113
9.1.2 Design of 3D Interaction Techniques.....	114
9.1.3 Task Taxonomy for Visual Analysis of Volume Data .....	114
9.1.4 Discussion.....	115
9.2 Limitations.....	116
9.3 Impact on Theory, Practice and Design.....	116
9.4 Future Work.....	117
<b>References .....</b>	<b>120</b>
<b>Appendix A. Experiment Documents: VR Study-I.....</b>	<b>129</b>
<b>Appendix B. Experiment Documents: VR Study-II .....</b>	<b>149</b>
<b>Appendix C. Experiment Documents: VR Study-III.....</b>	<b>164</b>
<b>Appendix D. Experiment Documents: Volume Cracker Study .....</b>	<b>177</b>
<b>Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study .....</b>	<b>187</b>
<b>Appendix F. Experiment Documents: Task Taxonomy Survey .....</b>	<b>200</b>

## List of Figures

Figure 1: Various real and simulated volume datasets we used in our studies. ....	1
Figure 2: Difference between a controlled experiment approach (b) and the traditional study approach (a) - (Bowman and McMahan 2007) .....	11
Figure 3: The mixed reality continuum as proposed by Milgram and Kishino (Milgram and Kishino 1994) .....	12
Figure 4: The three volume visualized microscopic computed tomography (micro-CT) datasets used in VR-Study-I.....	17
Figure 5: A participant experiencing the training dataset in the Viscube.....	20
Figure 6: Interaction between FOR and ST for M2 grade.....	26
Figure 7: Interaction between FOR and HT for M4 grade.....	26
Figure 8: Interaction between FOR and HT for F5 grade.....	27
Figure 9: Interaction between ST and HT for F1 grades.....	27
Figure 10: Interaction between FOR and ST for F5 time (in sec).....	27
Figure 11: Interaction between FOR and ST for F6 time (in sec).....	27
Figure 12: Interaction between FOR and HT for M2 time (in sec).....	27
Figure 13: Interaction between FOR and HT for M4 time (in sec).....	27
Figure 14: Interaction between FOR and HT for $\Sigma F$ time (in sec).....	27
Figure 15: Participants using two different Mixed-Reality (MR) simulator platforms in our VR studies I and II.....	36
Figure 16: The micro-CT datasets used in our studies.....	38
Figure 17: NVis SX111 Head Mounted Display .....	39
Figure 18: Interaction between FOR and HT for M4 grade .....	44
Figure 19: Interaction between FOR and HT for F4 grade.....	44
Figure 20: Box plot comparison of grades of mouse limb dataset from the CAVE experiment and the recent HMD experiment.....	49
Figure 21: Box plot comparison of grades of fossil dataset from the CAVE experiment and the recent HMD experiment .....	49
Figure 22: Isosurfaces of tracheal systems generated from micro-CT scans of beetles used in our VR system fidelity study.....	53
Figure 23: A participant in the FOR_ST_HT condition inside the CAVE.....	56
Figure 24: Interaction between FOR and HT for Grade in T3, a Quantitative Estimation Task.....	60



Figure 25. Interaction between FOR and HT for Grade in T12, a Pattern Recognition Task. ....	60
Figure 26. Average time in different conditions. ....	66
Figure 27. Average grade in different conditions. ....	66
Figure 28: The Volume Cracker interaction technique.....	72
Figure 29: User inspecting a dataset with Volume Cracker.....	73
Figure 30: The three interaction techniques we compared in our study. ....	77
Figure 31: Performance metrics for search tasks. ....	80
Figure 32: Performance metrics for pattern recognition tasks. ....	81
Figure 33: Bare Hand Volume Cracker – $\mu$ -CT scan of a beetle, cracking preview and cracked beetle, in opaque and x-ray modes. ....	84
Figure 34: Leap Motion controller.....	85
Figure 35: The bimanual-symmetric interface of the BHVC .....	86
Figure 36: A 3D mouse - the Space Navigator pro.....	88
Figure 37: <i>Turning-the-pages</i> with BHVC - the synchronous shrinking and expanding of the sub-volumes with hand movement. ....	89
Figure 38: Hiding a sub-volume - point-to-select a sub-volume, hit the escape key to hide, and remove the tracked hand. ....	90
Figure 39: Turning-the-pages in BHVC with one of the sub-volumes hidden. ....	91
Figure 40: The arbitrary slicing (AS) and the axis-aligned slicing (AAS) interfaces.....	93
Figure 41: A participant using the asymmetric BHVC interface.....	95
Figure 42: Average grades (with standard error bars) secured by the participants with the three interaction techniques.....	97
Figure 43: Current design of the task taxonomy framework .....	107
Figure 44: Box plot showing the match between the task description and the survey respondents' expectations.....	108
Figure 45: Box plot showing the respondents' division of time between the different tasks from our survey .....	109
Figure 46: Exploring the fibers in a micro-scaffold volume dataset.....	110
Figure 47: Analyzing the tracheal tubes in volume visualized micro-CT scan of a <i>Platynus</i> beetle.....	110

## List of Tables

Table 1. Conditions Experienced by the Eight Groups in the Experiment .....	22
Table 2. Relative Weights of the Tasks in each Dataset .....	24
Table 3. Significant Main Effects on Grades .....	25
Table 4. Significant Interaction Effects on Grades .....	25
Table 5. Significant Interaction Effects on Time .....	28
Table 6. Significant Main Effects on Difficulty .....	29
Table 7. Significant Interaction Effects on Difficulty .....	29
Table 8. Significant Main Effects on Confidence .....	31
Table 9. Significant Interaction Effects on Confidence .....	31
Table 10: Primary differences between the two MR simulator platforms .....	38
Table 11. Conditions Experienced by the Eight Groups in the Experiment .....	40
Table 12. Relative weights of tasks and abstract task categories .....	42
Table 13. Significant Main Effects on Grades .....	43
Table 14. Significant interaction effects of FOR and HT on confidence .....	44
Table 15. Significant effects from the two experiments. ....	46
Table 16. P-values for likelihood ratio and Pearson Chi-square .....	47
Table 17. P-values for Chi-square statistic in Mann-Whitney test .....	48
Table 18. Conditions experienced by the eight groups in the experiment, and their case-sensitive labels used in this paper .....	58
Table 19. Significant Main Effects on Grades .....	60
Table 20. Significant Interaction Effects on Grades .....	61
Table 21. Significant Main Effects on Time .....	62
Table 22. Significant Interaction Effects on Time .....	62
Table 23. Significant Interaction Effects on Perceived Difficulty of Tasks .....	63
Table 24. Significant Interaction Effects on Perceived Confidence .....	64
Table 25. Significant Interaction Effects on Post Questionnaire Ratings .....	65
Table 26. 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 .....	67
Table 27. Re-defining categories of tasks from our previous study (Laha <i>et al.</i> 2012) .....	69

Table 28. Mean errors, scores, and time taken for the task types .....	81
Table 29: Effects of interaction technique on grades.....	97
Table 30: Tasks with abstract categories and difficulty levels .....	100

## Definition of Key Terms

Here we define few important terms recurring in this document:

Volume Data or Volumetric Data - Data composed of voxels (x, y, z, v) on a 3D grid instead of 2D pixels, where each voxel has a numeric value based on some property of the object it is representing, like color, density, refractive index, or other material properties.

Volume Rendering or Volumetric Rendering - Technique(s) for rendering volume data such as decomposition, isosurface rendering, maximum intensity projection, semi-transparency, x-ray rendering, etc.

Region of Interest (ROI) - Region of a volume data that is of particular interest to a scientist, researcher, or a medical practitioner, due to the structure(s) lying within that region.

Segmentation - Manual or automatic process of marking out ROI in every slice of a volume, through the entire length of the volume data.

3D Interaction - Human-computer interaction in which the user's tasks are performed directly in a 3D spatial context.

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.

Head Tracking (HT) - The display of images based on the physical position and orientation of the user's head (produced by head tracking).

Stereoscopy or stereo - The display of different images to each eye to provide an additional depth cue.

Immersion - The objective level of sensory fidelity produced by a system.

Display Fidelity or Visual Immersion - Characteristics of the physical display, which can be broken into components such as field of regard, stereoscopy, display size, etc.

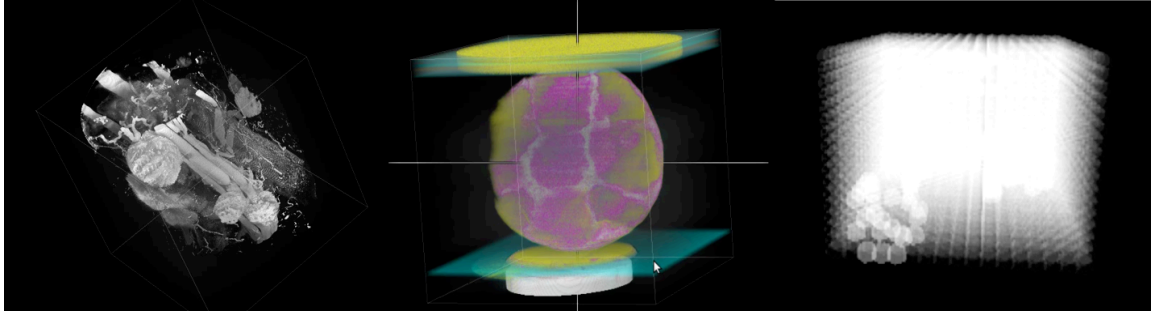
System Fidelity - The faithfulness of the system to match the real world through the fidelity of its various components; could be broken down into display and interaction fidelity.

Virtual Reality (VR) or Virtual Environment (VE) - 3D Immersive and Interactive computer generated environment.

Mixed Reality (MR) continuum - Range of displays the human eyes can perceive, connecting the real world on one end, and the virtual environment or virtual reality on the other end.

MR simulation - Simulation of any display on the MR continuum based on the levels of the various components of display fidelity characterizing the display.

MR platform - A VR hardware platform capable of producing various levels of the different components of display fidelity, used for MR simulation.



(a) A partially segmented medical data; (b) A raw and unsegmented fossil data; (c) A simulated raw volume

Figure 1: Various real and simulated volume datasets we used in our studies.

## Chapter 1. Introduction

Volume data is generated through processes such as computed tomography, magnetic resonance imaging, ultrasound, confocal microscopy, positron emission tomography, etc. (Kaufman 1990). Volume data, in its raw form, is composed of voxels ( $x, y, z, v$ ) 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, mapped in the rendering using some transfer function (Kaufman 1996). There are many techniques for voxel rendering such as decomposition, isosurface rendering, maximum intensity projection, semi-transparency, and x-ray rendering (Marmitt et al. 2006).

Visualization of volume data has been an important tool in a variety of application domains. Applications include geophysics, medical biology, paleontology, astronomy, biomechanics, solid modeling, etc. 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.

For effective analysis of a 3D volume (see Figure 1), users often need to look at the rendering from various viewpoints. For example, in Figure 1-a, we can see the top of a blood vessel at the periphery towards the back. If we want to know how it went through the volume, we will need to look from the back, by rotating the volume  $180^\circ$  about the vertical axis. Understanding a volume dataset thus requires the users to mentally integrate the various views to construct a 3D mental model of the volume.

Scientists also often need to peer inside the volume. This process involves finding the right voxels representing the structure they are looking for, either by disregarding or discarding the unwanted voxels, and then joining all the right voxels to understand the structure. For example, in Figure 1-c, we can see that there are a few voxels, more opaque than the others, towards the bottom left corner, which appear to be joined to a few more voxels inside the simulated volume. But we really cannot understand the structure as a whole, because the rest of the voxels are blocked from our view. To see the entire structure, we can remove the voxels from the front blocking our view to see the voxels hiding behind. The analysis could also involve understanding in context, when the other voxels and relevant neighboring structures play an active role in the analysis.

The following two sections describe the prior art in virtual reality and 3D interaction, related to this work. A detailed treatment is available in Chapter 2.

## 1.1 Virtual Reality Background

Traditionally, researchers analyze 3D volume renderings on desktop computer monitors or larger displays, which lack the more realistic visual stimuli found in more advanced systems, 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 systems with greater realism can be said to have a higher level of display fidelity or immersion, defined as the objective level of sensory fidelity provided by a system (Slater 2003). “Immersive” virtual reality (VR) systems produce sensory stimuli closer to those we experience in the real world. These systems may include fishtank VR (Arthur et al. 1993), head-mounted displays (HMDs), and surround-screen projection systems like the CAVE (Cruz Neira et al. 1993). The level of system fidelity is determined by a number of components e.g., field of regard, field of view, display size, display resolution, stereoscopy, realism of lighting, frame rate, and refresh rate (Bowman and McMahan 2007).

Immersion is related to display fidelity (McMahan et al. 2012), and different VR systems vary widely in their levels of immersion. Although in the VR community, we have differentiated *immersion* from *presence* very clearly (Slater 2003; Bowman and McMahan 2007), these terms could still confuse people sometimes, like in the phrase - “immersed in the virtual world”. To avoid such confusions, we shall use *system fidelity* while referring to the objective levels of the system components, and how closely they match with their real world counterparts. We shall further break down system fidelity into *display* and *interaction fidelity*, while referring specifically to the *display* and *interaction* components of a system.

Since VR systems with high levels of display fidelity can be costly and complex, decision makers need evidence for the benefits of display fidelity if they are to choose such a system. For researchers, understanding the effects of system fidelity (realism) is one of the fundamental questions in the field. Many people have suggested using virtual reality (VR) systems with higher levels of system fidelity for scientific visualization, including the analysis of volume data, since immersive VR is designed to display spatially complex structures in a manner easier to understand and explore. VR researchers have run experiments showing evidence of better task performance with CAVE over desktops for analyzing volume datasets (Zhang et al. 2001; Zhang et al. 2003; Prabhat et al. 2008). But since they failed to tie the significant findings with the components of system fidelity, such results cannot be generalized beyond the specific hardware platforms used in the studies. Also, since the tasks were specific to the datasets used, it is hard to generalize the findings beyond the domains from where the data originated.

## 1.2 3D Interaction Background

Scientists have been using mouse and keyboard interfaces for decades for analyzing volume datasets. The 2D movements possible through the mouse and keyboard allow them to explore the internal structures in a volume along any given axis. But often, the internal structures in a

volume dataset may not be axis-aligned. The outlines and the boundaries of the structures, as well as the gaps and the connections between them occur in 3D. The mismatch between the degrees of freedom required in exploring and that provided by 2D modalities might be fundamental to many of the problems inherent to volume data analysis.

3D interaction allows us to give inputs to a system in a 3D spatial context. There is anecdotal evidence showing the benefits of using 3D interaction techniques like two-handed (bimanual) interaction for interacting with volume data. But very little work has been done on developing and standardizing the various tools and techniques that can be used by scientists for the different tasks they perform with volume data. Also, there is very little empirical evidence supporting cases in which those techniques might be useful. As a result, interaction with and analysis of volume datasets have been relying on existing approaches. These are standardized over the last few decades, and are mostly designed with the 2D desktop display in mind.

The most standard method for analysis is called *segmentation* (3D-Slicer 2012; Amira 2012; GE-Measurement-and-Control-Solutions 2012; Xradia-Solutions 2012), in which a transfer function for mapping material properties is adjusted manually or automatically based on some thresholds, to mark out the region of interest (ROI) in the volume accurately. Manual segmentation is accurate but quite time-consuming. Corporate Labs like Mayo Clinic employ several full-time professionals for manual segmentation. Segmentation could also be done automatically, but often times the accuracy is less than desired. Thus, it would be useful to have alternative techniques that allow analysis of volume data and viewing of an ROI without requiring segmentation.

One alternative is to remove all the voxels that occlude the important ones. Traditionally, scientists have used the standard orthogonal or axis-aligned slicing (AAS) technique, which gives axial, coronal, and sagittal slice views (3D-Slicer 2012; Amira 2012; GE-Measurement-and-Control-Solutions 2012; Xradia-Solutions 2012). Researchers have also proposed more flexible slicing techniques, which, although very useful, suffer from the common problem of losing context as they remove part of the volume.

Focus+Context visualization and interaction techniques have arisen as a result, aiming to preserve the whole model while focusing on the details. But all these proposed techniques either assume prior definition of the ROI by the user, or distort the ROI at least partly in relation to the neighboring structures. Some techniques assume that semantic layers are defined for a volume dataset (McGuffin *et al.* 2003), but this is not typically the case unless segmentation has already been performed. Such assumptions for creating 3D interaction techniques might increase the gap between our research and the actual tools used by real-world domain scientists.

### 1.3 Problem Statement

Immersive VR and 3D interaction may be useful for volume data analysis; however, two specific problems remain:

1. It is not known what the effects of system fidelity are when analyzing volume data.
2. Existing 3D interaction techniques for volume visualizations do not allow the user to examine internal structures effectively while maintaining awareness of the overall spatial context.

## 1.4 Research Questions

This dissertation research is focused on addressing these problems by gathering empirical evidence for the effects of the components of system fidelity and novel 3D interaction techniques for improving analysis of volume visualizations. We have broken down this thesis statement into the following research questions.

### 1.4.1 Research Question 1

*How do various system characteristics or components of system fidelity impact the effectiveness of volume visualizations?*

Prior research has shown evidence of improved task performance with very high levels of system fidelity (e.g., CAVE) over desktop systems. But if we can achieve a sufficient level of accuracy in task performance for analyzing volume datasets with systems having a few components of system fidelity at higher levels, and others at lower levels, then a CAVE may not be necessary. The price of fully immersive hardware being significantly higher than other non-immersive commodity hardware, it is important to find the individual and combined effects of various components of system fidelity, both quantitatively and qualitatively, so that the cost to benefit ratio of acquiring such hardware is in the favor of the domain scientists, and is based on objective measures. To address these questions, we need to run controlled experiments, with varying components of system fidelity, and measure task performance under different conditions we want to compare. We run such studies using the MR simulation approach, which employs high-end VR platforms to simulate systems with various levels of system fidelity. Thus, it is also important to understand whether such experiments give valid results; i.e., we need to determine if different *MR simulator platforms* are able to produce similar results in task performance.

### 1.4.2 Research Question 2

*Can we design 3D interaction techniques for volume data analysis that will result in increased effectiveness as compared to traditional interaction techniques and existing 3D interaction techniques?*

The tasks that scientists perform with volume datasets are mostly 3D in nature, but traditionally they use interfaces based on a mouse and keyboard to interact with these volume datasets. These offer a 2D interaction medium, which may take more time just to get the correct viewpoint in 3D, and/or to manipulate the 3D volume datasets in various ways. There is a need to develop and explore novel tools and techniques of 3D interaction specifically for enabling scientists to easily navigate through and analyze their volume datasets in easy and intuitive ways. Further, we need to evaluate the effectiveness of these techniques for task performance and usability through empirical studies comparing them to existing approaches and techniques.

### 1.4.3 Research Question 3

*What is the best way to describe the tasks in volume visualizations such that the effects of system characteristics and novel 3D interaction techniques for a task would be general across multiple domains?*

To generalize the empirical results of controlled experiments run with volume datasets from particular domains, we need an abstract *task taxonomy* that cuts across the various domains from which volume datasets come. This task taxonomy is expected to have task categories



encapsulating the various tasks that domain scientists carry out during their daily research work. The task categories will be high-level abstractions (e.g., search tasks), which will make sense in fundamentally distinct domains (e.g., biomedical sciences and geophysics). Such taxonomy will allow VR and 3DI researchers to run empirical studies with non-expert participants but have results valuable to domain experts. This categorization will also allow domain scientists to relate closely to the empirical findings from our community, more than they would to our previous findings tied to fundamental tasks such as rotation, translation, scaling, etc.

## 1.5 Approach

### 1.5.1 For RQ1

To address the first research question, we ran three controlled experiments, comparing the different levels of system fidelity for tasks performed with volume datasets coming from at three different scientific domains: paleontology, medical biology, and biomechanics. The first experiment was performed in the CAVE, the second with the Nvis nVisor SX11 head-mounted display (HMD), and the third in the CAVE. From these experiments, we got insights into the effects of different components of system fidelity on abstract task categories of volume data analysis.

### 1.5.2 For RQ2

To address the second research question, we designed and developed the Volume Cracker technique, through several design studio sessions involving brainstorming, storyboarding, and design iterations. We first designed a prototype of the Volume Cracker with 5DT data gloves and IS-900 hand and head trackers. We evaluated it with simulated datasets, and novice participants. In the first experiment, we compared our volume cracker tool to two other interaction techniques for analyzing volume data. The first one is most widely used desktop approach – axis-aligned slicing – with three slicing planes oriented along three orthogonal axes for cutting through the data. The other one is the most widely known approach for analyzing internal structures in a volume – a 2D slice plane used as a cutting tool. The tasks were to search for randomly oriented letters of the alphabet, hidden in a simulated volume dataset, and to find the odd one out of three curved 3D structures. We used simulated voxels of varying density levels to represent the structures inside the volume. We found that the *volume cracker* had significant advantages over both of these for search and pattern recognition tasks.

Based on the results, we developed a second prototype with the Leap Motion controller. We evaluated it with real-world datasets and expert participants.

### 1.5.3 For RQ3

To address the third research question, we developed a volume data analysis task taxonomy primarily by interviewing scientists, researchers, and faculty from various scientific domains, who work on a daily basis with volume datasets. We refined the design through workshop position papers, interviewing and talking to more domain scientists and researchers, and by running a survey reaching out to a wider population of domain scientists, researchers, and graduate students who work with volume data on a daily basis.

## 1.6 Contributions

Our contributions can be divided in three main parts:

1. The VR controlled studies have produced empirically supported quantitative and qualitative insights on how system fidelity affects volume visualizations.
2. The design and evaluation of the Volume Cracker produced usability criteria and empirical guidelines to improve the effectiveness of interactive volume visualizations.
3. The task taxonomy framework helps to generalize empirical findings to multiple domains, allows non-expert participants in empirical experiments, and produces results in empirical studies that experts can easily relate to.

In addition to these specific contributions, our work provides the start of a three-way many-many-many mapping between the sets of components of system fidelity, 3D interaction tools and techniques, and volume visualization tasks, which could be used as a framework of reference for future research.

## Chapter 2. Related Work

In this chapter we will attempt to take a comprehensive look at the existing literature for using immersive VR for scientific and volume visualization, the MR simulation approach and its validity, 2D and 3D interaction techniques for volume data analysis, and attempts at creating task taxonomy for volume data analysis.

### 2.1 Immersive VR in Scientific Visualization

We can define Virtual Reality (VR), or Virtual Environment (VE) as an immersive and interactive computer generated environment (Sherman and Craig 2003; Bowman *et al.* 2005). Immersion (or VR system fidelity in this dissertation) is defined by Slater (Slater 2003) as the objective level of sensory fidelity produced by a system. Interactivity of a system would mean the ability to produce feedback or response in real-time to the user. Ivan Sutherland is often given the honor of creating one of the first VR systems, back in 1968 (Sutherland 1968). He created a head-mounted display (HMD) connected to the ceiling in a room, and allowed the users to occupy the same space as the virtual objects created with wire-frame graphics. Since then several VR hardware with various immersive levels have been created, one of the most important being the CAVE (CAVE Automated Virtual Environment) by Cruz-Neira in 1993 (Cruz Neira *et al.* 1993).

VR with its engaging qualities generated a lot of excitement. But beyond just fun, VR researchers were more interested in finding more serious applications of immersive VR. This was also important because even today, the cost of VR hardware remain high. So it is important to justify the cost to benefit ratio of VR hardware, for wider acceptance and use of VR as a technology for scientists and researchers.

As Bryson suggested (Bryson 1996), immersive virtual reality might offer a very good platform for analyzing spatially complex structures, as found in many scientific visualizations. There has been much anecdotal evidence of the benefits of immersive VR for task performance with 3D scientific data, or volume datasets.

Zhang et al. (Zhang *et al.* 2001) performed one 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 (Zhang *et al.* 2003) visualized in a CAVE, Zhang et al. reported that the stereoscopy and interactivity of the VR system aided in understanding complex geometric models.

Not all studies reported benefits of higher levels of display fidelity for analyzing volume datasets. Demiralp et al. (Demiralp *et al.* 2006) compared a fishtank VR with a CAVE for task performance with an abstract visual search task, with noisy “potato” objects. They reported that the users could perform the task significantly faster and with higher accuracy in the fishtank VR, as opposed to the CAVE environment.

Prabhat et al. (Prabhat *et al.* 2008) compared three display systems (desktop, fishtank VR, and CAVE), for various task performance with confocal datasets. They reported significant quantitative and qualitative benefits of the more immersive environments for analyzing volume

data. However, we were not aware of any controlled experiments evaluating the effects of individual components of display fidelity on the effectiveness of volume visualizations. Our first study filled that gap, and reported several effects of three components of display fidelity for analyzing volume visualized micro-CT datasets (Laha *et al.* 2012). We report the results of this study in detail in Chapter 3.

More recently, Chen *et al.* reported the effect of stereo and display size on five representative tasks performed with dense diffusion MRI (DMRI) data (Chen *et al.* 2012). They found that task completion time was not improved by the larger display, and stereo reduced accuracy of performance.

## 2.2 General Studies on the evaluation of the effects of VR system fidelity

For the last two decades, VR researchers have been running evaluation studies comparing VR systems with traditional desktop systems to evaluate the effects of VR system fidelity.

One of the first such studies was ran by Randy Pausch in 1993 (Pausch *et al.* 1993) using a HMD display system. He found a 42% reduction in time of a generic search task when performed with a HMD compared to a stationary display, with comparable display properties (resolution, field of view, etc.). He also found that training with HMD transferred to the stationary display with a 23% reduction in task completion time.

Arthur *et al.* reported results of experiments ran in 1993 with a fishtank VR system, which used a standard desktop display coupled with stereo and head tracking (Arthur *et al.* 1993). They found that head tracking with stereo produced an order of magnitude improvement in task performance with a tree-tracing task over a monoscopic and non-head tracked display. They also found that head tracking provided higher benefits than stereo. They also found that latency had a higher impact on task performance than lower frame rate.

On similar lines, Ware and Franck reported results on two experiments in 1996 (Ware and Franck 1996), for visualizing information net in 3D graphs. They found that head tracking combined with stereo improved task performance of perceiving abstract data in 3D, by a factor of 3. Head tracking alone produced an increase by a factor of 2.2, while stereo produced an increase by a factor of 1.6. They also found that structured 3D motion was more significant than stereo cues for improving task performance.

In 1997, Pausch *et al.* compared a HMD with a stationary monitor and a hand-based input device (Pausch *et al.* 1997), for searching the existence of camouflaged targets in a virtual room. They found that the desktop users took 41% more time than the HMD users, because they re-examined the areas they already searched. The researchers also found that the training transferred positively from HMD to desktop, but negatively from desktop to HMD.

Extending the findings of Pausch *et al.* to a desktop VR, Robertson *et al.* found (Robertson *et al.* 1997) that the time taken for searching targets, when absent matched the times predicted by standard desktop. In other words, the findings of Pausch *et al.* (Pausch *et al.* 1997) using HMD did not transfer to desktop VR.

Not all empirical studies found benefits of stereo and/or head tracking. In 1997, Barfield *et al.* ran a short study (Barfield *et al.* 1997) with twelve participants in which the participants were shown the virtual image of a curved wire, and then identified the correct representation of the

wire from one of three drawings. They found that neither stereo nor head tracking improved accuracy of their identification of the correct paper representation of the wire.

Later in 1999, in one of the first studies comparing the CAVE to other display technologies, Arns *et al.* 1999) for three structure detection tasks, and an ease of interaction task. They found that CAVE was better than the desktop for the structure detection tasks, while the participants preferred the desktop to the CAVE for the interaction task.

Ruddle *et al.* compared an HMD with a desktop display (Ruddle *et al.* 1999) for navigating virtual buildings with more than seventy rooms. They found that participants using the HMD developed significantly more accurate sense of relative straight-line distance than the participants using the desktop display. They also reported that participants took advantage of the head tracking provided by the HMD to look around in the VEs more than the participants did with the desktop monitor.

Ameya Datey in 2002 reported results of an experiment studying the effects of immersion on task performance in VR (Datey 2002). Participants in his study used either a desktop or a HMD, and a wand to detect dependency of radioactivity of objects, by navigating in a VE and moving a probe in different directions. Although there were no statistically significant results, he found that users generally performed better with the HMD than the desktop in terms of time of completion of tasks, and number of errors made.

Tan *et al.* compared large wall-sized displays with standard desktop monitor (Tan *et al.* 2003), for a reading comprehension task, and a spatial orientation task. They kept the visual angle (FOV) same by adjusting the viewing distance to each display. They found that users performed 26% better on the large display in a spatial orientation task. They thought that the improvement in performance could be related to higher presence of the users in the large display, for which they treated the spatial task as egocentric, as opposed to exocentric in typical desktop monitors.

Midttun *et al.* from Norway's Norsk Hydro (Midttun *et al.* 2000) build a CAVE system with three walls and a floor, for use in designing/drilling production wells and seismic exploration. They claimed that the VR system produced measurable improvements in many of their work processes by lowering cycle time costs, and increasing recovery. On similar lines, Gruchalla compared the CAVE with a desktop workstation for a well path planning task (Gruchalla 2004), which is a spatially complex three dimensional task. He found that fifteen out of sixteen participants could complete the task faster with the CAVE than with the desktop. He also found that the accuracy of the task performance was higher with the CAVE than in the desktop environment.

Shiratuddin *et al.* compared five different display environments for analyzing construction 3D models (Shiratuddin *et al.* 2004). Their experiment included a CAVE, an HMD, an Immersive-Workbench (IWB), a 1.5 meter hemispherical display, and a desktop monitor. The results of their study indicated higher immersive systems were better suited for analysis of 3D models.

Ball and North compared a large tiled display, with a resolution of 3840x3072 to two smaller displays, of resolutions 1560x2048, and 1280x1024 (Ball and North 2005). Based on the results of an exploratory study, they concluded that for finely detailed data, higher resolution displays (using physical navigation) significantly outperform displays with smaller resolution (requiring zoom and pan navigation).

Narayan et al. used an HMD, and a CAVE to explore how immersive VEs could improve collaborative task performance (Narayan *et al.* 2005). They compared different levels of stereo and head tracking, and found that stereo could have a positive effect on task performance.

Researchers have also compared different levels of display fidelity using a high immersive virtual reality system like the CAVE, to find effects of display fidelity on task performance. In one such study, Schuchardt and Bowman compared the front wall of a CAVE to its all four walls (three walls and the floor), for visualizing underground cave systems, which involved spatial understanding of complex 3D structures (Schuchardt and Bowman 2007). In their study, which involved 24 domain experts performing a variety of search and measurement tasks, they found that for detailed feature search tasks, the high immersive condition was significantly faster and more accurate than the lower display fidelity condition. In general, they found that the higher display fidelity condition improved accuracy and speed of overall task performance.

All these empirical results, although very useful in showing benefits of more immersive systems, has the primary drawback of analyzing the overall effects of display fidelity by comparing immersive vs. non-immersive systems. They do not evaluate the effects of individual components of display fidelity by controlling other factors (keeping them same). So they fail to determine which components of display fidelity resulted in the observed benefits, and do not evaluate whether a system with a moderate level of display fidelity might be sufficient to obtain the benefits. To address such questions, researchers use controlled experiments, which we discuss next.

### **2.3 Controlled Experiments for studying effects of Components of Display Fidelity**

In 2004, Bowman and Raja described a method of separating the components of display fidelity to separately study their effects on task performance, keeping the levels of other components of display fidelity constant (Bowman and Raja 2004). Their method used the CAVE as the VR platform. They suggested that by varying the number of walls of the CAVE displaying content at any time, we can change the level of display fidelity, and by enabling or disabling the head tracking of the user, we can control head based rendering. More recently, Bowman and McMahan also suggested (Bowman and McMahan 2007) that by using a high immersive VR hardware, like the CAVE or a HMD, we can evaluate the effects of individual components of system fidelity (like FOV, FOR, stereoscopy, head-based rendering, display size, etc.), while holding the other factors at a constant level. Their approach is shown in Figure 2.

This theory was based on the definition by Slater (Slater 2003) that immersion is the objective level of sensory fidelity produced by a system (implying that immersion is also measurable objectively and quantitatively), combined with the observation that the sensory fidelity of a system can be separated out in individual components of immersion, as described in (Bowman and McMahan 2007). It is important to note here that the concept of immersion, as defined here, is different from presence, which is the user's physiological response to a system, or the sense of "being there", which is not measurable, and can vary depending even on the time of the day, and the mood of the user.

Using this approach of controlled experimentation, researchers have gathered evidence of the effects of individual components of immersion or the components of display fidelity. In 2000, Kevin Arthur reported the results of a series of experiments carried out with a custom wide-FOV

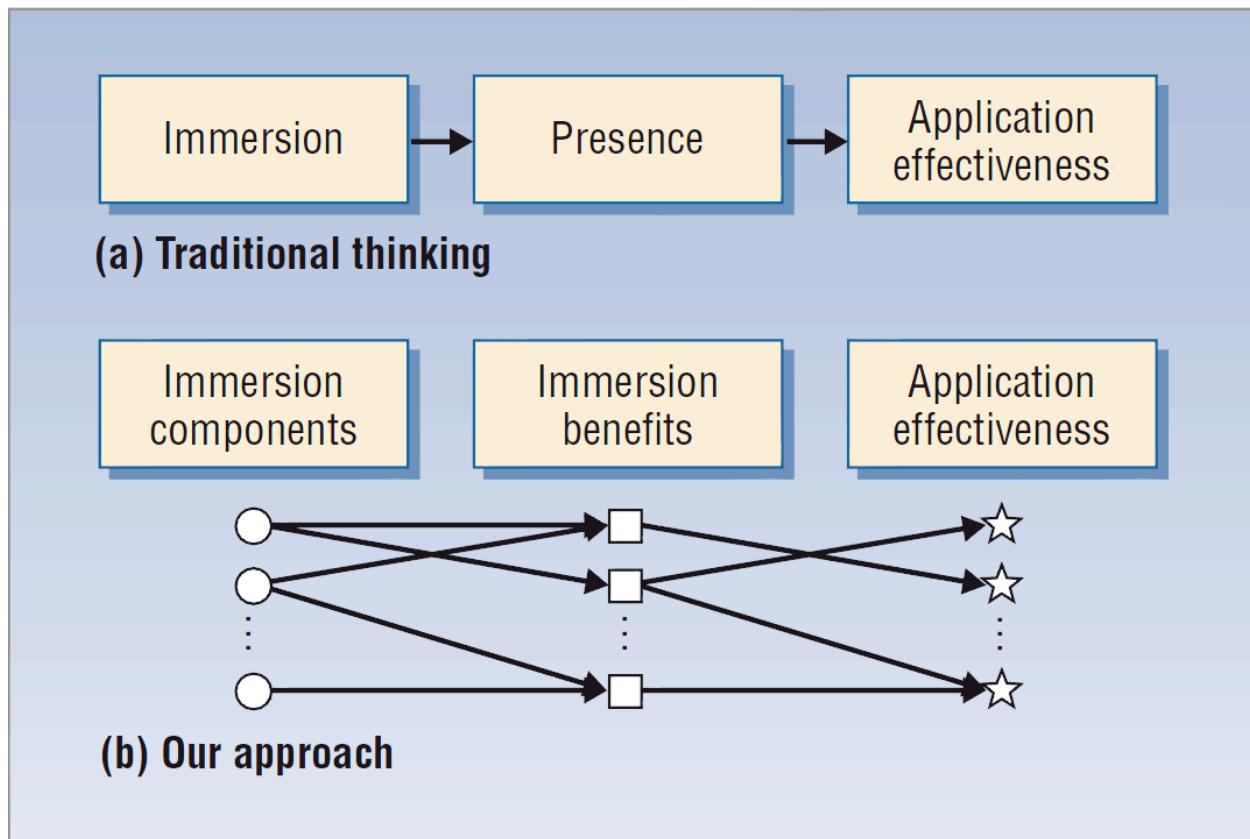


Figure 2: Difference between a controlled experiment approach (b) and the traditional study approach (a) - (Bowman and McMahan 2007)

(field of view) HMD (Arthur 2000), for a search task, and navigating through a simple maze like environment, while avoiding walls. He found that for both these tasks, the performance degraded from FOV 176° to 112° and further from 112° to 48°.

In another such controlled experiment, Ni et al. studied the individual and combined effects of display size, and resolution (Ni *et al.* 2006) in an information rich virtual environment (IRVE), using a flat panel LCD, a rear-projected screen, and a high-resolution display module. They found that higher resolution displays were better suited for search and comparison tasks in IRVE.

Polys et al. reported the results of another controlled experiment evaluating the effects of two information layout techniques, and two levels of software field of view (SFOV, with vertical angles of 60° and 100°) for search and comparison tasks in IRVE (Polys *et al.* 2007), using a single monitor and a tiled nine-panel display. They found that although SFOV didn't affect accuracy of task performance, the users significantly felt that low SFOV conditions were more difficult.

McMahan et al. ran a controlled experiment (McMahan *et al.* 2006) to find the effects of stereoscopy, and field of regard on single-user six degrees of freedom object manipulation task, with a CAVE. They also studied the effects of three 3D interaction techniques in the same study. The results indicated no significant effect of either field of regard or stereoscopy on task performance.

Researchers also ran empirical studies separately looking at the effects of display fidelity and interaction techniques. McMahan et al. reported on an experiment in 2006 (McMahan *et al.*



2006), studying different levels of two components of display fidelity (stereo and head tracking), and three interaction techniques (Homer, Go-Go, Do-It), for a six degrees of freedom manipulation task. They found a significant effect of the interaction technique on the speed of completion of the task. In another experiment, McMahan et al. looked at both display and interaction fidelities (McMahan *et al.* 2012) for strategy and task performance in a VR first-person shooter (FPS) game.

While researchers run controlled experiments to evaluate individual and combined effects of components of display fidelity, we may question whether the results are generalizable across VR platforms. The mixed reality simulation approach promises to answer this question by claiming that the results of VR controlled experiments are tied only to the components of display fidelity of the system and not to the system itself. We shall look at the MR simulation approach next, and the studies supporting its validity.

## 2.4 The MR Simulation Approach and Studies of its Validity

Milgram and Kishino proposed the idea of mixed reality (MR) continuum in 1994 (Milgram and Kishino 1994), as shown in Figure 3. The MR continuum seeks to connect complete real world on one end to complete virtual environment on the other end of the spectrum. Computer systems generating augmented reality are somewhere in the middle of the continuum.

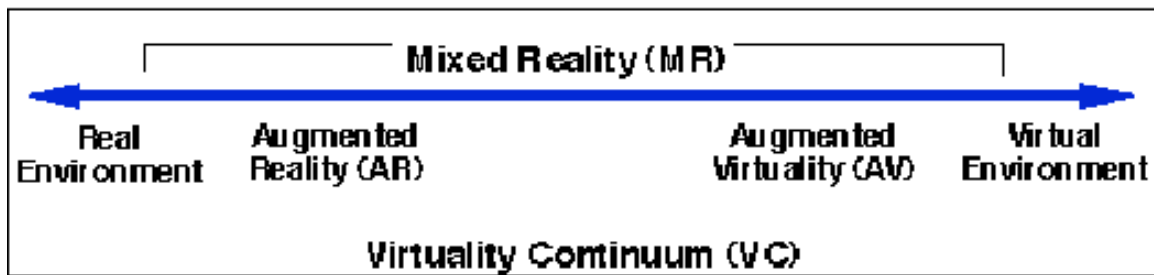


Figure 3: The mixed reality continuum as proposed by Milgram and Kishino (Milgram and Kishino 1994)

Since we can break down any computer system into components of sensory fidelity, and since the components of sensory fidelity are objectively measurable (Slater 2003), in theory we should be able to reproduce any system along this continuum if we can perfectly simulate the levels of sensory fidelity produced by that system. This simulation is termed as MR simulation (Bowman *et al.* 2012). Based on the MR continuum, MR simulation encompasses both VR as well as AR simulation, depending on which part of the spectrum the system belongs that we want to simulate. Thus, it promises to provide generalizable results no matter what hardware platform is used as the MR simulator.

However, some may question the real-world applicability and validity of the results of MR simulation studies. In particular, there is little evidence that experiments run on different MR simulator platforms will produce equivalent results. Prior studies have provided some such evidence by replicating an experiment from the literature and demonstrating that the effects of the simulator's latency were independent of other effects (Lee *et al.* 2010a), and by comparing results from an experiment with a simulated AR system to those from an actual AR system (Lee *et al.* 2012). In our work, we extended this prior work by running an experiment designed to obtain evidence that results from two different MR simulator platforms are similar (Laha *et al.* 2013). We discuss details of this paper in Chapter 4.



In the next section we look at prior work done for visualization and interaction techniques for volume data analysis.

## 2.5 Visualization and Interaction Techniques for Volume Data analysis

Previous approaches to exploring volume data can be broadly divided into two categories. The techniques or tools in the first category remove part of the volume to reveal hidden structures inside. The most standard approach is to use orthogonal or axis-aligned slicing (AAS), in which the user controls cutting planes aligned with the three orthogonal axes. The interface associated with this technique also allows the user to obtain so-called axial, coronal, and sagittal views along the three axes. The AAS technique has been actively and widely used by scientists and researchers around the world for decades as part of various free, open-source, and commercial software packages for volume data analysis, like MITO (medical imaging toolkit) (MITO-DICOM-viewer 2012), 3D Slicer (3D-Slicer 2012), Amira (Amira 2012), Avizo (Avizo 2012), and others. Commercial 3D imaging hardware manufacturers, like Xradia (Xradia-Solutions 2012), GE (GE-Measurement-and-Control-Solutions 2012), GE healthcare (GE-Healthcare 2012), Toshiba (Toshiba-Medical-Systems 2012), Siemens (Siemens-Healthcare 2012), Philips (Philips-CT-Scan 2012) provide the AAS interaction as a de facto technique in the factory software that comes packaged with their hardware.

Hinckley et al. designed a bimanual asymmetric interface allowing arbitrary 3D slicing (Hinckley *et al.* 1994), based on Guiard’s framework (Guiard 1987). Their technique used real-world props: a doll’s head in the non-dominant hand to control the volume data, and a clear plastic plate in the dominant hand for the cutting plane. Going beyond simple cutting planes, researchers have used deformable cutting planes (Konrad Verse *et al.* 2004), clipping based on arbitrary geometry (Weiskopf *et al.* 2003), and a filterbox tool (Mlyniec *et al.* 2011). Others have proposed cutting tools like a rasp and knife for surgery training (Pflesser *et al.* 1995), and scalpels, scissors, and forceps on volumetric meshes (Bruyns and Montgomery 2002). Sculpting metaphors have also been proposed (Wang and Kaufman 1995), and various sculpting tools like cutaway and ghost tools (Bruckner and Groller 2005), and for erasing, digging, and clipping were explored (Huff *et al.* 2006). All these techniques, although allowing the user to explore the volume in very useful ways, cause the user to lose spatial context.

This problem is addressed by the techniques of the second category, which are called Focus+Context techniques. They seek to preserve the entire volume at all times, while letting the user focus on the ROI or the important details in various ways. One such approach is to use a 3D magnification lens metaphor (LaMar *et al.* 2001), such as the Magic Volume Lens (Wang *et al.* 2005), Magicsphere with a MultiRes filter (Cignoni *et al.* 1994), or the Conformal Magnifier (Zhao *et al.* 2012), focusing on closer inspection of ROI in volume data. Some researchers have proposed focal-region-based feature enhancement (Zhou *et al.* 2002), importance-driven volume rendering (Viola *et al.* 2004), or energy optimization with distortion minimization (Wang *et al.* 2008). All these magnifying or enhancing techniques, although highlighting the ROI over the surrounding areas in the volume, distort the volume or the ROI at least partially with respect to its neighboring structures. Also, many of these techniques are intended for use with segmented datasets. Thus they assume a partial and existing solution of the problem they are trying to address.

Other Focus+Context visualization techniques include layered browsing of volume data, promoted through an array of deformation tools (McGuffin *et al.* 2003), and an exploded views

approach (Bruckner and Groller 2006). These assume semantic layers in the volume data, which may not always exist, as discussed by McGuffin et al. (McGuffin *et al.* 2003).

To address all these issues, we developed a 3D interaction technique called the Volume Cracker (Laha 2013). We discuss the design and evaluation of the volume cracker in Chapter 6 and a bare-hand prototype in Chapter 7. The design of the Volume Cracker relied on using the two hands of the user for interaction with a volume. So in the next section we will briefly look at the prior work done in bimanual interaction techniques.

## **2.6 Brief survey of Bimanual Interaction Techniques, and its use for Volume Data Analysis**

Bimanual interactions could be broadly classified as symmetric or asymmetric (Guiard 1987). Bimanual interactions could also be synchronous (where both hands complete the assigned tasks at the same time, like stretching a rubber band) or asynchronous (both hands perform the corresponding tasks at different times, like pulling a rope).

Researchers have argued that the two hands may not always work in parallel. Guiard presented a widely accepted framework for asymmetric bimanual interactions (Guiard 1987), suggesting that the dominant hand follows the reference frame provided by the non-dominant hand, the dominant hand works on a finer spatial and temporal scale, and that the non-dominant hand leads while the dominant hand follows. Kabbash et al. used Guiard's framework to create bimanual asymmetric techniques for Toolglass (Kabbash *et al.* 1994), and found that bimanual techniques improve performance, by reducing the number of operations, and cognitive load. Hinckley et al. designed a bimanual asymmetric interface for neurosurgical visualization (Hinckley *et al.* 1994), based on Guiard's framework. Their technique used real world props for controlling the volume data with a doll's head in the non-dominant hand, and a clear plastic plate in the dominant hand for the cutting plane. Cutler et al. developed bimanual direct manipulation techniques for the Responsive Workbench (Cutler *et al.* 1997), based on Guiard's framework, and suggested that the most interesting bimanual interactions are coordinated and asymmetric.

In favor of symmetric bimanual interactions, Kelso argued for of the coordinative coupling between the two human hands (Kelso *et al.* 1979), claiming that the brain produces simultaneity of action in bimanual tasks by organizing functional groupings of muscles constrained to act as a single unit. One of the earlier studies comparing bimanual and unimanual inputs by Buxton et al. (Buxton and Myers 1986) showed that bimanual method for navigation/selection outperformed single handed method for several measures, and the speed of performance in a positioning task was strongly correlated to parallelism or symmetry between the two hands. In a study of graphical object manipulations, Hauptmann (Hauptmann 1989) found that users expressed scaling and rotations through bimanual symmetric gestures (like using a steering), suggesting a natural preference for bimanual actions. The Polyshop designed by Mapes et al. (Mapes and Moshell 1995) proposed symmetric bimanual techniques for scaling, rotating, and stretching graphical objects.

Investigating factors affecting symmetric bimanual interaction, Balakrishnan and Hinckley reported that increasing task difficulty, divided attention, and lack of visual integration, all cause the brain to revert to sequential style of interaction, rather than parallel (Balakrishnan and Hinckley 2000). An interesting finding in their study was that parallelism is not a requirement for

performance to be symmetric, implying that level of symmetry does not easily degrade due to increased task difficulty or divided attention.

Results from an experiment by Hinckley et al. (Hinckley *et al.* 1997) suggested that the two human hands together create a perceptual frame of reference, which is independent of visual feedback. Further investigating interactions between visual feedback and kinesthetic reference frames in bimanual interactions, Balakrishnan and Hinckley reported that visual feedback is clearly dominant over bimanual kinesthetic reference frames when both are present, but they can independently guide bimanual actions (Balakrishnan and Hinckley 1999).

In bimanual interactions, due to integration of the sub-tasks with the two hands, the cost of task switching is lowered, suggesting performance improvement at the cognitive level. Leganchuk et al. compared two bimanual techniques with a unimanual technique, and observed that bimanual interactions bring significant manual, and cognitive benefits to the users (Leganchuk *et al.* 1998). Owen et al. reports that for a curve matching task, as the cognitive demands increase, bimanual technique provides greater performance benefits over unimanual technique (Owen *et al.* 2005).

In spite of all these evidences, very few researchers focused on designing or standardizing bimanual techniques for volume data analysis, with the exception of Hinckley et al. (Hinckley *et al.* 1994). In her doctoral dissertation work, Ulinski explored and created guidelines for designing bimanual selection techniques for volume data (Ulinski 2008). She found that symmetric and synchronous techniques gave best performance for selection tasks in volume data.

Recently, there has been some effort in leveraging symmetric but asynchronous bimanual techniques for medical data exploration (Mlyniec *et al.* 2011). Jackson et al. created a prop-based tangible interface for volume data analysis (Jackson *et al.* 2013). Their interface, although lightweight, relies on the affordances of the prop and does not leverage the freedoms of movement of the human hand. Researchers are also exploring a variety of free-hand gestures for point cloud selection (Bacim *et al.* 2014).

Our design of the Volume Cracker technique (Laha 2013) relies on using the two hands of the user for interaction with a volume. It lowers the cognitive load on the users through leveraging the muscle memory of the user by using postures from the real world (Leganchuk *et al.* 1998; Balakrishnan and Hinckley 1999). We discuss the design and evaluation of the volume cracker in Chapter 6 and a bare-hand prototype of the volume cracker in Chapter 7.

The evaluation of the volume cracker also leveraged tasks from a volume data analysis task taxonomy that we are designing. So in the last section below of the related work chapter, we shall take a look at the prior work done in developing visualization task taxonomies.

## 2.7 Generic and High Level Visualization Task Taxonomies

There have been several prior attempts at creating high-level generic task taxonomies for information visualization. In this section we touch upon some of the important ones, and discuss on what findings we are basing the design of our task taxonomy.

One of the first task taxonomy for information visualization was designed by Shneiderman (Shneiderman 1996). He proposed a data type taxonomy with seven data types (one-, two-, three-dimensional data, temporal and multi-dimensional data, and tree and network data), and seven tasks (overview, zoom, filter, details-on-demand, relate, history, and extracts).

Zhou and Feiner proposed a visual task taxonomy (Zhou and Feiner 1998) for interfacing high level presentation intents with low-level visual techniques. In their approach they characterized visual tasks by their visual accomplishments, and implications.

Morse et al. proposed a domain-independent taxonomy of visual tasks (Morse *et al.* 2000), for testing of visualization in isolation from the rest of the system. They proposed defining tests based on the high-level visual task taxonomy to generalize the evaluation results to newer system rather than basing the results on tasks that could be achieved only with older, less capable systems. They discussed methods to exhaustively test capabilities of an information retrieval system based on their task taxonomy.

Amar and Stasko presented a task based framework or task taxonomy (Amar and Stasko 2004) for bridging analytic gaps in information visualization. Their approach was to seek generalization for design and evaluation of information visualization systems, across users, data sets, and domains.

Tory and Moller presented a high-level visualization task taxonomy (Tory and Moller 2004), for categorization of visualization algorithms, rather than data. Their categorization was based on design model, which considers the user's conceptual model. They claim that such a categorization is more flexible than other approaches, and helps to understand how the domains of information and scientific visualizations might relate and overlap.

Valiati et al. proposed a visualization task taxonomy (Valiati *et al.* 2006), based on previous designs, and observation of tasks performed by users with two visualization techniques in a multidimensional dataset. They suggested that such a taxonomy would be useful for guiding the selection of visual, and interaction metaphors, as well as usability testing of visualization techniques.

In spite of all this work, the volume visualization community is still lacking a comprehensive taxonomy for visual analysis of volume visualizations. This is identified as the first and foremost scientific visualization research challenge by (Keefe and Isenberg 2013). We developed a task taxonomy of visual analysis tasks with volume datasets by closely working with many domain scientists, and evaluated the design through an online survey; we present our results in Chapter 8.

## 2.8 Conclusion

In this section we looked at the existing literature for empirical studies on the effects of immersive VR on a variety of task performance, a few controlled experiments and their results, empirical and controlled studies for evaluating the effects of VR on volume data analysis, the MR simulation approach and studies towards its validation, visualization and interaction techniques for analyzing volume datasets, a brief overview of bimanual interaction techniques, and prior attempts at creating visualization task taxonomies. In the next chapter, we discuss the first project we completed for answering research question - 1 (see section 1.4.1).

## Chapter 3. VR Study-I: Effects of Display Fidelity on Visual Analysis of Volume Data

To address research question - 1 (see section 1.1), we ran three controlled studies. In this chapter we summarize the motivation and the findings from the first experiment.

Prabhat and his colleagues, in 2008, showed benefits of more immersive systems for analyzing confocal datasets (Prabhat *et al.* 2008). This study and a few others ran previously (Zhang *et al.* 2001; Zhang *et al.* 2003) reported significant benefits of more immersive systems for volume data analysis. But they couldn't tie the significant results to the various components of the systems they studied. We want to be able to port these and other empirical findings to systems other than the ones used in the studies, and also to domains other than the ones from where the data was used in these experiments. This inspired us to design and run our first controlled experiment, described below (Laha *et al.* 2012).

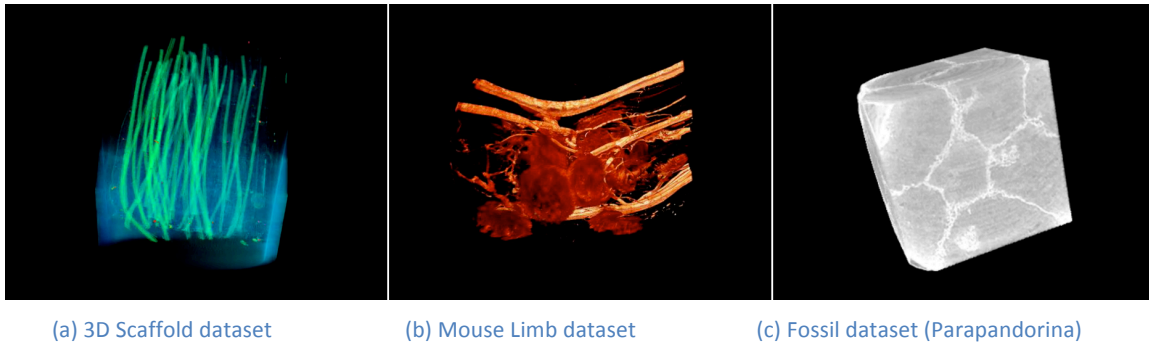


Figure 4: The three volume visualized microscopic computed tomography (micro-CT) datasets used in VR-Study-I.

### 3.1 Experiment

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

#### 3.1.1 Goals and Hypotheses

Broadly, our goal is to understand the influence of the level of display fidelity 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 display fidelity for understanding visualizations of volume data?*

Besides the general benefits of higher levels of display fidelity, we want to find the effects of individual components of display fidelity (Bowman and McMahan 2007) on visual analysis of volume datasets, which gives us our next question:

2. *What are the effects of components of display fidelity on understanding visualizations of volume data?* In this study, we chose to explore the effects of three specific components of display fidelity 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 display fidelity, as prior studies have shown (Schuchardt and Bowman 2007):

3. *How are different visual analysis tasks affected by components of display fidelity?* 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 display fidelity when performing visual analysis tasks with volume datasets:

4. *When performing visual analysis of volume data, what level of display fidelity 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 display fidelity.

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

1. *There are significant positive effects of higher levels of display fidelity for analyzing volume data.* Prior research has shown these benefits partially already (Zhang *et al.* 2001; Zhang *et al.* 2003; Prabhat *et al.* 2008).
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 display fidelity will have significant benefits for complex spatial tasks, but not for others.* The extra spatial and proprioceptive cues offered by higher levels of display fidelity could benefit spatial analysis tasks involving detailed spatial judgments. However, lower levels of display fidelity are likely to be sufficient for simpler tasks.
4. *Users will prefer higher levels of display fidelity.* Higher display fidelity 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.

### 3.1.2 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 (Figure 4-a) used in bone regeneration studies (Sensharma *et al.* 2009). 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 (Sensharma Oct 2011), imaged at the major knee joint of the mouse (Figure 4-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 (Figure 4-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 (Schiffbauer *et al.* 2012). The visualization that participants viewed was of an incomplete fractured specimen.

### 3.1.3 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 (Bowman and Raja 2004).

#### 3.1.3.1 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.

#### 3.1.3.2 Software framework

Our software framework for distributed volume rendering consisted of DIVERSE (Kelso *et al.* 2003) at the bottom layer for clustering and distributed rendering, 3D Visualizer (Billen *et al.* 2008) for volume rendering of the micro-CT datasets, and VRUI (Kreylos 2008) 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.

#### 3.1.3.3 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.



Figure 5: 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. Figure 5 shows a head-tracked user examining the training dataset in the Viscube.

#### 3.1.4 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 display fidelity 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 3.1.7). 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 3.1.5), 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 3.3.1 and 3.3.2).

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.

### 3.1.5 Exploratory Study

We conducted an exploratory study for the purpose of evaluating whether the level of display fidelity 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 (Sensharma *et al.* 2009; Schiffbauer *et al.* 2012; Sensharma Oct 2011) and their tasks, always presented in the same order, were evaluated with two levels of display fidelity, 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 display fidelity 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 display fidelity condition for these tasks.

### 3.1.6 Design

In this controlled experiment, we chose to study the effects of three components of display fidelity, 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 3.2). 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.

### 3.1.7 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 (Ekstrom *et al.* 1979), 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.

### 3.1.8 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 (Ekstrom *et al.* 1979). 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 (Figure 4-a) such that the novice participants were trained in the different strategies used by experts for analyzing the two main micro-CT datasets (Figure 4-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 (Figure 4-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 display fidelity. 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 (Figure 4-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.

### 3.2 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  $\sum M$  and  $\sum F$  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 3.3.1.

### 3.2.1 Grades (accuracy in task performance)

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

Table 3. Significant Main Effects on Grades

Task: source	$\chi^2$	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
$\Sigma$ 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
$\Sigma$ 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
$\Sigma$ M: HT	4.792	0.029	HT on better
M4: ST	22.746	< 0.0001	ST on better
$\Sigma$ 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 display fidelity on the grades received by the participants (see Table 4). These interactions are plotted in Figures 6-9).

Table 4. Significant Interaction Effects on Grades

Task: source	$\chi^2$	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.

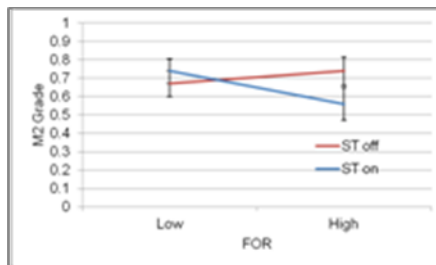


Figure 6: Interaction between FOR and ST for M2 grade.

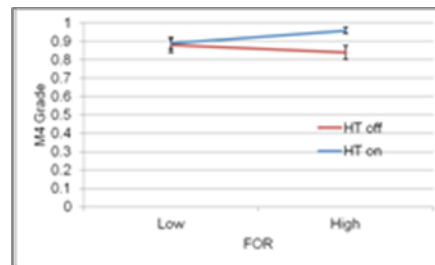


Figure 7: Interaction between FOR and HT for M4 grade.

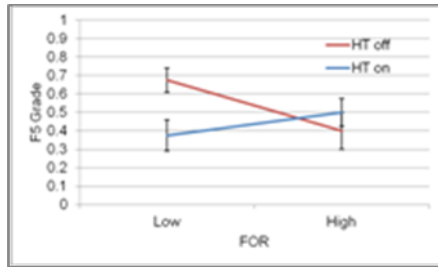


Figure 8: Interaction between FOR and HT for F5 grade.

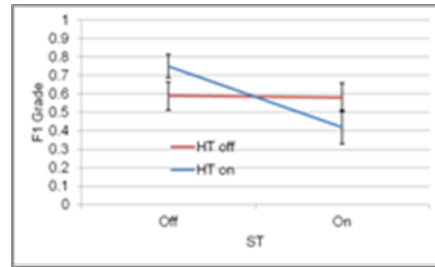


Figure 9: Interaction between ST and HT for F1 grades.

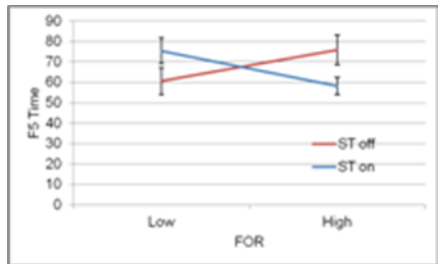


Figure 10: Interaction between FOR and ST for F5 time (in sec).

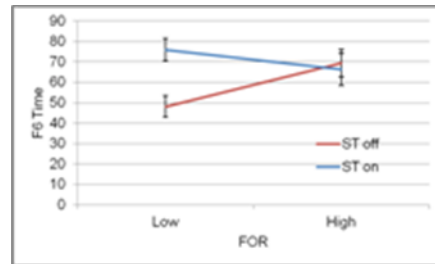


Figure 11: Interaction between FOR and ST for F6 time (in sec).

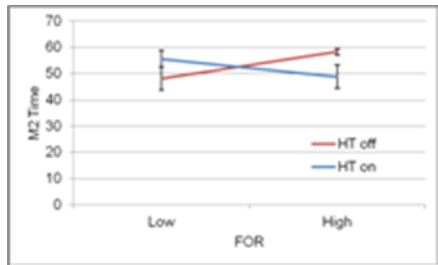


Figure 12: Interaction between FOR and HT for M2 time (in sec).

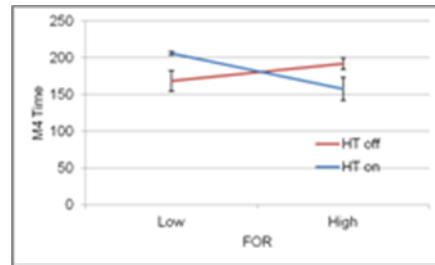


Figure 13: Interaction between FOR and HT for M4 time (in sec).

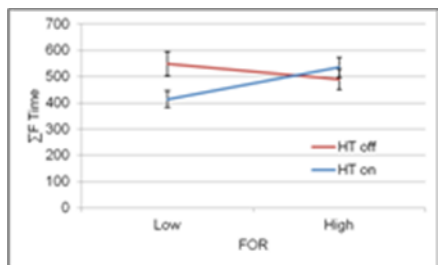


Figure 14: Interaction between FOR and HT for  $\Sigma F$  time (in sec).

### 3.2.2 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 display fidelity on the time taken by the participants in F5 (Figure 10), F6 (Figure 11), M2 (Figure 12), M4 (Figure 13),  $\Sigma M$ , and  $\Sigma F$  (Figure 14).

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
$\Sigma M$ : 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



$\Sigma F$ : 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  $\Sigma M$ , the FOR high HT on and FOR low HT off were significantly better than the FOR low HT on condition. Interestingly in  $\Sigma F$ , the FOR low HT on condition was significantly better than the FOR high HT on and FOR low HT off conditions.

### 3.2.3 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	$\chi^2$ 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 display fidelity on the difficulty levels perceived by the participants (see Table 7).

Table 7. Significant Interaction Effects on Difficulty

Task: source	$\chi^2$	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 & HT	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
			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 ( $p = 0.031$ ).

### 3.2.4 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	$\chi^2$	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 display fidelity on the confidence levels (see Table 9).

Table 9. Significant Interaction Effects on Confidence

Task: source	$\chi^2$	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 ( $p = 0.004$ ).

### 3.2.5 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 display fidelity 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 display fidelity 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.

### 3.2.6 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.

### 3.2.7 Summary of important results

Looking across the different metrics from above, we summarize the main effects of the three components of display fidelity (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.

## 3.3 Discussion

We found that most of the effects of high display fidelity 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 display fidelity, 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 display fidelity 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.

### 3.3.1 Effects of display fidelity on task categories

Similar to the Schuchardt et al. study (Schuchardt and Bowman 2007), 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 display fidelity 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 display fidelity 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 display fidelity 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 (Schuchardt and Bowman 2007), spatially complex search tasks seem to benefit strongly from higher levels of display fidelity.

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 display fidelity 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.

### 3.3.2 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 Table 2), 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 (Ritman 2004; Stock 2008). 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 (Lee *et al.* 2010b).

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.

### 3.3.3 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.

### 3.4 Conclusion

Revisiting our research question - 1 (see section 1.4.1), we found that most of the benefits of high display fidelity for analysing volume data are positive. Of the three components of display fidelity that we evaluated in our controlled study, FOR had the most positive effects on the widest range of tasks.

However, levels of display fidelity have differential effects on different tasks. If we re-categorize the tasks (see Table 27) based on our generic task categorization (Chapter 8), we find that FOR helped search (M1, M3, F4) and shape description (F1), while stereo was disruptive in those tasks (F1, F2). We believe that the extra space provided by higher FOR helped to reduce the clutter in the dense 3D texture rendering and aided in search tasks. But the better depth perception from stereo didn't help at all in describing the shape of the compact fossil (F1) or understanding the boundaries of the cells (F2). The lack of surfaces, and the closely matching density values of the boundaries might have even confused the participants causing the lower accuracy in those tasks. However, the higher depth perception through stereo was helpful in spatial judgment (M4), where the gaps between the blood vessels and their orientation in 3D were critical for accuracy of the answers.

Better motion parallax through head tracking helped in search and spatial judgment involving isolation of structures (M1) or understanding the gaps and orientation between structures (M4). Further, head tracking with higher FOR helped again in spatial judgment (M4) and pattern recognition (F5). The extra space from high FOR probably helped in leveraging the benefits of nuanced motion parallax, such as better depth perception in judging the small gaps and matching the patterns in those tasks.

Users showed mixed preferences for the different levels of display fidelity. 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.

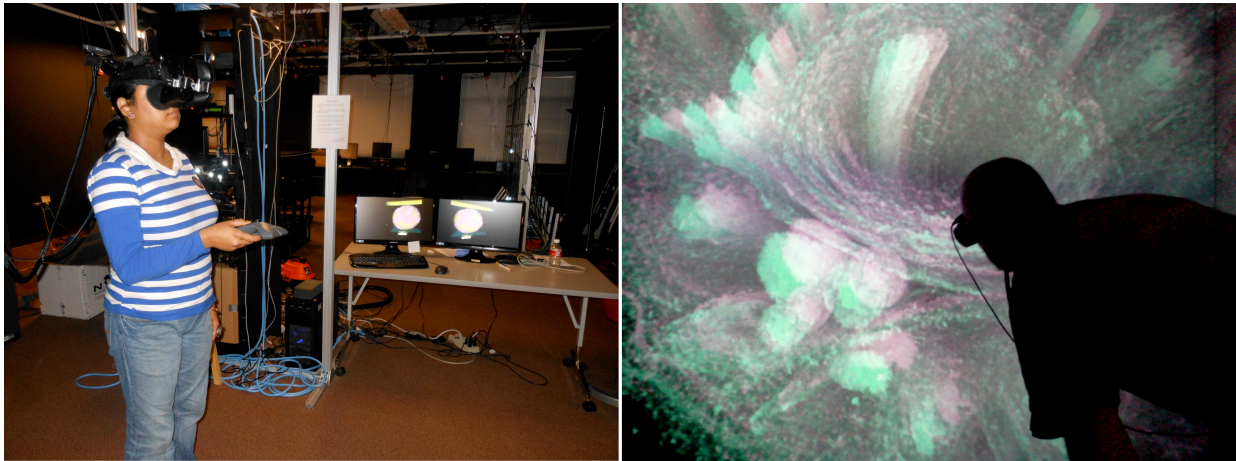
Although we found evidence of high FOR improving task performance for analysing volume data, we do not yet know what level of FOR is needed to provide these benefits. An interesting follow up study could vary FOR with three or more levels to examine these effects more closely (see Chapter 4).



## Chapter 4. VR Study-II: Validation of the MR Simulation Approach for Evaluating the Effects of Display Fidelity on Visual Analysis of Volume Data

Although VR-Study-I (see Chapter 3) produced significant results tied to one or more components of a VR system, the findings were only tied to the components from a single hardware platform (the 4-sided CAVE), which we used in that experiment. To have stronger arguments on whether the findings can be generalized across hardware platforms based on the system components, we wanted to repeat the experiment in a second display. In doing so, we attempted to validate the concept of mixed-reality (MR) simulation. The concept of MR simulation (Lee *et al.* 2010a; Lee *et al.* 2012) suggests that we can recreate any system on the MR continuum (Milgram and Kishino 1994) if we are able to exactly recreate the fidelity levels of the various components that go on to make that system. Thus, the validation of MR simulation will give us a strong argument in favor of claiming that the findings from VR empirical studies are generalizable across any VR (or more generally, MR) platforms.

In VR-Study-I, we were also limited by the CAVE hardware in that we couldn't recreate 360° FOR. As we had observed several significant effects of FOR in VR-Study-I (both single-factor and combined with HT), we wanted to study the effects of FOR more closely. We thus recreated the visualization of the datasets from our VR-Study-I in the Nvis nVisor SX11 head-mounted display (HMD), which allowed a full 360° FOR (see Figure 15). With four levels of FOR (90°, 180°, 270°, 360°) and two levels of HT (on/off), we had eight independent conditions for VR-Study-II, which we describe here (Laha *et al.* 2013).



(a) A participant wearing the Nvis SX111 HMD (b) A participant inside the Viscube (4-sided CAVE)

Figure 15: Participants using two different Mixed-Reality (MR) simulator platforms in our VR studies I and II.

### 4.1 Experiment

We designed a controlled experiment to reproduce most of the conditions from our previous experiment and also to find more granular results on the interaction effects of FOR and head tracking for analyzing volume datasets.

#### 4.1.1 Goals and Hypotheses

Our primary goal in this study is to understand if our prior results on task performance with visual analysis of volume data (Laha *et al.* 2012) still hold when the experimental conditions are recreated with a different MR simulator platform. Thus, our first research question is:

1. *Are there differences in the findings for various experimental conditions when different MR simulator platforms are used to run the experiment?*



Our earlier study (Laha et al. 2012) used a four-sided CAVE as the simulator platform; we decided to use a high-end HMD with important differences from the CAVE platform in the current study.

In our previous study (Laha et al. 2012), we had two levels of FOR (*high* or 270 degrees and *low* or 90 degrees). In several cases, we found significant interactions between FOR and head tracking (HT), with FOR high/HT on and FOR low/HT off proving to be better than the other two combinations. We also found several significant individual effects of FOR. With only two levels of FOR, however, the effects of the highest possible level (360 degrees) and of moderate levels (e.g., 180 degrees) were unknown. This leads to our next research question:

2. *What are the individual effects of FOR and its interaction with HT on visual analysis tasks with volume datasets?*

In this study, we chose to have four levels of FOR (90, 180, 270, and 360), and two levels of HT (on and off).

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

1. *We do not expect to find a significant difference between the findings of an MR simulation experiment run on a CAVE and those from an experiment run on an HMD.*

In theory, since the level of display fidelity is an objective description of a VR system (Slater 2003), the effects produced by any MR simulator platform, which is characterised by particular levels of display fidelity components, should be comparable. Thus, if we simulate the experimental conditions as closely as we can, using the different MR simulator platforms, then we should get similar results. However, other differences between the platforms, such as FOV, weight, accommodation distance, and the presence or absence of seams on the display, could potentially affect the results. The primary differences between the platforms are shown in Table 10. We hypothesize that the effects will come primarily from the variables being studied, and not from these differences in the platforms.

2. *The combination of the highest level of FOR with HT on will produce the best results, followed by the combination of the lowest level of FOR with HT off.*

We hypothesize that the trends from our previous experiment (Laha et al. 2012) will continue when new levels of FOR are considered.

#### 4.1.2 Datasets

Computed Tomography (CT) performed at the microscopic (10-6) level, or micro-CT, produces 3D internal imaging of objects, and is useful in various disciplines such as biology, palaeontology, and medicine. Traditionally, researchers have used desktop displays to visualize and analyze micro-CT data in volumetric format. As good visualization is essential for the analysis of such datasets, scientists have shown great interest in evaluating VR platforms for analyzing their datasets (Laha et al. 2012).

We worked with domain scientists to identify three datasets actively used in their work. The first one is a 3D Scaffold dataset (Figure 16-a) used in bone regeneration studies (Sensharma et al. 2009). 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.

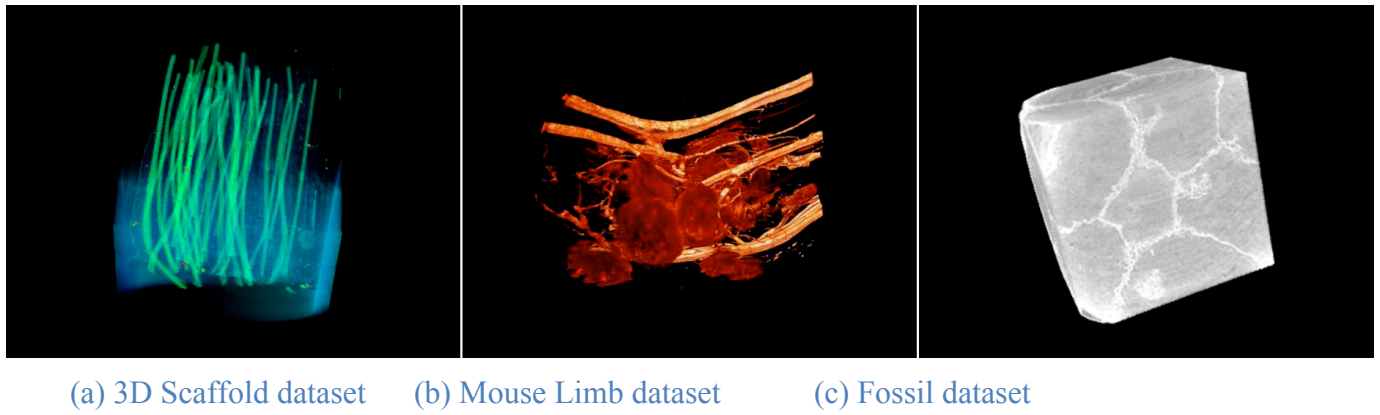


Figure 16. The micro-CT datasets used in our studies.

The second dataset was a mouse limb (Sensharma et al. 2011), imaged at the major knee joint of the mouse (Figure 16-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 (Figure 16-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 (Schiffbauer et al. 2012). The visualization that participants viewed was of an incomplete fractured specimen.

Table 10: Primary differences between the two MR simulator platforms

Factor	CAVE	HMD
Horizontal FOV	90° (with stereo glasses)	102°
Resolution	1920×1920 per wall	1280x1024 per eye
Weight worn on head	85 grams	1.3 kilograms
Stereoscopic display technology	Infitec stereo	Separate displays for each eye
Accommodation distance	Approx. 1.5 m	Infinity
Seams between displays	Visible seams	None
Occlusion of body and surrounding environment	No occlusion	Full occlusion

### 4.1.3 Apparatus

#### 4.1.3.1 Hardware and software

We used the NVis SX111 head mounted display (HMD) as our MR simulator platform (Figure 17). It offers a FOV of 102° by 64°, with 1280x1024 pixels per eye. Head movements were tracked by a wired head tracker of an Intersense IS-900 tracking system, which also provided a wireless wand device with a joystick and five buttons. A participant using the system is shown in Figure 15-a. A participant using the MR simulator from our previous experiment (a four-sided CAVE-like system) can be seen in Figure 15-b. Table 10 shows differences between the CAVE and the HMD we used.



Figure 17: NVis SX111 Head Mounted Display

We used DIVERSE (Kelso *et al.* 2003) to get data from the head tracker and the wand from the IS900 system. The open source 3D Visualizer (Billen *et al.* 2008) gave us a platform for interactive volume rendering, with stereo capabilities for the two screens of the HMD. We used a customized version of VRUI (Kreylos 2008) for the specific 3D interaction needs of our experiment, as described in the following section.

#### 4.1.3.2 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. These interactions were identical to those used in the prior experiment (Laha *et al.* 2012).

To correctly simulate the head tracking conditions from our previous study (where we used a four-sided CAVE-like environment (Laha *et al.* 2012)), we disabled positional head tracking for conditions where head-tracking was off. Positional head tracking affects the rendering of the volume based on the position of the head tracker. The rotational head tracking was enabled, even in the non-head tracking condition, because in a CAVE-like setting without head tracking, head rotations still allow the user to see views of the dataset in different directions. In head-tracked conditions, both positional and rotational head-tracking were enabled.

#### 4.1.4 Tasks

We used the same set of tasks from our previous study (Laha *et al.* 2012), but with a few key modifications (in appendix). Tasks were either quantitative (counting features) or qualitative (describing characteristics). In the quantitative tasks, participants gave their answer verbally, and the experimenter recorded the response on paper.

For qualitative tasks (deviating from the previous experiment in which the participants answered verbally), the participants were shown a choice of five response options, from which they marked the most appropriate one. We chose to do this to have more objectivity to the analyses of the results, as previously

we found that open-ended responses to descriptive tasks often produced a wide array of responses, some of which were difficult to interpret and grade objectively.

In our previous project, we had worked with micro-CT researchers to identify tasks that are of actual importance to their research, to make sure that any benefit of display fidelity that we found could be used by them and others in their community. Since we planned to run the studies with novice participants (to avoid confounds based on prior knowledge level) we ensured that the tasks were of real technical importance to experts but at the same time not so cryptic so as to confuse novices. We assumed a basic knowledge of what blood vessels, cells, bones, and other simple biological structures look like.

To train the participants, we had three quantitative and three qualitative tasks with a training dataset (Figure 16-a). The tasks for the two main datasets with the suggested strategies and new multiple response options for the qualitative tasks are shown in the appendix.

The tasks in each dataset were different. But as before, we categorized them in abstract task categories (see Table 2) and counterbalanced the order of the datasets so that each combination of dataset and experimental condition was studied.

#### 4.1.5 Design

This controlled experiment was primarily designed as a follow up to our previous experiment (Laha et al. 2012). In this experiment we wanted to closely study the effects of two independent variables, FOR, and HT, keeping all other factors constant. FOR had four levels: 360, 270, 180, and 90 degrees. At level ‘x’, the user could view  $x^\circ$  of the virtual environment by rotating her head about the vertical axis. HT had two levels: on and off. At the ‘on’ level, both rotational and positional HT was enabled. At the ‘off’ level, only rotational HT was enabled. The different conditions with the levels of FOR and HT are shown in Table 11.

For producing the four levels of FOR, we created two virtual black walls. The black walls extended infinitely in the vertical direction. Horizontally, they merged four inches behind the head position, and formed a horizontal angle corresponding to the FOR. The walls moved (changing position, but not orientation) with the user’s head. While this is not exactly like the different levels of FOR in a CAVE-like display, it ensured that the user could not move his head through the walls. Conditions with 360-degree FOR had no black walls.

Table 11. Conditions Experienced by the Eight Groups in the Experiment

Group#	First Condition (Mouse-Limb)		Second Condition (Fossil)	
	FOR	HT	FOR	HT
1	360	On	360	Off
2	360	Off	360	On
3	270	On	270	Off
4	270	Off	270	On
5	180	On	180	Off
6	180	Off	180	On
7	90	On	90	Off

With four levels of one variable and two levels of the other, we had eight possible conditions. We chose to vary FOR between subjects and HT within subjects, as in the previous experiment. This allowed us to study whether individuals used different strategies to explore the datasets with and without HT. Although all participants experienced both levels of HT, we consider those who experienced HT on first to be a separate group from those who experienced HT off first, since the two datasets were not comparable in terms of complexity. All participants first performed tasks with the mouse limb dataset followed by the fossil dataset (Table 11).

As before, the dependent variables were the amount of time taken for each task and the responses of the participants to each task, recorded and graded offline by the experimenter 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 seven-point scales.

#### 4.1.6 Participants

We recruited 65 voluntary unpaid participants for our study, all of whom reported no prior experience in analyzing volume visualized micro-CT datasets. Most of the participants were recruited through a university wide recruitment system, and got awarded two credits in a psychology course for their participation. Four of them were pilot participants. We dismissed 13 participants based on below-threshold scores on a spatial ability test (Ekstrom *et al.* 1979). This gave us a total of 48 participants, distributed uniformly in eight study groups (six participants per group), with comparable spatial ability scores in each group. The overall average spatial ability of the participants in this study (8.35, max score 20) was lower than that (10.95, max score 20) in the previous study (Laha *et al.* 2012). In this study, 26 males and 22 females participated, all undergraduate or graduate students. They were 18 years to 41 years old, with an average age of 21 years.

#### 4.1.7 Procedure

Our study was approved by the Institutional Review Board of our university. Before beginning the study the participants signed a standard informed consent, informing them of their right to withdraw at any point during the study. Next, participants filled out a background questionnaire capturing information on their demographics, experience with VR, and experience analyzing CT and micro-CT datasets. Following that, they took the spatial ability test (Ekstrom *et al.* 1979) discussed above. The participants were then given a brief background of the purpose of the study, introduced to the hardware, and trained with the various 3D interactions they were about to use.

The tasks in the training dataset (Figure 16-a) trained the participants in the different expert strategies that domain scientists use. The training introduced the participants to the various interactions with the HMD and the wand, how to analyze a volume visualized micro-CT dataset using that system, and how to complete quantitative and qualitative tasks. The participants trained on the same condition in which they would experience the first (mouse limb) dataset. The participants also completed three rotation tasks, about the three orthogonal axes, with the joystick of the wand, to make sure they could comfortably use the rotations when needed, without outside assistance. The training took around 15-20 minutes.

After the training, participants were asked to take a short break. Then they started analyzing the mouse limb dataset (Figure 16-b). The participants were asked to be as accurate as possible in their responses. They were informed that there was a maximum amount of time for each task. For the quantitative tasks, they were asked to let the experimenter know when they were ready to answer. The experimenter recorded the time using a stopwatch. For the qualitative tasks, they were asked to analyze the dataset for the entire available time, after which they were shown five options. After every task completion, the experimenter

also recorded the perceived level of difficulty and confidence level in two separate seven-point scales. The details of the tasks are in the appendix.

The participants then rested for a short while, and again underwent training in the condition they would use to analyze the fossil dataset (Figure 16-c) in. They performed seven tasks with the fossil, in the same manner as the mouse limb, and the experimenter recorded their responses in the response sheets.

As in the previous study, if the participants digressed too much from the expected strategy for a particular task, we guided them towards the correct expert strategies. The appendix lists the main strategies for each task identified by the domain experts. We thus tried to emulate expert strategies as closely as possible.

Finally, the participants completed a post-questionnaire capturing their opinions for both the head-tracked and non-head-tracked conditions on seven-point 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 viewpoints.

The datasets in each condition were rendered at the same initial position and orientation in front of the participants. Each question was read out loud to the participants, using consistent wording.

## 4.2 Results

In this section, we first present the significant results from our recent controlled experiment. We then present the comparison of the results of our current study with those from our previous study with the CAVE-like system. We first compared the significant results from the two studies. We then compared the grade metric in all the conditions from both experiments in which all the components of display fidelity were at the same level.

Table 12. Relative weights of tasks and abstract task categories

Mouse 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 search feature	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



In the current study, task time was analyzed as a numeric continuous variable, while the other measures (grade, difficulty, and confidence) were considered to be numeric ordinal variables. To understand the significant main effects and the two factor interaction effects of our independent variables (FOR and HT), we ran a two-way analysis of variance for the time metric, and an Ordinal Logistic Regression based on a Chi-square statistic for all other metrics.

For the sake of brevity, we shall use the task numbers as defined in the appendix; e.g., ‘M1’ will denote the first task with the mouse limb dataset, and ‘F4’ will denote the fourth task with the fossil dataset. We used our previous classification of the tasks in the abstract categories as shown in Table 2. Table 2 also shows the relative weights of the tasks (totalling 1.0 for each dataset) determined by domain scientists, based on the perceived relative importance of the tasks to their own research. We used these weights to calculate the weighted totals  $\sum M$  and  $\sum F$  for the mouse limb and fossil datasets respectively.  $\sum M$  and  $\sum F$  helped us to evaluate the overall effects on the tasks with a particular dataset.

#### 4.2.1 Significant results from current experiment

Here we report all the significant main effects and interaction effects of the independent variables FOR and HT in our present experiment with the HMD system. For the interaction effects, we also present graphs to compare them with those from the previous experiment.

##### 4.2.1.1 Grades (accuracy in task performance)

The significant main effects of FOR and HT on the grades received by the participants are shown in Table 13. We found significant interactions of FOR and HT on the grades received by the participants in two cases. The first case is the effect on M4 grade ( $\chi^2_{df=3}=10.371$ ,  $p=0.016$ ) and is shown in Figure 18. The left graph is from the original experiment with a CAVE-like system; the right graph is from the present experiment. Overall, the M4 grades show similar trends in both experiments. Grades were better with higher FOR when HT was on, but were better with lower FOR when HT was off. The mean and variances of the grades are also comparable. Additionally, in the HMD experiment, we learned that the grades reached the lowest level at FOR 270 with HT off, and didn’t change significantly from FOR 270 to 360 with HT on.

Table 13. Significant Main Effects on Grades

Task: source	$\chi^2$	DF	p-value	Note (higher grade is better)
F1: FOR	7.983	3	0.046	270>360>180>90
F3: FOR	11.849	3	0.008	180>360=90>270
F4: HT	8.342	1	0.004	on > off
$\sum F$ : HT	9.967	1	0.001	on > off

The second significant interaction of FOR and HT on F4 grades ( $\chi^2_{df=3}=8.672$ ,  $p=0.034$ ) is shown in Figure 19. Again, the left graph is from the previous experiment with a CAVE-like system; the right graph is from the current experiment. The F4 grades in the two graphs are comparable at FOR 270 and 90 for both HT on and off. In the HMD experiment, the best results were achieved with FOR 360 and HT on, and the worst results with FOR 180 and HT off. We found that three of the six participants in group five (with FOR 180 and HT off condition in the fossil dataset) failed the task completely. As a result the data point probably became an outlier in our study.

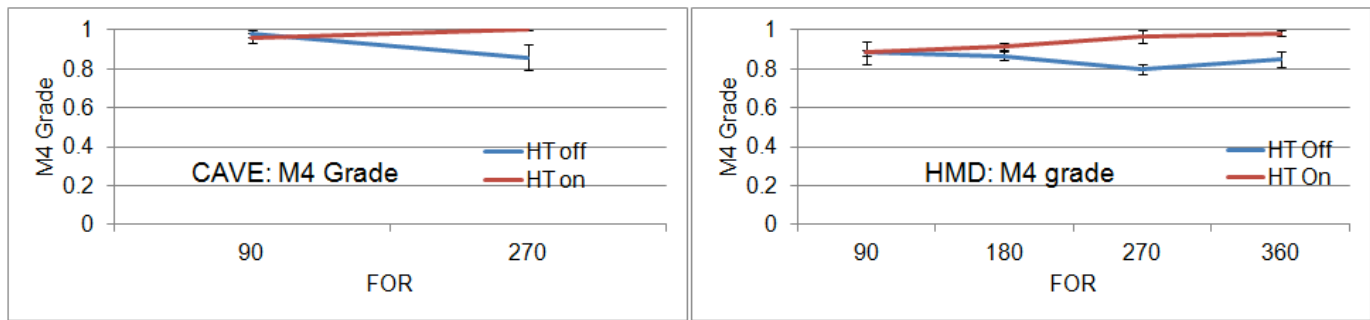


Figure 18: Interaction between FOR and HT for M4 grade

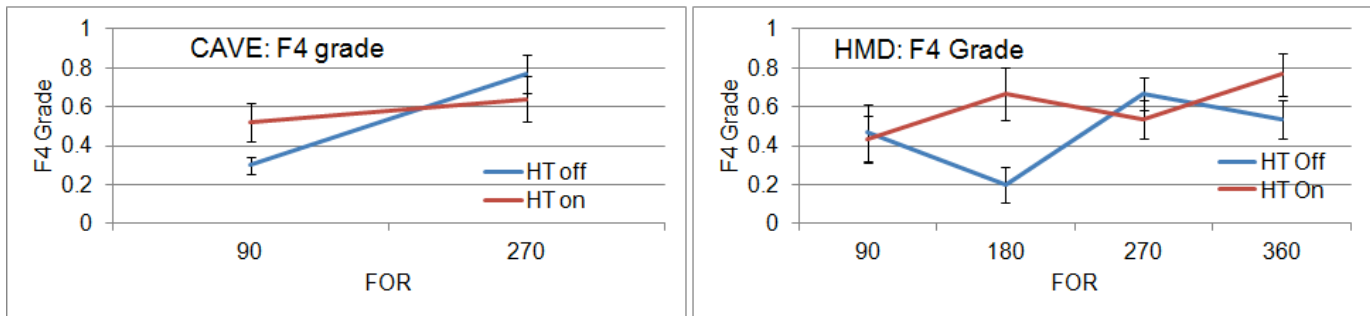


Figure 19: Interaction between FOR and HT for F4 grade

#### 4.2.1.2 Task completion time

We found a significant main effect of FOR on F4 time ( $F(3, 40) = 5.4773$ ,  $p=0.003$ , power=0.9149) in the HMD experiment. A post-hoc t-test ( $t=2.021$ ) indicated that the task performance at FOR 360, and FOR 270 was significantly faster than that at FOR 90, and FOR 180, with the fastest mean time achieved with FOR 360.

#### 4.2.1.3 Subjective metrics

There were no significant main or interaction effects of FOR or HT for the perceived difficulty metric in the current experiment.

For perceived confidence levels of F6, we found a significant main effect of FOR ( $\chi^2_{df=3}=8.394$ ;  $p=0.0385$ ), and of HT ( $\chi^2_{df=1}=4.58$ ;  $p=0.0323$ ). Confidence levels were highest with FOR 180, decreased with FOR 90 and FOR 270, and were lowest with FOR 360. Confidence levels were higher with HT on.

There were significant interaction effects between FOR and HT for three tasks: F4, F5, and F6. Table 14 shows the  $\chi^2$  and p-values of the interaction effects and the means of the different conditions.

From the table, we see that the participants consistently had higher confidence levels for three conditions: FOR 90 with HT off, FOR 360 with HT on, and FOR 180 with HT on, and consistently had the lowest confidence for the condition FOR 360 with HT off.

Table 14. Significant interaction effects of FOR and HT on confidence

Task: source	$\chi^2$	DF	p-value	Mean values in descending order (higher is better)	
F4: FOR & HT	10.552	3	0.014	FOR 90 HT off	5.3
				FOR 360 HT on	5.2
				FOR 180 HT on	4.7



F5: FOR & HT	8.858	3	0.031	FOR 180 HT off	4.3
				FOR 270 HT off	4.2
				FOR 90 HT on	4.2
				FOR 270 HT on	3.7
				FOR 360 HT off	3.2
				FOR 90 HT off	6
				FOR 180 HT on	5
				FOR 360 HT on	4.8
				FOR 180 HT off	4.5
				FOR 90 HT on	4.5
F6: FOR & HT	9.689	3	0.021	FOR 270 HT on	4.3
				FOR 270 HT off	4
				FOR 360 HT off	3.5
				FOR 180 HT on	6.2
				FOR 90 HT off	5.8
				FOR 360 HT on	5.3
				FOR 180 HT off	5.2
				FOR 270 HT off	5.2
				FOR 90 HT on	5
				FOR 270 HT on	4.3
				FOR 360 HT off	3.5

#### 4.2.1.4 Post-questionnaire results

In our post-questionnaire, we captured subjective ratings of user's perception, as in our previous experiment, on a seven-point scale. The users felt that head tracking significantly improved ( $\chi^2_{df=1}=11.784$ ;  $p=0.0006$ ) the ease of getting the view they wanted, and exploring the dataset in general. In the previous experiment with CAVE as the MR simulator platform (Laha *et al.* 2012), we also found the same result ( $\chi^2_{df=1} = 3.854$ ,  $p = 0.0496$ ).

#### 4.2.1.5 Effects of spatial ability and gender of the participants

We ran pairwise correlation analyses (nonparametric Spearman's  $\rho$ ) between the spatial abilities, and gender of participants and the different metrics in our study. We found a few significant correlations with spatial ability: M2 grade ( $\rho=0.42$ ;  $p=0.003$ ), M3 time ( $\rho=0.3146$ ;  $p=0.0294$ ), F1 grade ( $\rho=0.3693$ ;  $p=0.0098$ ), and F4 time ( $\rho=0.2882$ ;  $p=0.047$ ). Accuracy was higher but performance was slower with higher spatial ability for these tasks.

Significant correlations of gender were with F2 time ( $\rho=-0.3131$ ;  $p=0.03$ ), and M4 difficulty ( $\rho=0.4573$ ;  $p=0.001$ ). Females were faster in F2, but found M4 more difficult than males.

#### 4.2.2 Comparison of MR simulator platforms

We designed our current experiment so that the display fidelity conditions had significant overlap with those from a previous study (Laha *et al.* 2012), which was run on a CAVE-like system. The 90- and 270-degree FOR conditions in our current study matched with the low and high FOR conditions, respectively, in the previous study. Both studies had the HT on and off conditions. We also had conditions with stereo on in the previous study, and all the conditions in our current study had stereo on. Thus, we had two similar levels each of two independent variables in both experiment, giving us four independent conditions to compare.

In the current study, for the qualitative tasks, we introduced a multiple choice response system, which was different from the previous study. Further, we asked the participants to analyze the dataset for a fixed amount of time for every qualitative task. Thus, we could not directly compare the results of the two studies for the qualitative tasks. The quantitative tasks remained the same from the previous study. There were three quantitative tasks in each of the two datasets in our study, giving us six independent tasks for comparison.

With four comparable display fidelity levels, and six independent tasks to compare, we had 24 sets of comparable results from both experiments that we could use to provide evidence for the validity of the MR simulation approach.

##### 4.2.2.1 Comparison of significant effects

Table 15 compares the significant effects found in the experiments using the two different MR simulation platforms, for the quantitative tasks only, organized by metric (columns) and variable (rows). The cell with significant results found in both experiments is shaded. Note that the results in the table are slightly different from the results reported in our prior paper (Laha *et al.* 2012), because we ran a new two-factor analysis of the earlier data using only the conditions with stereo on, and because FOR is treated here as an ordinal, rather than nominal variable. In addition, these effects are different than those reported above, because this analysis for the current experiment used only the data from the FOR 90 and FOR 270 conditions. Only one of the seven significant effects from the CAVE experiment was reproduced in the HMD experiment. Three new significant effects were seen in the HMD experiment that did not occur in the CAVE experiment.

Out of 65 non-significant effects in the CAVE experiment, we found 62 in the HMD experiment. Overall, then, out of the 72 significant and non-significant results (6 tasks x 3 effect types x 4 metrics), we found 63 in the current study. The similarity between the results suggests that the first hypothesis has some validity. We discuss possible causes for the differences in significant effects in section 5.1.

Table 15. Significant effects from the two experiments.

Grade		Time		Difficulty		Confidence	
CAVE	HMD	CAVE	HMD	CAVE	HMD	CAVE	HMD

FOR	F4	F4		F2
HT	F4	M4	M4	F4
FOR*HT	M4	M4		F4
	F4			

#### 4.2.2.1.1 Contingency Analysis results

Out of the four metrics in our study, grade was the most important, as it represented accuracy in task performance. We ran contingency analyses to find the relation between the two sets of grades in each comparable condition from the two studies, as the grades were considered ordinal data (Agresti 2010).

Table 16. P-values for likelihood ratio and Pearson Chi-square

Tasks	FOR 270		FOR 90	
	HT ON	HT OFF	HT ON	HT OFF
M1	0.2956	0.411	0.1308	0.2028
	0.402	0.4974	0.2307	0.2909
M3	0.1446	0.0044*	0.0995	0.2264
	0.2592	0.0266*	0.2399	0.387
M4	0.2551	0.0356*	0.4278	0.0076*
	0.3384	0.0981	0.5169	0.0219*
F2	0.2542	0.2264	0.3883	0.448
	0.3519	0.387	0.52	0.569
F4	0.5543	0.2328	0.3268	0.1847
	0.6879	0.3820	0.5102	0.3447
F7	0.6179	0.1308	0.2475	0.0233*
	0.621	0.2307	0.3611	0.0601

In Table 16 above, the upper value in each cell is the likelihood ratio, and the lower value is the p-value for the Pearson Chi-Square statistic. P-values less than 0.05 (marked with an asterisk) lead us to reject the null hypothesis  $H_0$ : *The ordinal values in the two series come from the same distribution*. In other words, significant p-values indicate that we did find a difference in grades between the two experiments. Out of the 24 different cases, no significant difference was found in 20 cases, while four cases, all from tasks with HT off, had a significant difference.

#### 4.2.2.2 Mann-Whitney test results

We also ran a Mann-Whitney test to compare the medians of the grades from the two experiments. This is a non-parametric test parallel to comparing the means of two independent populations (Agresti 2010).

Since the grade metric was categorical data, comparing medians is more appropriate rather than comparing means.

The p-values in the Table 17 are those for the Chi-square test statistic, the null hypothesis  $H_0$  being that the group means or medians are in the same location across groups. In other words, insignificant p-values ( $>0.05$ ) indicate that we did find a difference between the two MR simulators. Of the 24 different cases, no significant difference was found in 23 cases. The only significant difference was found for task M4 with FOR 90 and HT off. We also report the power of each test in Table 17. Lower power indicates higher error (type II) in claiming that the two means are the same.

The box plots in Figure 20 compare the grades of the 12 cases from the mouse limb dataset, and those in Figure 21 compare the grades of the 12 cases from the fossil dataset. A discussion on how the graphs compared is in section 4.3.

Table 17. P-values for Chi-square statistic in Mann-Whitney test

Tasks		FOR 270		FOR 90	
		HT ON	HT OFF	HT ON	HT OFF
M1	Prob>ChiSq	0.1073	0.2844	0.8157	0.077
	Power	0.3412	0.1748	0.0691	0.3661
M3	Prob>ChiSq	0.6363	0.0926	0.3993	0.7728
	Power	0.1403	0.2427	0.0906	0.0609
M4	Prob>ChiSq	0.3613	0.5069	0.3129	0.0105*
	Power	0.1281	0.1253	0.1655	0.9140
F2	Prob>ChiSq	0.1844	0.3865	0.5069	1.0
	Power	0.2944	0.1188	0.0999	0.05
F4	Prob>ChiSq	0.508	0.509	0.5778	0.3513
	Power	0.0967	0.112	0.0778	0.1502
F7	Prob>ChiSq	0.6374	0.8157	1.0	0.3061
	Power	0.0691	0.0522	0.0598	0.0862

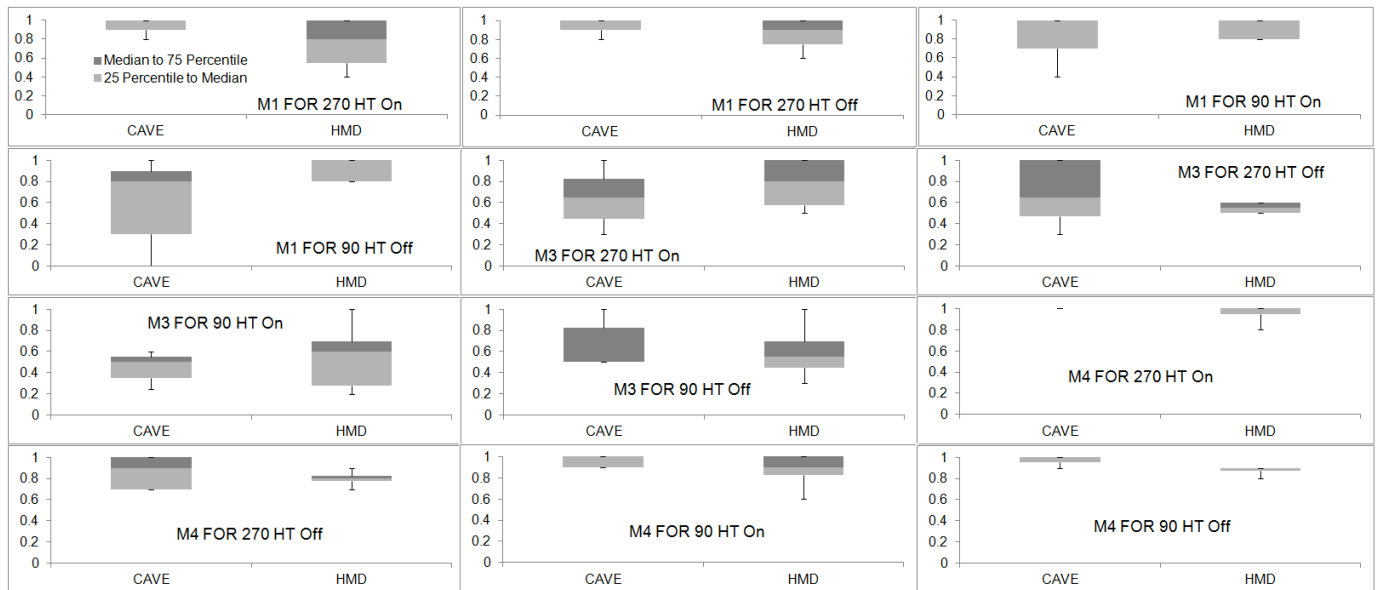


Figure 20: Box plot comparison of grades of mouse limb dataset from the CAVE experiment and the recent HMD experiment

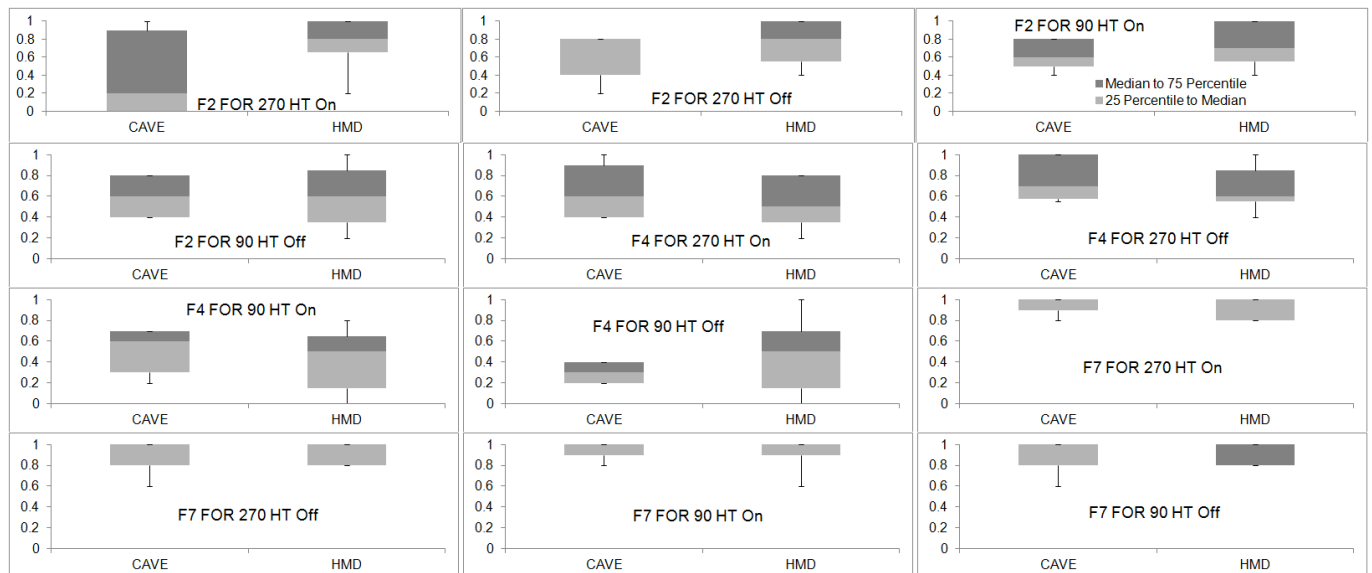


Figure 21: Box plot comparison of grades of fossil dataset from the CAVE experiment and the recent HMD experiment

## 4.3 Discussion

### 4.3.1 MR Simulation Validity

Looking back at our research questions for this study (section 3.1.1), we have found some evidence for, and some evidence against, our first hypothesis that MR simulation produces valid results even with very different simulator platforms (see Table 10).

It should be noted that our studies have limited scope, examining only two components of display fidelity, a few task types, and a particular application domain (volume visualization). It is not possible to definitively prove the validity of MR simulation with these experiments.

In favor of validity are the results comparing the grades for the quantitative tasks in the two experiments. Of the 24 different combinations of system characteristics and tasks that we compared from the two experiments, we found that the measured grades were not significantly different in 20 cases (83.3%) with a contingency analysis and 23 cases (95.8%) with a Mann-Whitney test. This is evidence that, overall, our two MR simulator experiments produced consistent results for task performance in individual conditions, even though the two MR simulator platforms were quite different.

The box plots shown in Figure 20 and Figure 21 give a quick and comprehensive visual comparison of the grades obtained by the participants in the two studies. In a few cases, we notice a big variance in the grades from one of the MR simulator platforms, such as for task F2 with FOR 270 and HT on. It is hard to make any judgments in these cases. But in many cases we can judge that the plots are very similar. These plots are visual evidence of the similarity between the grades obtained from the two MR simulator platforms, to complement the statistical analyses.

Half of the tasks from the two datasets were qualitative in nature, which changed from the previous study to become multiple-choice questions. As verbal reports would require more cognition to form the answer, it might have been easier and/or quicker for the participants to answer these in the HMD experiment. We thus traded off the chance to compare the results of these tasks between the MR simulator platforms, for more objectivity in the results of this study.

On the other hand, we found that most (six out of seven) of the significant effects found in the original experiment were not reproduced in the current study (section 4.2.2.1). This argues against the validity of MR simulation experiments run on simulator platforms that are highly different than the actual target platforms. It is important to understand what might have caused these differences. Table 10 listed the differences in the MR simulator platforms we used. We also noted that our experiments had a relatively small number of participants in each condition, differed in the male to female participant ratio, and used participants with different mean spatial ability scores.

We have no direct evidence to tell us which of the differences in the simulator platforms or participants might have caused the differences in results between the two studies, but we can speculate. We believe that participant spatial ability may have played an important role, especially in the lower-display fidelity conditions, but spatial ability alone is probably not enough to explain the differences. Based on observation and experience with the two systems, we believe that occlusion, weight, and stereo display technology may have been the most important technological differences between the platforms.

The total occlusion of the real world (including the participant's body) in the HMD platform means that users are more tentative about turning and walking, and are more likely to feel disoriented. Prior MR simulation studies measuring presence, and anxiety using a virtual 'pit' have found that the ability to see the user's body (Juan and Pérez 2009; Gandy *et al.* 2010) or a lack of it (Meehan *et al.* 2002; Juan and Pérez 2009) could affect presence, and thereby movement of users in the environment.

The weight and general encumbrances of the HMD platform also contribute to these phenomena. Many of the users felt that the HMD was heavy, and were concerned about its tethering. In particular, in the non-head tracking conditions, the users in the HMD experiment might not have moved as much as the participants in the CAVE experiment, who could see their whole body, as well as lacked the tethering, and the weight of the HMD on their shoulders. We wonder whether having a lighter and wireless HMD simulating the various conditions in our MR simulation would have produced results more similar to those from the CAVE experiment.

The Infitec stereo technology used in the CAVE study is a passive stereo approach that (anecdotally) results in color perception problems in some users. Stereo effects may be more pronounced or easier to view in the HMD platform, perhaps reducing the need to rely on head tracking.

We intentionally chose two MR simulator platforms that were very different, in order to test how far we could go without compromising the validity of the results. It seems that the differences in our CAVE and HMD simulator platforms were too pronounced, so future studies should use MR simulators that have fewer differences from the target platform.

#### **4.3.2 Effects of FOR**

For our second research question regarding the effects of FOR, we have significant results from this study supporting our second hypothesis. We found that HT was more effective with higher FOR (360 and 270). The combination of high FOR with HT on was significantly better than the 90 FOR with HT off for M4 grade, which was a spatially complex search task. Additionally, we observed that grades improved considerably from FOR 180 to FOR 270 with HT on for M4. These results suggest that for complex search tasks, the combination of high FOR with head tracking provides the best conditions for accurate visual analysis, probably because it allows users to physically walk around the dataset, rather than rotating it, to view it from different directions.

To take advantage of an increased FOR, a user needs to look from different distances in different directions. In the current study, we found in most tasks that the participants moved much less than in the study in the CAVE. Thus, the only significant effects of FOR that we found were on tasks F1 and F3, both of which were of general description type (see Table 12).

We also observed that participants' confidence levels were consistently higher with conditions similar to desktop displays (FOR 90 HT off) or to the real world (FOR 360 HT on) than other conditions (section 4.2.1.3). This is consistent with the results of other studies that have found benefits of the most familiar conditions (McMahan *et al.* 2012).

#### **4.3.3 Implications for design**

From our two studies, we can extend our previous guidelines for designing VR systems for improving effectiveness for visual analysis of volume data (Laha *et al.* 2012):

- VR systems with FOR 270 degrees or above with head tracking are useful for spatially complex search tasks with volume datasets.
- VR systems with fewer encumbrances might produce more significant benefits of higher display fidelity for visual task analysis with volume datasets.

For researchers studying the effects of display fidelity on visual task analysis with volume data, we recommend the use of the MR simulation approach for creating more generalizable results, but the MR simulation platform used in these studies should be similar to the target platform for the application.

## 4.4 Conclusions

Our goal in this research was to examine the validity of MR simulation by comparing the results from two similar experiments using quite different MR simulator platforms (CAVE and HMD). Revisiting our research question - 1 (see section 1.4.1), we have presented a variety of evidence both for and against the validity of the MR simulation approach for empirical studies with volume datasets. It appears from this study that the effects of the simulator platform cannot be fully discounted, and that it's likely best to use an MR simulator platform that is more similar to the platform being simulated.

A secondary goal of this work was to learn more about the effects of FOR on visual analysis tasks with volume datasets (see section 1.4.1). We observed that beneficial effects of FOR depend on physical movement. VR hardware such as HMDs that have greater weight, less visibility of the real world, and other encumbrances can cause reduced movement and might lessen the positive effects of FOR on visual analysis tasks. Even with the HMD platform, however, we still found at least one task category where high levels of FOR combined with head tracking resulted in improved task performance.

The re-categorization of the tasks (see Table 27) based on our generic task categorization (Chapter 8) shows that the extra space provided by the higher FOR again improved search (mildly in M1) and shape description (F1). Better depth perception through motion parallax provided by HT aided in spatial judgment (M4) and pattern recognition (F5); in both cases the gaps between two structures (between blood vessels or between cell boundary and intercellular body) were important for the correct response. Together, extra space for analysis (higher FOR) and better depth perception through motion parallax (HT) aided in tasks involving arbitrary 3D gaps between two similar structures (M4) or searching for presence or absence of structures (F4).

To have broader evidence of the effects of display fidelity on volume visualization, we want to know the effects on other rendering styles as well, such as isosurfaces (see Chapter 5).



## Chapter 5. VR Study-III: Effects of VR System Fidelity on Analyzing Isosurface Visualization of Volume Datasets

While both the previous studies (VR-Study-I, in Chapter 3, and VR-Study-II, in Chapter 4) produced significant results tied to system components, the rendering of the data in both cases used 3D texture (semi-transparent rendering). However, there are numerous other modes of rendering a volume, and we wanted to explore the combined effects of system fidelity and volume rendering technique on task performance. We also have been using datasets from medical biology and paleontology domains. To have broader and generalizable results we wanted to use datasets from some other domain. This inspired our VR-Study-III in which we used micro-CT datasets from biomechanics domain (see Figure 22). Our goals were to study the effects of FOR, ST and HT on visual analysis of isosurface rendering of volume data, and evaluate the effects of system fidelity on a wide variety of abstract task types.

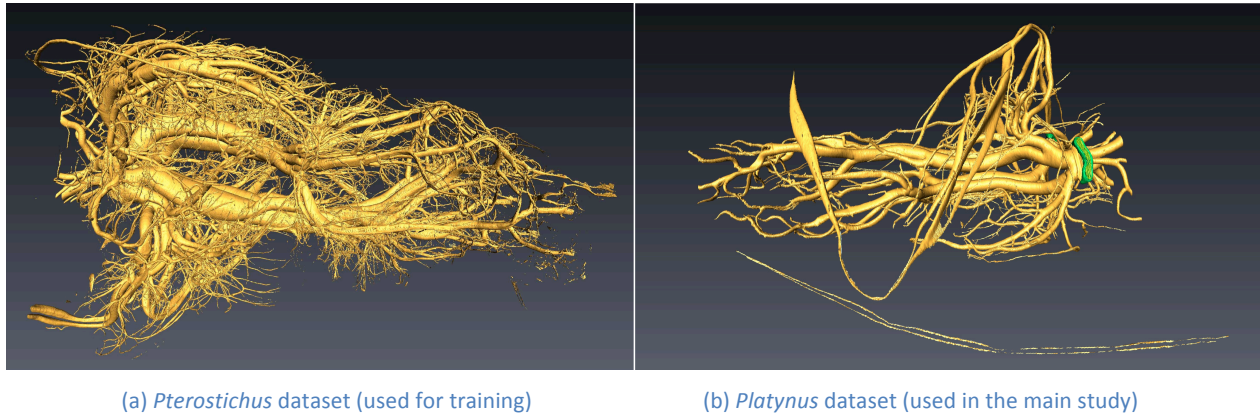


Figure 22: Isosurfaces of tracheal systems generated from micro-CT scans of beetles used in our VR system fidelity study.

From VR-STUDY-III we found several single-factor and multi-factor effects on a wide variety of abstract tasks with volume visualizations. We found that higher levels of system fidelity produced improved task performance. In particular, stereo had the strongest effects on task performance (among FOR, ST and HT) on several search and spatial judgment tasks. FOR improved performance in two spatial judgment tasks, and HT improved confidence in one search task. We also compared our findings with those from our VR-STUDY-I, and got some indications that the effects of VR system fidelity may vary depending on the rendering technique used to visualize the volume. We describe the details of VR-Study-III here (Laha *et al.* 2014).

### 5.1 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.

#### 5.1.1 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 (Slater 2003; Bowman and McMahan 2007). 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) (Chen *et al.* 2012; Laha *et al.* 2012; Ragan *et al.* 2013), 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 (90° and 270°), 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 3.1.4) 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 (Laha *et al.* 2012).

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 (Zhang *et al.* 2001; Zhang *et al.* 2003; Prabhat *et al.* 2008); but there are some results against this as well (Demiralp *et al.* 2006).
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 (Laha *et al.* 2012; Ragan *et al.* 2013), while others challenge it partially (Chen *et al.* 2012).
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 (Schuchardt and Bowman 2007; Laha *et al.* 2012). 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 5.1.4).

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 (Bowman and McMahan 2007).

### 5.1.2 Datasets

Micro-CT ( $\mu$ CT) 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  $\mu$ CT devices are not as powerful as  $\mu$ CT conducted at 3rd-generation synchrotron light sources, which yields the highest quality  $\mu$ CT data currently available (known as SR- $\mu$ CT, (Socha and De Carlo 2008; Westneat *et al.* 2008)). 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- $\mu$ CT datasets (see Figure 22) 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 (Westneat *et al.* 2003; Socha *et al.* 2008); 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 (Socha *et al.* 2010).

Although SR- $\mu$ CT 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 the x-rays did not resolve these, 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 (Chapman 1998).



Figure 23. A participant in the FOR\_ST\_HT condition inside the CAVE.

We used Avizo<sup>1</sup> 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 5.1.3.1).

### 5.1.3 Apparatus

#### 5.1.3.1 Hardware and Software

We used a four-screen CAVE-like system (Figure 23) (Cruz Neira *et al.* 1993) with three rear-projected 10' by 10' walls, and a top-projected floor, each with passive Infitec<sup>2</sup> 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<sup>3</sup>, which also tracked a wireless wand with five buttons and a joystick.

We used open source software to interface with the hardware. DIVERSE (Kelso *et al.* 2003) provided support for distributed rendering on our cluster of computers running the CAVE system. VRUI (Kreylos 2008) 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<sup>4</sup> lab for rendering the isosurfaces of the volumes in the CAVE.

#### 5.1.3.2 User Interactions

We provided users a grab interaction with six degrees of freedom about the absolute position of the grab. Pressing the trigger button at the bottom of the wand with the index finger could activate this. In addition to the grab action, the users in the head tracked conditions (HT; see Table 18) could also use positional head tracking to get different viewpoints around the datasets

---

<sup>1</sup> <http://www.vsg3d.com/avizo/overview>

<sup>2</sup> <http://www.infitec.net/>

<sup>3</sup> <http://www.intersense.com/>

<sup>4</sup> <http://idav.ucdavis.edu/~okreylos/ResDev/KeckCAVES/Applications.html>

based on their absolute head movements inside the CAVE system, which gave them an added mode of interaction.

#### 5.1.4 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, palaeontology, 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
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 (Forsberg *et al.* 2009; Laha *et al.* 2012; Laha *et al.* 2013). The arguments supporting this choice include expert participants self-reporting as novices in prior studies (Laha *et al.* 2012), and the fact that volume data analysis requires significant training (Lee *et al.* 2010b), indicating that many of these domains have scientists who are similar to novices (Stock 2008). 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.

### 5.1.5 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 270° (all four walls of the CAVE system used to render the isosurfaces) and 90° (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 18. 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 (Bowman and McMahan 2007), at the same level (Slater 2003; Bowman and Raja 2004). 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.

### 5.1.6 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 (Ekstrom *et*



*al.* 1979)) 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.

### 5.1.7 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 (Ekstrom *et al.* 1979). 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 (Figure 22-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 (Figure 22-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.

## 5.2 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

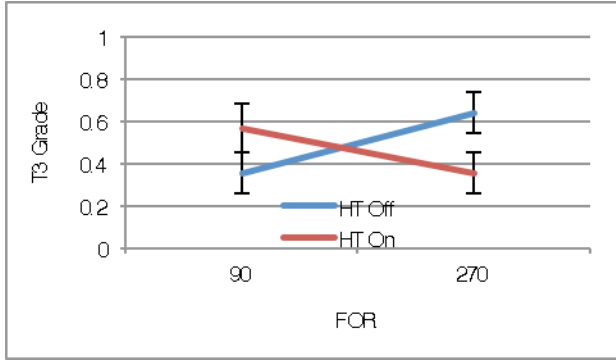


Figure 24. Interaction between FOR and HT for Grade in T3, a Quantitative Estimation Task.

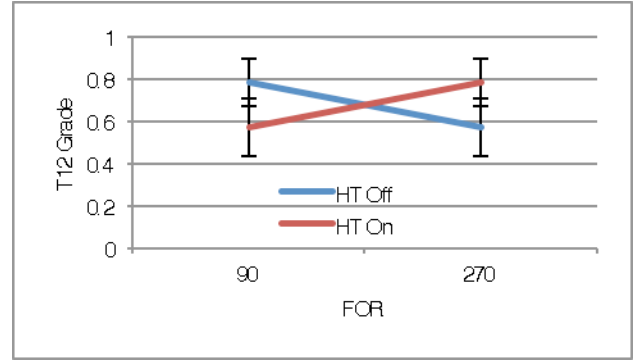


Figure 25. Interaction between FOR and HT for Grade in T12, a Pattern Recognition Task.

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 (Prabhat *et al.* 2008; Laha *et al.* 2012), 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 5.3.2).

The significant main and interaction effects of the display components on the various task types are summarized in Table 26.

### 5.2.1 Grades (task performance accuracy)

We observed five significant main effects of FOR and ST on the grades obtained by the participants, shown in Table 19 below; higher levels of these improved accuracy of task performance in all cases.

Table 19. Significant Main Effects on Grades

Task: source	$\chi^2$	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 20, and shown in Figure 24 and Figure 25, 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 ( $p=0.0469$ ).

### 5.2.2 Completion time (speed of task performance)

There were several significant main effects of FOR and ST on the task completion time, shown in Table 21. Higher levels of VR system fidelity components improved speed of task completion in each case.

We observed a significant three-way interaction effect of FOR, ST and HT on the speed of completion of task T5 (see Table 22). 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.

### 5.2.3 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 23. 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 ( $p=0.0313$ ).

### 5.2.4 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 24. 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 20. Significant Interaction Effects on Grades

Task: source	$\chi^2$	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

T12: FOR & HT	4.9773	1	0.0257	for_ht	0.36
				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 21. 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 22. 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 23. Significant Interaction Effects on Perceived Difficulty of Tasks

Task: source	$\chi^2$	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

T11: FOR, ST & HT	5.211	1	0.0224	FOR_st_HT	6.4
				FOR_ST_ht	6.7
				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

#### 5.2.5 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 25. 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 360° field of regard. Also, few of the participants thought zooming was helpful in certain tasks, but more usable with surrounding visuals (higher FOR).

Table 24. Significant Interaction Effects on Perceived Confidence

Task: source	$\chi^2$	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 25. Significant Interaction Effects on Post Questionnaire Ratings

Effect: source	$\chi^2$	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

### 5.3 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, Figure 26 and Figure 27 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 26 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 5.2 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.

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.

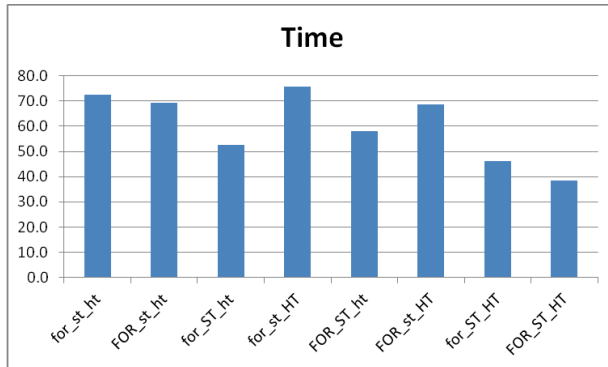


Figure 26. Average time in different conditions.

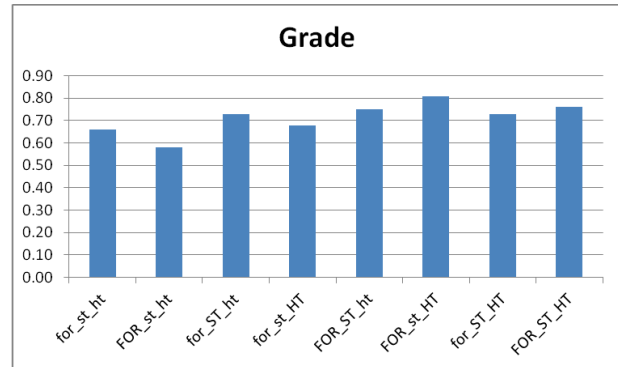


Figure 27. Average grade in different conditions.

#### 5.3.1 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

Table 26. 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—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—O—O		
T15 (quantitative estimation)												
T7 (shape description)					X					O—O—O		
T12 (pattern recognition)	O—O—O											

(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 (Laha *et al.* 2012; McMahan *et al.* 2012). 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.

### 5.3.2 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, 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 (Laha *et al.* 2012; McMahan *et al.* 2012), 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.

### 5.3.3 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 (Laha *et al.* 2012). 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 Figure 22-b), as through a dense network of vessels (Ware and Franck 1996; Barfield *et al.* 1997). 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 (Laha *et al.* 2012).

Table 27. Re-defining categories of tasks from our previous study (Laha *et al.* 2012)

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 3.1.4), as shown in Table 27.

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 (Laha *et al.* 2012)) 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 (McMahan *et al.* 2012; Ragan *et al.* 2013).

Our prior study also found significant three-way interactions between FOR, ST and HT for shape description tasks (Laha *et al.* 2012), 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.

#### 5.3.4 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.

### 5.4 Conclusion

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- $\mu$ CT volume datasets. Revisiting our research question - 1 (see section 1.4.1), 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 analysing isosurfaces, while FOR might improve analysis of semi-transparent volume rendering.

In VR study-I, the 3D texture rendering produced a very dense visualization of the volume datasets, which the participants analysed. The extra virtual space provided by the higher FOR might have helped to unclutter the dense data, which might be the reason why FOR improved the performance so much. But in this study, with isosurface rendering, the data was not that dense; so FOR was not able to provide as much improvement in performance. The isosurfaces spaced out the volume data in separate layers and surfaces of the tracheal tubes. The analysis tasks involved spatial judgment and understanding the gaps between the tubes. As a result, stereo with the better depth perception cues provided a lot of benefit.

From 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. We as a community of VR and visualization researchers need to identify and define abstract task categories (Chapter 8) cutting across various scientific domains, so that we can leverage that framework to empirically evaluate VR system fidelity on task performance with volume visualization (Laha and Bowman 2012).

In the following chapters (Chapter 6 and Chapter 7) we describe design of 3D interaction technique for answering research question - 2 (see section 1.4.2).

## Chapter 6. 3DI Design-I: The Volume Cracker Technique

As we discussed in Chapter 2, existing interaction techniques for analyzing volume data can be broadly grouped under two categories. The ones from the first group remove occluding voxels to expose internal structures. These techniques use slice planes or other regular (or irregular) geometries as cutting or sculpting tools. The techniques from the second group (Focus+Context) preserve all voxels at all times, but highlight the regions of interest in a volume using innovative visualization. Both these classes of techniques have disadvantages, and mostly do not work with raw volume datasets. This inspired us to come up with a list of usability criteria broadly outlining the interaction needs of volume data analysts. We held a series of design studios and discussions, within our 3DI research group and with domain scientists, and identified the high-level metaphor of cracking as the design inspiration for the Volume Cracker technique (see Figure 28). In this chapter, we describe the design and evaluation of this technique.

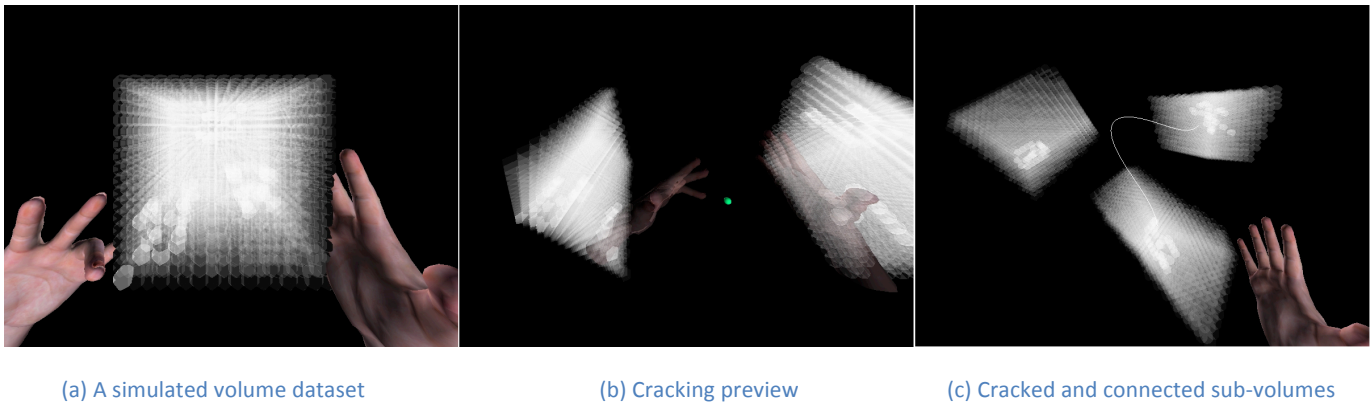


Figure 28: The Volume Cracker interaction technique.

### 6.1 Volume Cracker Design

#### 6.1.1 Goals

We set out to design a new interaction technique for volume data analysis that would address the problems of existing techniques. We wanted to allow the users to view internal structures while keeping all the voxels visible and not distorting the dataset in any way, and not requiring any prior segmentation.

During our interviews with researchers and our prior empirical studies for evaluating task performance with volume data, the researchers expressed a strong desire for 3D interaction techniques to directly interact with and analyze volume datasets for various research tasks they perform on a regular basis. They were of the opinion that having a technique where they could use their hand(s) would be more natural, faster, and easier than using indirect tools. Thus, our goal was to design 3D interaction technique(s) using direct manipulation.

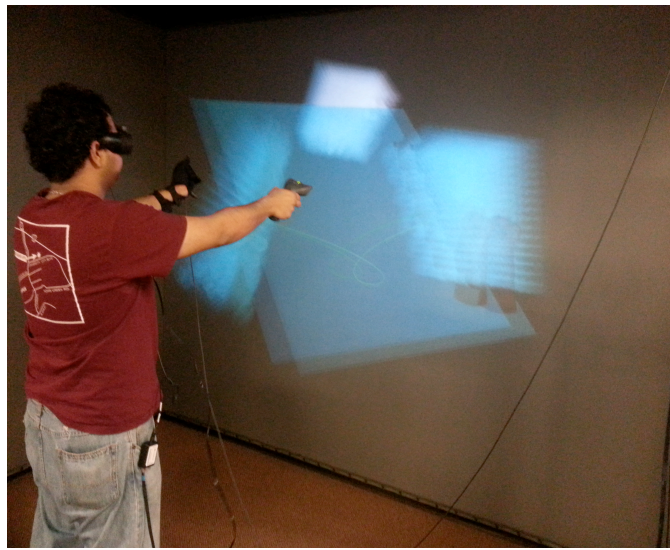


Figure 29: User inspecting a dataset with Volume Cracker.

## 6.1.2 Design

### 6.1.2.1 High Level Metaphor Selection

We used design studio sessions in the context of a university course to seek innovative ideas for interaction techniques mapped to a set of task categories (Laha and Bowman 2013a). We asked participants to brainstorm ways to use two hands to carry out such tasks in the real world (e.g., searching for seeds in cotton is similar to search tasks in volume data). We also drew inspiration from techniques found in the literature (McGuffin *et al.* 2003; Huff *et al.* 2006; Mlyniec *et al.* 2011). In particular, Zhou *et al.*'s magic story cube (Zhou *et al.* 2004) inspired our use of a cracking metaphor.

This resulted in a shortlist of metaphors to analyze more closely. These included knife, cracker, peeler, hinge spreader (McGuffin *et al.* 2003), and box spreader (McGuffin *et al.* 2003) techniques. In the knife metaphor, the dominant hand cuts open slices from a volume held by the non-dominant hand. The cracker technique uses two hands to crack open a volume, like cracking open a book. The peeler uses the fingers to peel off outer layers from a volume, revealing the inner layers. The hinge spreader uses a hinge, while the box spreader uses a resizable box to open up the volume. We created paper prototypes or storyboards of each technique to understand and discuss the affordances that each technique provides. To select the final metaphor, we looked at a set of usability criteria.

### 6.1.2.2 Usability Criteria

For the requirements of our technique, we had good reference from prior empirical studies in the literature with volume datasets (Laha *et al.* 2012; Laha *et al.* 2013), and from our interview sessions with domain scientists, from various domains such as geophysics, paleontology, medical biology, and biomechanics. The usability criteria were, in decreasing order of importance:

1. Allow the user to understand what is hidden inside a volume
2. Allow the user to look at smaller chunks of a volume more closely, to identify the regions of interest (ROIs)
3. Allow the user to maintain the spatial context
4. Do not distort any part of the volume
5. Do not assume that the data has predefined semantic layers

6. Make the results of user actions clear and predictable
7. Allow the user to quickly reverse or cancel their actions

Of the shortlisted metaphors, the hinge spreader and the box spreader remove occluding voxels, making the ROI more clearly visible, but violating the criterion to maintain the spatial context.

The cracker, peeler, and knife metaphors preserve all the voxels in the volume, while breaking it in useful ways for the user to analyze. The peeler, however, is based on the assumption of semantic layers in the volume.

The knife technique allows cutting along specific lines through the volume, like we cut open an apple using a knife, while we hold it with the non-dominant hand. The cracker allows cracking open a volume with two hands, and also cracking sub-volumes recursively. Both these techniques satisfied the first five usability criteria. The cracker technique, however, offers the possibility of informing the user of the result of the cracking action prior to cracking (see section 6.1.2.3 below), through a “cracking preview,” satisfying the sixth usability criterion. It was difficult to see how a knife metaphor might offer such a feature, since multiple cuts may be required to cut out a piece of the volume.

Further, while an asymmetric technique (knife) might offer more precision in the cut that we make to analyze the ROI closely, based on Guiard’s framework (Guiard 1987), a symmetric technique (cracker) could be faster, as we use the degrees of freedom offered by two hands instead of one (Leganchuk *et al.* 1998).

Based on this analysis, we chose to design and develop the cracker metaphor. The next section discusses various design decisions that we incorporated in the Volume Cracker, and how those address the various usability criteria.

#### 6.1.2.3 Design Details

The basic idea of cracking emerged from the first usability criterion—to understand what is inside a volume. Intuitively, like we crack open a book, we wanted to crack open a volume with our two hands, to look at the things inside. Cracking preserves all the voxels in a volume, and does not assume layers in a volume. It also does not distort any part of the volume.

We designed a cracking preview (see Figure 28-b), which showed up as long as the midpoint between the two hands was within the volume. This was created and updated in real time by dividing the volume into two halves on either side of the plane orthogonal to the line joining the two hands, and then by displacing each voxel by a fixed distance along that line, towards the hand closer to the individual voxel. The preview showed how the voxels (along with the internal structures that they form) moved from one sub-volume to the other interactively as the position and orientation of the crack moved. It was designed to make the resulting sub-volumes fully predictable, and ties back to the sixth usability criterion above. Interestingly, from user feedback, we found that the cracking preview also provided a quick and dirty analysis of the various internal structures as well.

Cracking required simultaneous closing of both fists, and broke the volume into two sub-volumes as shown by the preview. The user could choose to crack recursively, until she had separated out the internal structures in blocks of ROIs, which she wanted to analyze separately. To preserve the context after cracking (Balakrishnan and Hinckley 2000), we displayed a line joining the two sub-volumes (third usability criterion). Separating the ROIs in connected sub-

volumes served the second usability criterion—looking at smaller chunks of volume or ROIs more closely.

To address the final usability criterion, we introduced a bimanual grab function in the Volume Cracker. This allowed the users to grab connected sub-volumes, move them farther apart or closer together, and also join them back (when their bounding boxes overlapped, and the grab was deactivated). We added visual feedback for activation of bimanual grab (blue), and joining of sub-volumes (green).

If the user cracked a sub-volume into two sub-sub-volumes, the line joining the original sub-volumes disappeared, while a new line joining the two new sub-sub-volumes appeared. The users could only rejoin connected sub-volumes, which maintained the hierarchy of sub-volumes.

The other action available in the original design was a single-hand grab on the entire dataset, cracked or otherwise, for manipulating its position and orientation directly. The fist did not need to be within the volume to activate the single-hand grab.

From demo sessions, we found that there were structures on the newly revealed surfaces of both sub-volumes that users wanted to analyze simultaneously, but they were having trouble because the surfaces were facing each other. This became another important usability criterion as even the domain scientists want to look at separate ROIs together. Also, this would help to figure out if there was an internal structure that got split in two during cracking.

We thus introduced rotations of the sub-volumes in the bimanual grab function, which gets activated right at the time of cracking. This allowed the users to grab the connected sub-volumes with both hands, and open those up like a book (while cracking) to look at the interior surfaces simultaneously. We connected the rotated sub-volumes with Hermite curves, defined and updated in real time by the normal vectors coming out of the two hands at any point during a bimanual grab (see Figure 30-c).

We initially allowed the users to reposition and reorient the sub-volumes however they wanted. This easily got confusing because if a structure is broken in two parts, it will be almost impossible to understand it as a whole if the divided structures get rotated about the axis running longitudinally through the unbroken structure, as in that case, the outlines of the structure on the broken faces would align improperly against each other.

In the final design, therefore, we use constraints to preserve the relative orientation of the structures between the two sub-volumes. We limit the movements of the sub-volumes to be along a plane and rotations of the sub-volumes to be around the normal vector to this plane (Figure 30-c). We define the plane at the time of cracking, as the plane containing the vector joining the two hands, and the vector joining the center of the original volume to the viewpoint. We selected these vectors, as at the time of cracking, the resulting plane allows the user to open up the sub-volumes like a book, so that the interior surfaces can face the user. This also serves the seventh usability criterion much better, as the users can quickly reverse the action of cracking if desired.

To avoid confusion, the cracking preview only shows where the crack will be, and not how the sub-volumes can be rotated after cracking. In other words, the preview shows what sub-volumes will be created, and then the separation and rotation of the sub-volumes is done after cracking. Although this means the cracking preview does not show the final state of the sub-volumes after cracking, the user can still quickly reverse the action if desired.



We use head tracking in combination with the Volume Cracker, since it helps the user to get different viewpoints very easily around a volume (Laha *et al.* 2012), and serves the first three usability criteria well. We also use stereoscopic display for better depth perception.

## 6.2 Evaluation

### 6.2.1 Goals and Hypothesis

Our primary goal was to find out whether the Volume Cracker (VC) would improve quantitative task performance for volume data analysis, as compared to standard techniques. Thus, our first research question was:

1. How does VC perform as compared to a standard 2D and a standard 3D interaction technique for volume data analysis?

Further, we also wanted to gather qualitative insights on the usability and the design of VC, represented by our second research question:

2. How usable and easy to learn is VC?

Corresponding to the two research questions, we had two hypotheses:

1. *VC will produce significantly better task performance than other standard techniques for volume data analysis.* The design choices that we made while designing VC were all geared towards making volume analysis task performance faster and more accurate.
2. *VC will have comparable usability and learnability to existing techniques for volume data analysis tasks.* Based on the real-world metaphor of cracking with human hands, we believed VC would be easy to learn and intuitive to use.

To test our hypotheses, we designed an experiment to evaluate the VC and compared it against two existing interaction techniques for volume data analysis.

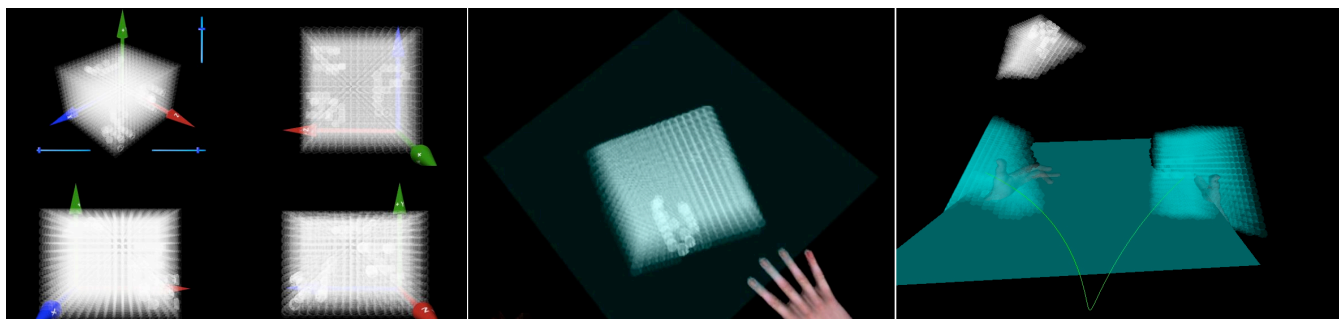
### 6.2.2 Preliminary Evaluation by Experts

Before running a more formal study, we invited four domain experts to evaluate VC. They included: a biomechanics professor who analyzes micro-CT scans of beetles and snakes, a medical biology doctoral student who works with micro-CT and nano-CT scans, a molecular biology doctoral student working with volume datasets generated from proteins, and an engineering science and mechanics doctoral student working with volume datasets.

We demonstrated a VC prototype to the experts. They appreciated being able to crack open a volume and segregate any region of interest to analyze the internal structures. All of them gave us examples of analysis tasks from their research domains where they felt the technique would be an improvement over the standard tools and techniques they use regularly.

We also verified with them that the rendering of our simulated volume datasets (see section 6.2.4) were comparable to the real datasets they are used to. We made the simulated datasets as large as we could ( $17^3$  voxels) without sacrificing real-time interactivity, which is needed for VC to work smoothly. They opined that if we observed any benefits in the speed of task completion by using VC with these datasets, it would further improve with real volume datasets. Based on their feedback and input, we created structures inside the simulated datasets similar to the ones in





(a) Axis Aligned Slicing (AAS) technique

(b) Arbitrary Slicing (AS) technique

(c) Volume Cracker (VC) technique

Figure 30: The three interaction techniques we compared in our study.

real datasets, and we used realistic analysis tasks from a volume analysis task taxonomy (Laha and Bowman 2013a).

### 6.2.3 Study Design

There are several existing interaction techniques for volume data analysis (see section 2.6). The wide variety of available techniques makes it difficult to evaluate the benefits of a new design in an absolute sense, but it highlights the need for empirical studies to compare the designs. We chose to compare VC against the most widely used desktop 2D interaction technique, and a widely known 3D bimanual interaction technique (Hinckley *et al.* 1994; Billen *et al.* 2008).

We called the desktop 2D interaction technique the “axis-aligned slicing (AAS) technique.” In AAS, the user looks at four views of the same data simultaneously. One is a 3D view of the volume. The user can slice it along three orthogonal axes, the views along which are called the axial, sagittal, and coronal views.

We termed the 3D interaction technique the “arbitrary slicing (AS) technique.” AS is currently used by several 3D volume visualization software packages like 3D Visualizer (Billen *et al.* 2008). In the AS, the user can rotate the volume data to any arbitrary 3D orientation. The user also has a slicing or cutting plane, which can also be rotated to any orientation, and can be used to slice through the volume to look at any arbitrary cross section of the volume.

### 6.2.4 Implementation

We implemented each of the three techniques in Vizard<sup>5</sup>. Because of implementing techniques like VC that manipulate individual voxels, we chose to create simulated volume datasets in the shape of a cube with  $17^3$  simulated voxels (represented by spheres). We varied the transparency between the various simulated voxels just like in real volume datasets. We verified with domain scientists that these datasets were comparable to the ones they use (see 6.2.2).

In our AAS technique (see Figure 30-a), we created four views, showing the simulated volume from four different viewpoints. The top left view showed a perspective view; the top right view looked down the vertical axis; the bottom left view looked down the left horizontal axis; the bottom right view looked down the right horizontal axis. We placed three sliders in the top left view, for slicing along the three axes. The three sliders and the corresponding cutting planes

<sup>5</sup> <http://www.worldviz.com/products/vizard>

could be used in combination, by clicking and dragging with a wireless desktop mouse, placed on a table in front of the display. The handle of each slider bar spanned 20x10 pixels. None of the participants reported or appeared to have any trouble selecting them with the mouse pointer.

In our AS technique (see Figure 30-b), the user could grab using her left hand. She had a cutting plane attached rigidly to a six-degree-of-freedom (6DOF) tracked IS-900 wand, held in her right hand. The plane removed the voxels from the simulated volume that were on the rear side of the plane. The user could also make the cutting plane static with the wand, and use the left hand grab to analyze the various cross sections easily. Mapping the cutting plane to a wand held in the user's hand allowed the user to manipulate the plane easily to orientations difficult to achieve with just the human right hand.

The VC technique used in our study (Figure 30-c) was implemented as described in section 6.1.2.

We ran the evaluation with a rear projected Visbox-X2 display, a 10 by 7.5 foot display wall with passive stereo capabilities provided by Infitec. An Intersense's IS-900 tracking system provided 6DOF position and orientation tracking of the head and hands of the users in the AS and VC techniques, with a hand tracker for the left hand, and a wand held in their right hand. We used a 5DT data glove for detecting the grab action with the left hand. The right hand grab action in the VC technique was mapped to the trigger button click of the wand.

We used the same display (see Figure 29) in our study for all three techniques. We used stereoscopic rendering and head tracking in AS and VC. However, we used monoscopic rendering with no head tracking in AAS, trading off the experimental control for ecological validity. Since three of the four views in AAS showed just 2D slices, it is unlikely that stereoscopic rendering, and head tracking would have made much impact.

#### 6.2.5 Participants

We recruited 17 unpaid voluntary participants, of whom three were pilot participants, one had equipment problems, and one retired due to sickness. The final study thus had 12 participants, of whom two were female. Their ages ranged from 19 to 33 years, with an average of 22 years. All the participants were graduate or undergraduate students, either native or fluent in English, and self-reported no prior background in volume data analysis (see selection of novices vs. experts in section ). The ordering of the three techniques (AAS, AS, VC) was counterbalanced, giving six different orderings. We had two participants with each ordering.

#### 6.2.6 Tasks

We are designing an abstract volume data analysis task taxonomy incorporating tasks from various domains under one umbrella (Laha and Bowman 2013a). The experts we invited in our lab for a qualitative evaluation of the VC reflected upon some of these task categories (see 6.2.2). Based on their suggestions, we initially chose four categories in which we hypothesized VC would improve task performance. After running three pilot participants, and informally testing the tasks ourselves with the three techniques, we chose two task types for the final study.

The first was a search task. In a real volume dataset, scientists search for blood vessels, joints in a tracheal network, bones, soft floating tissues, tumors, and many other structures (Laha *et al.* 2013). We chose to use letters from the English alphabet as an abstraction for such structures. This allowed us to recruit novice participants (see 6.3.1). We believed that this task was

important for evaluation, because a task defined by experts, and used in a prior study of volume data analysis (Laha *et al.* 2012) also asked participants to describe the shape of an internal structure as a letter from the alphabet. In each dataset, we created 13 letters, of various size and thickness, randomly oriented, and distributed in the simulated volume. Of the 13 letters, four letters were the same. The participants were told that there were at least two, but no more than six of that particular letter, and were asked to find all the instances of that letter.

The second task was a pattern recognition task. Domain scientists often seek patterns in different parts of volume datasets. For this they try to match complex shapes and clusters—they correlate and compare them to shapes in other parts of the dataset. To simulate this task, we created three closely matching coiled structures, all randomly oriented in 3D. We made the structures overlapping and interweaving in 3D, like in real volume data. One of these was a little different from the rest. The task was to find the odd one out.

We did not repeat the datasets between techniques to avoid learning effects. We also wanted to have two trials for each task, for each technique. Thus, we created six datasets for search tasks, and six more for pattern recognition tasks. We also created a training dataset with many letters and two coiled structures.

#### 6.2.7 Procedure

The Institutional Review Board of our university approved the study. After arrival, the participant signed an informed consent form, and filled out a background questionnaire. Then she was given a short introduction of the motivation behind the study, and went through a training session in the technique that would be used first. During the training, she spent three minutes practicing a search task, and three minutes practicing a pattern recognition task, with the training dataset. Next, she used the first technique to complete two search tasks and two pattern recognition tasks, each with different datasets. We created multiple simulated datasets of the same complexity level to avoid any learning effects between the techniques, and the datasets were counterbalanced between the participants and techniques. She was provided three minutes to complete each task. The experimenter used a stopwatch to record the time taken to complete each task. The same experimenter recorded the time for each participant, and we assumed that any time measurement errors were distributed evenly among the participants and techniques. The same experimenter recorded the response of each participant for each search task. For the pattern recognition task, the experimenter directly recorded whether the answer was correct or incorrect.

During the search task, the participants were asked to confirm their answer if they sounded unsure of what they were answering. This was done to make sure that the novice participants were not making passing guesses at what apparently looked like the structures they were searching for, but it also avoided over-counting. The experimenter reminded the participant before every search task that the letters could be big or small, and thick or thin.

After completing the tasks with the first technique, the participant had a short round of training with the second technique, with the training dataset. Then she completed the two search tasks and two pattern recognition tasks with the second technique. The participant then repeated the cycle with the final technique. Following that, the participant completed a post-questionnaire capturing her feedback on ease of use, ease of learning, and preference, all rated on a 1-7 scale for each technique, in addition to a few other usability and strategy-related questions for VC. The experimenter also recorded the participant's responses to different usability and design-related questions in a free-form interview.

From the pilot participants, we gathered useful strategies for the three techniques, which we passed on to the participants in the main study, in a way to give them a little bit of expertise that they wouldn't have. For the search task with VC, we recommended that the participants crack through the middle of the simulated volume to reduce the clutter of the dataset. For the pattern recognition task with the VC, we recommended that the participants separate the three structures into three sub-volumes, for easier comparison. For the AS, we recommended that they orient the plane towards the back of the simulated volume, so that the structures or letters towards the front became more distinguishable. We also suggested that they keep the plane static, parallel to the view, and through the middle of the dataset, and then use head tracking to look at both sides of the sliced dataset. With the AAS, we recommended using all three sliders together.

### 6.2.8 Results

The score metric was of ordinal numeric type, while the time was continuous numeric. The quantitative metrics in the post questionnaire (ease of use, ease of learning, and liking) were all of ordinal numeric type. For the time metric, we ran a one-way analysis of variance (ANOVA). For the other metrics, we ran an ordinal logistic regression based on a Chi-square statistic. If we found significant difference between the techniques, we employed post-hoc Tukey's HSD tests (for differences in least square means) for the time metric, and the two-sided Wilcoxon Signed rank test (with Bonferroni corrections), for all other metrics.

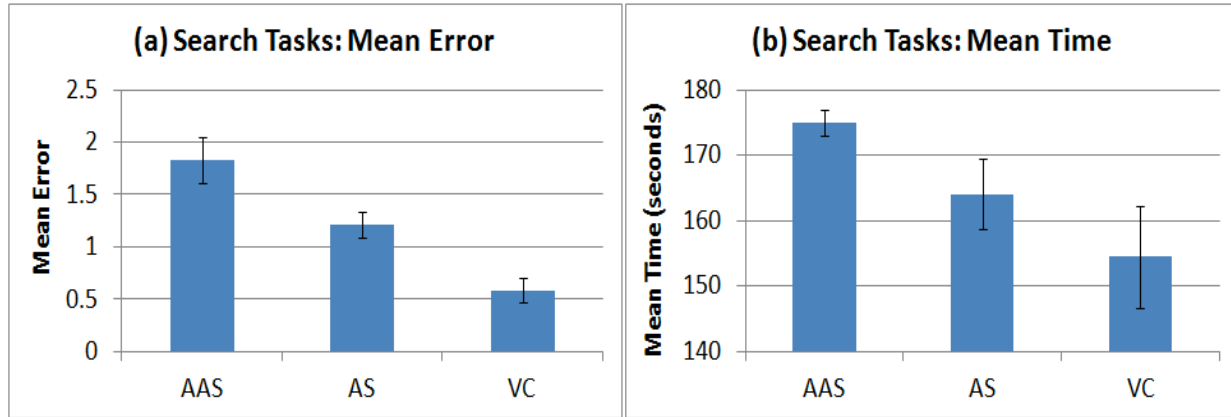


Figure 31: Performance metrics for search tasks.

#### 6.2.8.1 Search Tasks

The score metric for the search tasks had responses ranging from zero to four. We took the difference of the participants' response from four, which gave us the error produced by the participants. We found a significant effect of technique ( $\chi^2_{df=2} = 25.6202$ ;  $p < 0.0001$ ) on error. Post-hoc tests indicated that the errors decreased significantly from AAS to AS ( $p=0.0348$ ), AAS to VC ( $p=0.0003$ ), and AS to VC ( $p=0.0027$ ), as illustrated by the graph in Figure 31-a. The mean errors of participants with the three techniques are in Table 28.

We also found a significant effect of technique ( $F(2, 69) = 3.3915$ ;  $p = 0.0394$ ) on time. A post-hoc test indicated that VC was significantly faster than AAS for search tasks, as seen in Figure 31-b. There was no significant difference between the AAS and AS, or the AS and VC pairs. Mean times are in Table 28.

Table 28. Mean errors, scores, and time taken for the task types

Interaction Technique	Search Task Mean Error	Search Task Mean Time (seconds)	Pattern Task Mean Score	Pattern Task Mean Time (seconds)
AAS	1.833	175.00	0.2917	149.46
AS	1.208	164.08	0.7500	150.88
VC	0.583	154.45	0.8750	168.92

### 6.2.8.2 Pattern Matching Tasks

The score metric for the pattern recognition task was either 1 (correct) or 0 (incorrect). We found a significant effect of technique ( $\chi^2_{df=2} = 20.1322$ ;  $p < 0.0001$ ) on score. Post-hoc tests indicate that the score increased significantly from AAS to VC ( $p=0.0003$ ), but neither AS and VC nor AAS and AS scores were significantly different (see Figure 32-a; mean scores in Table 28).

We found a significant effect of technique on time ( $F(2, 69) = 3.5537$ ;  $p = 0.0340$ ), although post hoc tests found no significant difference in the time taken between the different techniques (also see Figure 32-b). Mean time spent is in Table 28.

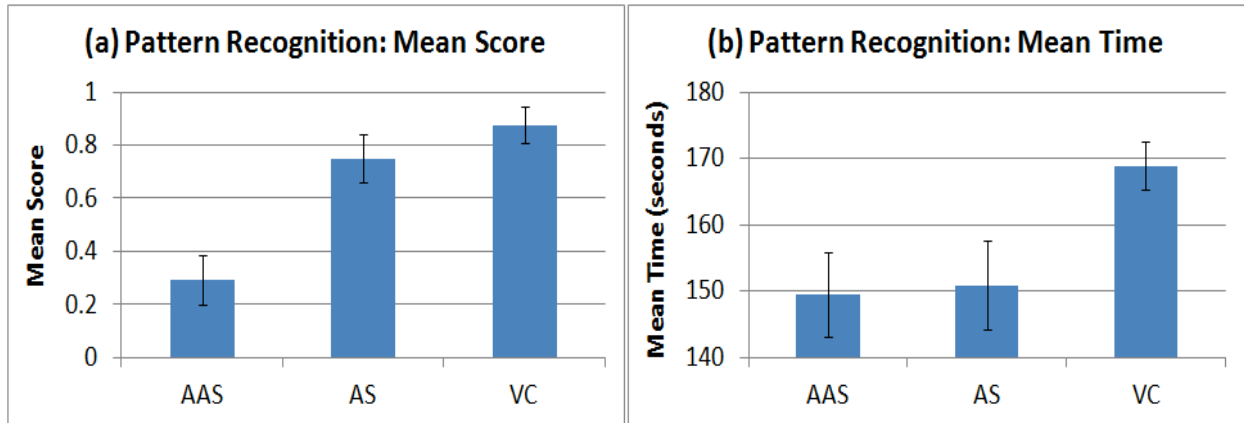


Figure 32: Performance metrics for pattern recognition tasks.

### 6.2.8.3 Post Questionnaire Results

The participants reported no significant difference between the techniques for ease of use or ease of learning. However, the participants reported a significant difference ( $\chi^2_{df=2} = 17.5265$ ;  $p = 0.0002$ ) in their preference for the three techniques—they liked VC significantly more than AAS ( $p=0.0411$ ).

## 6.3 Discussion

Looking back at our research questions, we have found evidence that partially supports and partially refutes our first hypothesis that VC improves task performance over standard techniques

for search tasks and pattern recognition tasks. For the search tasks, VC had the lowest mean error and was also the fastest, which strongly supports the first hypothesis for the search tasks.

For the pattern recognition tasks, VC had the highest absolute score (0.875), but was the slowest, which could have been due to a speed-accuracy tradeoff. With VC, participants took quite some time to crack the volume such that the three structures were in separate sub-volumes. We believe the time spent on this would reduce with more practice. The AAS had significantly lower times of completion because several participants gave up on the pattern recognition task with AAS, as they felt it was really difficult. The graphs plotting the mean scores and errors (Figure 32-a, Figure 31-a) indicate that in the AAS the participants scored very close to the score that would have resulted from random chance (2.0 for search error, 0.33 for pattern recognition score).

We believe that the specific design choices that we made (section 6.1.2.3) supporting our different usability criteria (see section 6.1.2.2), contributed to the observed benefits of the VC technique, like the basic function of cracking a volume in two halves served to instantly reduce the clutter inside the volume, so that several participants could identify all the letters inside correctly in as little as 40 seconds. Most of them took the remaining time just to join the sub-volumes back, break at some other point, and recount to confirm to themselves, before giving a final answer.

For the pattern recognition task with the VC technique, once they had broken the volume to separate out the shapes in distinct sub-volumes, it was quite easy to determine the right answer. Out of 24 trials, they were incorrect in only 3 cases. Looking back closely at the results, we found that in two of those three incorrect cases they ran out of time. Also, we constrained the bimanual movement only on a plane as described in section 6.1.2.3. For the pattern recognition task, if we removed the constraints, the participants could actually rotate the sub-volumes so that they could align the coiled structures to verify visually if they matched.

The post-questionnaire results support our second hypothesis. The participants liked VC significantly better than the standard AAS technique. Even though the VC technique challenged the participants with more gestures to recall and use (single-handed grab, crack, bimanual grab, join) than both the AAS (orthogonal slicing) and the AS technique (unconstrained slicing, single-handed grab) combined, the participants reported no significant difference in learning or usability for the three techniques. VC was easy to learn, participants preferred it, and it had good performance in most cases. We believe that the use of a real world metaphor that was natural to the users contributed to these results.

Participant comments included concerns about the lack of precision while cracking, blocking of the view by the hands, and difficulty in understanding the depth of the hands. The lack of precision might have also contributed to the slow task performance with VC. We are planning to address these issues in the next iteration of VC, by using an asymmetric interface, and reducing visibility of the 3D models of hands.

### 6.3.1 Novice vs. Expert Participants

We initially evaluated our design with experts (see 6.2.2), but chose to run the main study with novice participants (see 6.2.5). This decision is in line with arguments made by researchers running empirical studies for evaluating task performance with volume datasets (Forsberg *et al.* 2009; Laha *et al.* 2012; Laha *et al.* 2013). There are not very many expert users of volume datasets, and researchers find it hard to locate domain scientists as study participants. The field



of micro-CT is relatively new, domain scientists have little experience in analyzing volume data (Stock 2008), and training activities are actively being pursued (Mishra *et al.* 2012). Laha *et al.* found that none of the experts in their study self-reported as *experts* in the background survey questionnaire (Laha *et al.* 2012).

We could argue that experts might perform better than novices with the AAS technique. But the results of our study provides evidence that AAS at the least requires a lot of training for *search* tasks, much more than what is needed for AS and VC. Also, as discussed in 6.2.7, we gave the novices useful interaction strategies in each technique in a way to give them a little bit of expertise when they began to use each technique.

### 6.3.2 Simulated Datasets vs. Real Datasets

Using simulated volume datasets for evaluation studies is not ideal, but given our rendering hardware constraints, we believed it to be a reasonable choice. We kept the size of the datasets as big as possible ( $17^3$ ) without sacrificing real-time interactivity, but they were much smaller than real data ( $256^3$  or higher). We verified that the design of the simulated datasets closely corresponded to real datasets by evaluating them qualitatively with scientists from four different domains (see 6.2.2). We designed the simulated structures inside these datasets to be similar to those from real data, with internal objects coiling and overlapping in 3D. All four scientists who verified the structures opined that they closely resemble structures from real datasets that they work with. They were also of the opinion that the results we found would translate to their individual domains, but also encouraged us to find ways to work with real datasets. Currently, we are working on rendering real datasets in such a way to access and manipulate individual voxels interactively as needed by the VC, to prototype and evaluate VC with real datasets.

## 6.4 Conclusion

The contribution of this chapter is the Volume Cracker (VC), which is a bimanual interaction technique for generating exploded views of volumes (Bruckner and Groller 2006). VC works directly on raw volumetric datasets, prior to any form of segmentation or definition of selection objects or ROI. We discussed in detail the various decisions that we made while designing the Volume Cracker, and reported the results of a user study comparing VC with the AS and the AAS techniques.

Revisiting our research question 2 (see section 1.4.2), we found that VC had significant advantages over both the other techniques for search and pattern recognition tasks. We have contributed a deeper understanding of the usability criteria for volume data analysis interaction techniques, and a proven exemplar design satisfying those criteria.

To address some of the issues with VC, we designed a bare-hand version of VC with the Leap Motion controller. We evaluated the new VC design with real volume datasets and expert participants using the task categories from our abstract task taxonomy (see Chapter 7).

## Chapter 7. 3DI Design-II: The Bare Hand Volume Cracker Technique

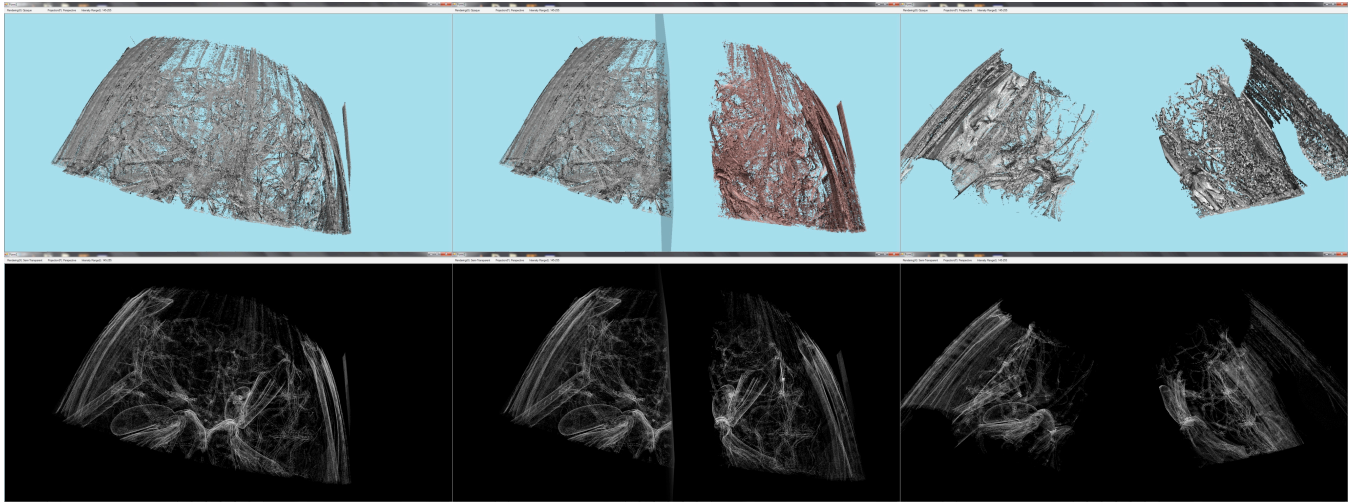


Figure 33: Bare Hand Volume Cracker –  $\mu$ -CT scan of a beetle, cracking preview and cracked beetle, in opaque and x-ray modes.

The Volume Cracker (VC; see Chapter 6) is a free-hand bimanual interaction technique that works directly on raw volumetric datasets (Laha and Bowman 2013b). Here we define *free-hand* techniques as ones where the hands of the user move freely in 3D space. This is designed to aid the visual analysis of raw scanned data obtained from modalities such as CT, MRI, PET, etc., before going through any data processing or cleaning pipeline, such as *segmentation*.

VC allows to recursively crack open a volume using intuitive real-world gestures, for isolating smaller regions of interest, which has the features an analyst would want to explore closely. This helps to reduce clutter in a dense volume, but explore hidden and occluded structures inside. By preserving all of the volume at all times, VC overcomes the problems of prior approaches, which relied on removing occluding portions of the volume, but in the process removed valuable context information important for an overall understanding.

The first prototype of the volume cracker was developed using 5DT data gloves and Intersense IS900 hand trackers. For evaluation of the design, we used simulated raw volumetric datasets (Laha and Bowman 2013b). In the evaluation study, we compared volume cracker to axis-aligned slicing and arbitrary slicing techniques. The results showed that volume cracker has significant advantages over those techniques for search and pattern recognition tasks.

Although the first prototype of volume cracker showed significant potential for improving task performance, the results could be challenged in several important ways. The study participants were non-experts, and the datasets were simulated (as opposed to real volume datasets). Also, the volume cracker prototype used the tethered 5DT data gloves and hand trackers. The cumbersome wearable trackers might have restricted the movements of the users, affecting many design decisions and the evaluation of the performance in important ways.

In this chapter, we introduce a Bare-Hand version of the Volume Cracker technique, nicknamed BHVC. We define *bare-hand* techniques as ones where the hands of the users are tracked without using any wearable or tethered device. This work extends the volume cracker design in important ways, and addresses the limitations of the prior evaluation study.



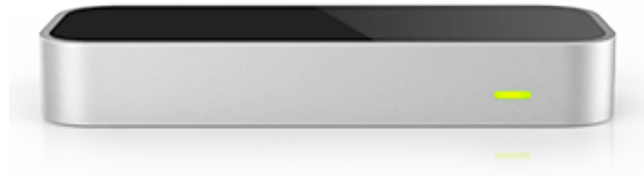


Figure 34: Leap Motion controller

## 7.1 Bare-Hand Volume Cracker

### 7.1.1 Overcoming prior problems with the Volume Cracker

We made some important changes in the design and the interface of the volume cracker. By using the Leap Motion controller (or simply, the Leap)(Leap-Motion 2012), we were able to avoid the cumbersome tethering, localization, and the high price range of the wired IS900 trackers, and the 5DT data gloves, used in the previous VC prototype. The tracking of the bare human hands and the fingers with high accuracy and low latency by the Leap inspired us to name this version of the VC the Bare-Hand Volume Cracker or simply, the BHVC.

We used the constructor SDK (cSDK) from NGRain (NGRAIN-constructor-SDK 2014) to render real-volume datasets. The cSDK rendered our datasets in real-time frame rates, while allowing us access to individual and groups of voxels at run-time. This was critical for the design of the BHVC.

We worked closely with a biomechanics faculty, who analyzes the fluid flow in the tracheal tubes of insects. We used datasets from his lab (see Datasets), and designed the current interface and evaluation of the BHVC around the tasks important to his research (see Evaluation). We were also able to secure the help of 12 researchers from his lab to participate in our evaluation study of the BHVC.

### 7.1.2 Design Challenges with the Leap

The Leap is a low-priced hardware that provides 6DOF tracking of bare human hands and fingers. Its accuracy-latency level is comparable to some of the best 3D trackers currently available. But it has several limitations, primarily stemming from its infrared camera use. It sits on a desk and has an inverted cone-shaped field-of-view (FOV), which limits the boundary for the users' hand movement. The best tracking requires placing the users hand at least five inches above the device, avoiding bright sunlight, avoiding metallic objects on the forearms, and leaning forward (it detects human heads). The detection of a hand orientation by the Leap depends a lot on the number of fingers tracked. This causes problems with detecting orientation of closed fists, and reduces the postures available for 3D interaction.

Occlusion is another important source of problems. Since the tracking by the Leap needs clear line of sight, one hand can occlude the other, or palm can occlude the fingers, or fingers can occlude other fingers. These problems limit the hand and wrist movements, and severely limits the gestures available for designing with the Leap.

Finally, the Leap assigns hand and finger ids to the tracked body parts. These may randomly vary between frames posing interesting design challenges. The jitter between frames can also cause

the palm normal (calculated by the Leap) to invert, and the number of fingers detected to change unexpectedly, between frames.

In spite of all these problems, its unique capabilities of tracking bare-hand and fingers inspired us to work around the problems and come up with new design decisions.

### 7.1.3 Bimanual-symmetric interface design for BHVC

We first attempted to directly port the prior VC design (Laha and Bowman 2013b) with the Leap and cSDK. The main features of VC were the unimanual grab and selection modes, the bimanual volume-cracking preview mode, and the cracking gesture. Mapping the corresponding gestures directly with the Leap was not possible as the detection of closed fist orientation with the Leap was far from reliable, but the grab gesture needed 6DOF tracking of the closed fist. Further, bimanual interaction with the Leap was severely impaired due to occlusion problems, while the strength of the volume-cracking preview arguably came from the 3D movement of two hands in free space.

We started exploring the bare-hand interaction design space, balancing intuitiveness with the choices that the Leap offered. System control commands could be mapped to the different number of fingers detected by Leap. But the fluctuations in the detected number of fingers (sometimes varying by one or two between frames) demanded a gap of at least two or more fingers between meaningful postures, for distinct and robust detection by the Leap. The two postures we found reliably distinct were with one finger open and with more than one finger open (open-palm).

The grab mode of VC can use a lot more freedom of rotation than the selection mode. The Leap finds it quite hard to detect orientation changes when tracking single finger. So we decided to use single-finger selection and open-palm grab (or its equivalent). Although this may appear counter-intuitive (in real-world, we grab with our fist closed, and not with open-palm), we decided to trade-off intuitiveness for freedom of movement and tracking.

Since 3DOF absolute rotation with the Leap is limited to around 100 degrees about x and z-axes,

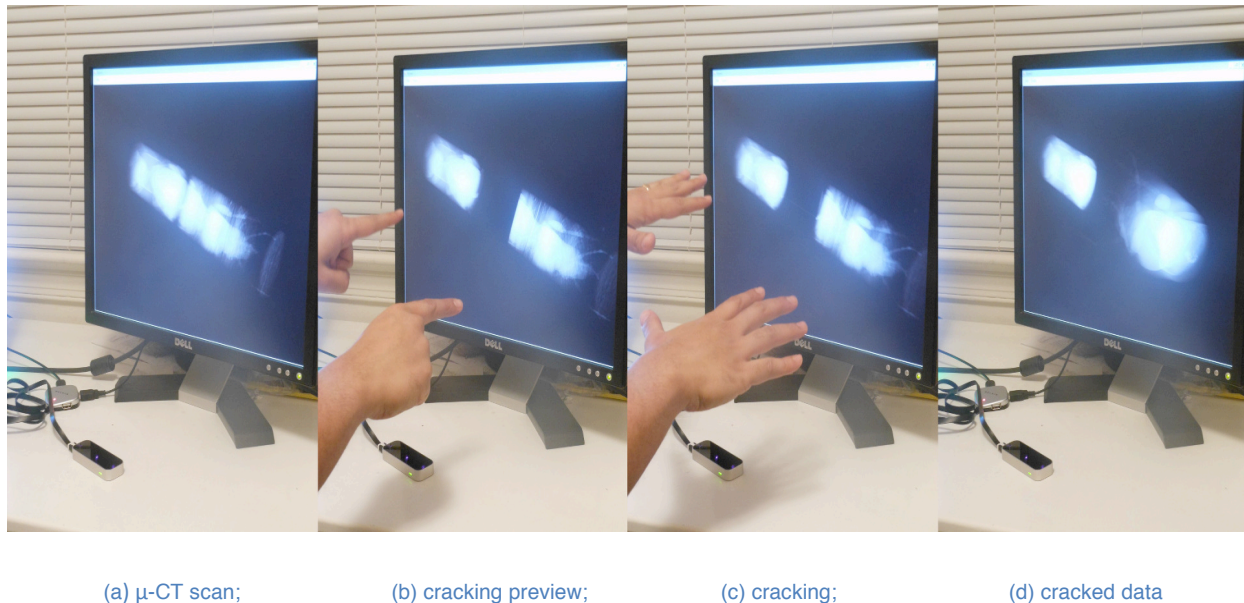


Figure 35: The bimanual-symmetric interface of the BHVC

we tried position-controlled non-isomorphic 3D rotation (amplified twice, and thrice) and rate-controlled 3D rotation (starting beyond a threshold). In the amplified non-isomorphic 3D rotation, as the user rotated her hand in the real world, she saw the camera in the virtual world rotating twice or thrice the amount of rotation in the real world. But the quick rotation was either too fast to follow, or the tracked space for rotation fell short of what was needed.

In the rate-controlled 3D rotation, we kept a neutral zone above the Leap, within which when the user kept her hand, no rotation happened. When she tilted the hand in some direction, we started the camera rotation if the tilt-angle went beyond a certain threshold. By experimenting with different threshold values, we found a certain value about each axis that had a good intuitive-performance combination. We finally settled for the 3DOF rate-controlled rotation, with two levels of speed (two different thresholds), around each of the three orthogonal axes. We used 3DOF position-controlled translations.

For the bimanual cracking preview and cracking gesture, we had two choices: single-finger and open-palm postures with each hand. We wanted to have bimanual grab activated immediately after the cracking occurs, like in the original volume cracker. We thus used the bimanual single-finger pointing in the cracking preview mode. We chose the cracking gesture as the simultaneous opening of the fists to the open-palm position, from the bimanual single-finger cracking preview. The cracking preview mode suffered from the occlusion problems with the Leap, but far less with the bimanual single-finger posture than with bimanual open-palms.

We preserved other features of the VC such as restricting the movements of a sub-volume on a plane, joining back of the sub-volumes, and the hierarchy of cracking.

#### **7.1.4 Evaluation of the bimanual-symmetric design of BHVC**

Our collaborating biomechanics faculty tested a prototype of this interface with his  $\mu$ -CT scanned beetle datasets (see Datasets and Figure 35). Our approach was to let him try it out first, and then carry out some important tasks he would typically perform to answer some of his research questions. He faced difficulty while trying out the different features of the BHVC. We iterated with him for several weeks, modifying our design based on his advice, and our assessment of his problems. But the usability issues kept recurring.

We reviewed the problems he faced, in discussion with him, and separately as designers of 3D user interfaces. We found that part of his problems was due to his lack of familiarity with 3D input devices, and 3D rate-controlled rotation was not very easy for him. There was also fatigue associated with holding both hand in space all the time. Problems with the Leap like the limited FOV, occlusion (made worse by bimanual tracking), etc., compounded these. The lack of intuitiveness of our gesture set (like open-palm for grab) caused more issues.

A final set of problems arose with the wide array of features in BHVC. This probably increased the cognitive load, while carrying out symmetric or parallel interaction. We found that the researcher wanted to make very precise cracks to isolate the regions of interest (ROIs) in the volume dataset. From our observations, we found that the nuances of the symmetric style of interaction, and the many features of the BHVC interfered quite a bit with his attention, while he wanted to focus on the dataset, and its internal structures.

Investigating factors affecting symmetric bimanual interaction, Balakrishnan and Hinckley reported that increasing task difficulty, divided attention, and lack of visual integration, all cause the brain to revert to sequential style of interaction, rather than parallel (Balakrishnan and



Figure 36: A 3D mouse - the Space Navigator pro

Hinckley 2000). For ecological validity, we used a regular desktop, which caused mismatch between the visual and the interaction workspaces. The analysis tasks with the beetle were fairly difficult, and needed good attention level even from the biomechanics faculty.

We were inspired to explore the options for an asymmetric interface for BHVC, where we can rest one hand on the desk for an absolute 3DOF rotation (instead of free-hand rate-controlled rotation), and use a higher reliability with the Leap through single-hand tracking. This would also allow less fatigue because such an interface doesn't require holding both hands in space all the time.

#### 7.1.5 Bimanual-asymmetric interface design for BHVC

In designing an asymmetric interface, our first priority was trying to decide how to divide the different interaction requirements between the dominant and the non-dominant hands. We decided to follow Guiard's framework for division of labor in bimanual asymmetric interactions (Guiard 1987). He postulated that the dominant hand follows the reference frame provided by the non-dominant hand, the dominant hand works on a finer spatial and temporal scale, and the non-dominant hand leads while the dominant hand follows.

Of the interaction features in the BHVC, we observed that the camera rotations and translations applied to the whole dataset at all times. These work on a global, spatial and temporal scale, and provide the frame of reference for the more nuanced interactions specific to the volume cracker. Similarly, the mode switches (or the gesture commands) for system control also operate globally.

We thus decided to separate out the global 6DOF manipulations, and system control commands to a different input modality, and use that with the non-dominant hand. We chose to use the Space Navigator Pro, from 3Dconnexion (3D-Connexion 2014) for this purpose, which is a 3D mouse (see Figure 36). It provided a 6DOF joystick control, which we could directly use for camera manipulation. We used the buttons on its keypad for our system control commands.



We used the Leap for the functionalities specific to the VC. Now the Leap tracked just the dominant hand. We designed the asymmetric interface of the BHVC inspired by a magnet-metaphor. We imagined the dominant hand acting like a magnet, and the voxels, in a volume, acting like small metallic pieces. As the hand hovered around the dataset in 3D, it attracted more or less of the voxels, depending on how close or far it was from the dataset. To keep the interface simple, we defined the boundary of magnetic attraction as a plane, positioned at fixed distance from the hand, towards the volume, and oriented perpendicular to the line joining the center of the hand and the center of the volume. We showed a transparent slice plane at that position and orientation, as a visual cue to the user.

The biomechanics faculty found that this interface was a lot more intuitive and easier to control than the previous symmetric version. But there were still some design issues specific to BHVC. For example, the cracking preview was unconstrained in 3D. When the cracking preview showed the volume breaking into two along the viewing axis (the line of sight of the viewer), it blocked the cracking face of the other sub-volume, by the sub-volume in front. This was making it difficult for him to precisely judge where to crack. So we made the cracking plane parallel to the view by keeping the depth of the hand at the same depth as the center of the volume (see cracking preview in Figure 33).

Another problem was related to selecting a sub-volume in 3D. Our interface lacked all depth cues such as stereo, shadows, etc., except for perspective rendering. He was struggling with selecting sub-volumes in 3D, as it was difficult to gauge the depth of a hand compared to that of the volume. This was increasing the difficulty level of the overall BHVC experience. So we decided to use *point-to-select*, using the direction of the ray from the camera and passing through the position of the hand, instead of the earlier absolute position mapped selection.

Finally, to improve the precision of the crack, and to give him more control over isolation of the regions of interest, we introduced an important feature in the BHVC. We called this the *turn-the-pages* feature (see Figure 37), inspired by a book metaphor. After we crack a volume, now we could point to either of the sub-volumes, and activate the *turn-the-pages*. Like we turn pages in a book, now we can turn the layers of voxels from one sub-volume to the other. With the

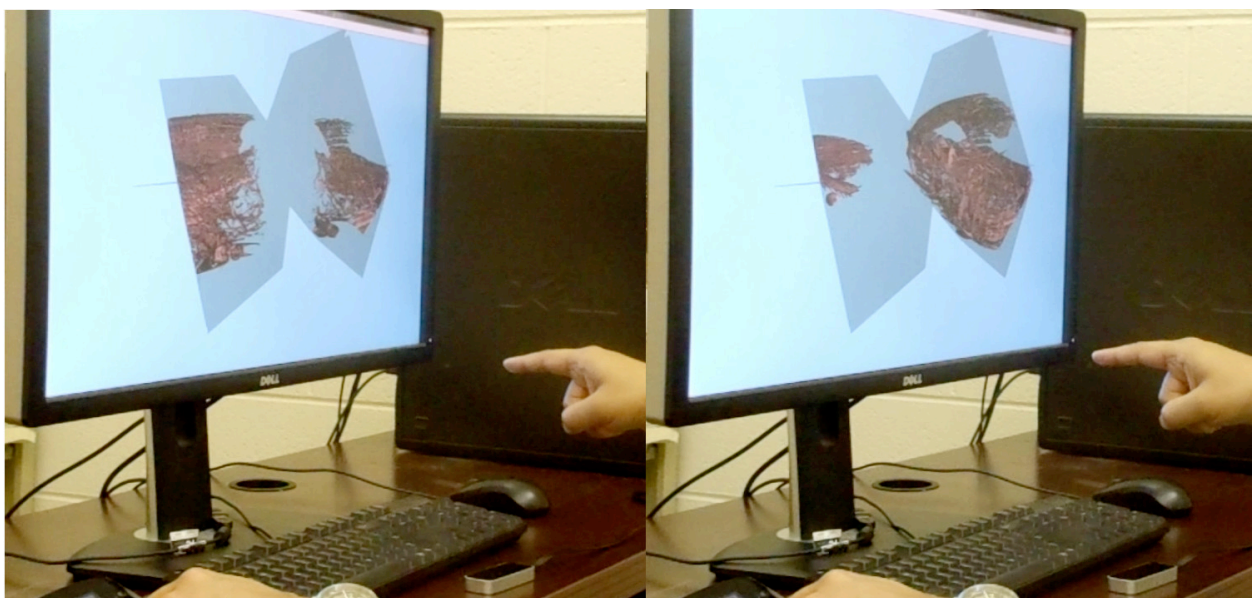


Figure 37: *Turning-the-pages* with BHVC - the synchronous shrinking and expanding of the sub-volumes with hand movement.

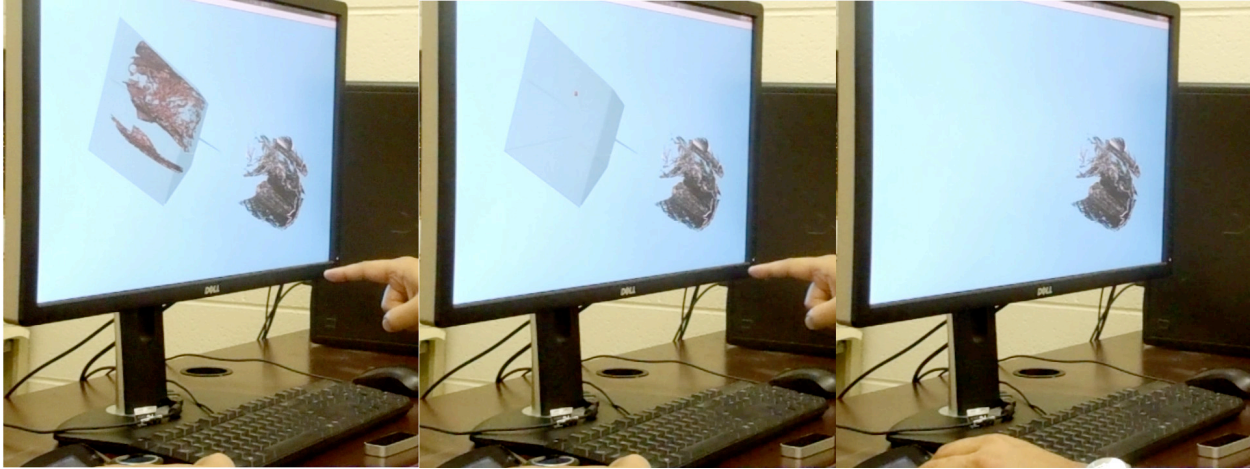


Figure 38: Hiding a sub-volume - point-to-select a sub-volume, hit the escape key to hide, and remove the tracked hand.

movement of our hand, the two sub-volumes synchronously shrunk and expanded (position-mapped to the hand above the Leap, along one axis). This feature combined the power of precise re-adjustment of the crack and the ease of intuitive position-based control.

Based on a few more design studios with the researcher, we introduced a few more features to make the BHVC interface more usable and comprehensive for his tasks:

1. The previous unconstrained cracking process gave way to an automated simulation of cracking of the volume to the sub-volumes, each oriented at 45 degrees to their original orientation before cracking (see cracked beetle in Figure 33). This arrangement opened them up like we open a book to see the pages inside, which was precisely the original purpose of the cracking metaphor. This avoided the unnecessary but required bimanual interaction from the user right after cracking to arrange the individual sub-volumes at proper position and orientation, best for viewing the cracked faces. To preserve conformity, we introduced a similar automated joining back simulation.
2. Another important change was a follow-up to the previous change. As we had automated the cracking process, there was no real need of individual sub-volume manipulation. So we removed the grab action for constrained movement of individual sub-volumes.
3. We introduced a feature to hide selected sub-volumes (see Figure 38). Once a volume is broken into multiple sub-volumes, and the user wants to focus on just one of them, the other sub-volumes might clutter or occlude the view unnecessarily. This feature allowed the user to hide (or unhide) sub-volumes at will. We showed an outline for a hidden volume, to avoid visually losing its location.
4. Combining the *turn-the-pages* feature with the hidden volumes gave a new experience (see Figure 39). Once a sub-volume is hidden, the turn-the-pages would show the growing or shrinking of only the other sub-volume. This helped focusing on getting the regions of interest more accurately, as now there was less visual clutter.
5. Instead of having a fixed center of rotation at the origin of the virtual world, or at the mid-point of all the visible sub-volumes, we decided to have the center of rotation at the center of the screen, but at an average depth of all the visible sub-volumes. This kept rotating sub-volumes within the visible screen space.

With all these modifications, the interface appeared to be sufficiently ready for a formal evaluation study by experts.

## 7.2 Evaluation

### 7.2.1 Study Design

We designed a study to evaluate the performance of our asymmetric Bare-Hand Volume Cracker (BHVC) technique against standard interaction techniques used by researchers in volume data analysis. The first prototype of the volume cracker (Laha and Bowman 2013b) was evaluated against the most widely used 2D interaction technique (axis-aligned slicing or AAS) and a widely known 3D interaction technique (arbitrary slicing or AS). We kept our evaluation design similar as we thought that it would allow us to know how our design performs as compared to baseline 2D and 3D interaction techniques.

We wanted to have expert participants in our study. It is very difficult to recruit very many experts in the domain of biomechanics. A within-subjects design would have given us more statistical power. But we wanted to keep enough variety in the tasks (see Tasks) to have meaningful results for our collaborating biomechanics faculty. We thus settled with a between-subjects design with three groups for each of the three techniques BHVC, AS and AAS. With the available number of experts (see Participants), we were able to secure four sets of data in each study group.

### 7.2.2 Goals and Hypotheses

Our main goal is to find out how BHVC performs quantitatively compared with the AS and AAS techniques in terms of task performance. So our primary research question is:

1. Which one of BHVC, AS and AAS is the overall best in terms of task performance for analyzing raw volume datasets?

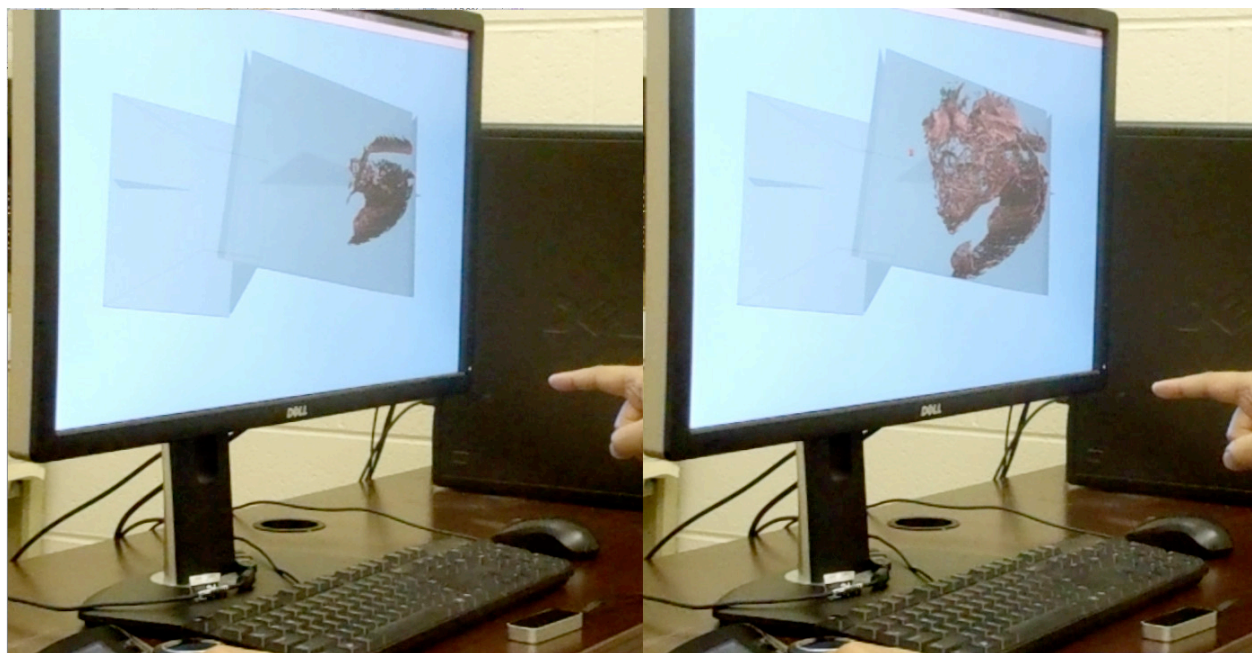


Figure 39: Turning-the-pages in BHVC with one of the sub-volumes hidden.



In our research, we are also gathering visual analysis tasks from multiple scientific domains in an abstract task taxonomy. Our next goal is to understand the mapping between those abstract task categories and the interaction techniques such as the BHVC, AS and AAS. Thus our next research question is:

2. Is the task performance with each technique consistent within an abstract task category?

We also wanted to capture some qualitative feedback on the usability of our interfaces for the three techniques that we prototyped, from an expert perspective, giving us the next research question:

3. How does BHVC compare with the standard interaction techniques in terms of user preference, usability and ease of learning?

Based on our design principles, and prior results (Laha and Bowman 2013b), we had the following hypotheses for our research questions.

H1.a. *BHVC will be more accurate than both AS and AAS* - BHVC is designed to isolate and analyze smaller regions in a whole volume. This should allow high accuracy in search tasks inside a volume. Spatial judgment tasks, involving close analysis of some internal structure or region of interest within a volume from different viewpoints, should also benefit from it. AS gives quick insights into occluded sub-volumes, but would probably be less accurate due to the associated loss of context. AAS uses axis-aligned approach while most internal structures in a volume dataset in general are not be axis-aligned. So we expect it to be far less accurate than VC and AS techniques.

H1.b. *BHVC would be slower than the other two techniques* - The number of functionalities in BHVC is more than in AS and AAS. This might cause the BHVC participants to be slower than others, if they want to try out the different features in general. Using the *joining back* functionality more frequently than the *turn-the-pages* for getting precise cracks could also slow down performance.

H2. *Task performance will be roughly consistent within each task category* - The purpose of the abstract categorization is to group similar tasks. Table 30 shows the grouping of the tasks from this study. The scope of this research question is well beyond this study. But we hoped to get some useful insights to inform our future research.

H3. *BHVC will be liked more than the other techniques, but will have comparable usability ratings and ease of learning* - BHVC, based on the real world metaphor of cracking, will be enjoyable. Again, reliance on a real-world action means the learning curve won't be steep. We iterated quite a bit on the design of BHVC and the other techniques, focusing on making all of them usable.

### 7.2.3 Hardware and Software

We used a custom configured machine with Intel's Xeon CPU X5667 (64-bit) running at 3.07 GHz, with 48 GB RAM, an NVidia Quadro FX 5800 graphics card, and a Dell's 27 inch monitor. We also used the Leap Motion controller or the Leap (Leap-Motion 2012) (Figure 34) and the Space Navigator Pro or the 3D mouse (3D-Connexion 2014) (Figure 36), for 3D input.

We used NGRAIN's cSDK (NGRAIN-constructor-SDK 2014) to implement prototypes of BHVC, AS and AAS. We integrated the APIs for the Leap and the 3D mouse with the cSDK,



which pooled the inputs from the different hardware into a common application. We used the joystick and a few buttons on the 3D mouse pad. We used the Leap to track just one hand, in all cases.

#### 7.2.4 Implementation

The joystick in the 3D mouse pad allowed 6DOF movements of the dataset, in the top-left quadrant of the AAS (see c and d in Figure 40), and in the AS (see a and b in Figure 40) and the BHVC interfaces (see Figure 33). The backward or forward tilt of the joystick pitched the data up or down; the left or right twist caused the data to yaw left or right; the left or right tilt rolled the data left or right. The movement of the whole joystick along any direction translated the dataset in that direction. We kept some thresholds on the joystick to start the rotations, to separate those from the translations. But there was a very small window within which the joystick allowed simultaneous rotations and translations.

In the AAS interface (c and d in Figure 40) we divided the screen into four quadrants. The top-left quadrant showed the dataset from any viewpoint around it. The other three quadrants had fixed views oriented along three orthogonal axes. Using the Leap, the user could select any one of the three quadrants out of the top-right and the bottom two, depending on the position of the tracked hand above, and to the right or left of the Leap. The user could activate the slice mode, for the direction of the selected quadrant, by pressing the *shift* key on the 3D mouse pad, and deactivate slicing by pressing the *escape* key. In the slicing mode, we position-mapped the slice plane to the forward-backward movement of the hand, with the position of the hand right above the Leap always corresponding to the slice plane position halfway through the dataset. Pressing the *C*, *S*, and *A* keys on the keyboard would flip the direction of view (and the slice plane orientation) in each of the three quadrants. We chose the letters *C*, *S* and *A* to denote correspondence with the coronal, sagittal and axial views, used in some versions of this software in domains such as medical biology.

In the AS interface (a and b in Figure 40), we showed a transparent slice plane at the position and orientation of the hand, tracked by the Leap. The slice plane removed the volume behind the plane. The user could freeze the slice plane at any instant, to hold the selected sliced view, by pressing the *shift* key on the 3D mouse pad. While frozen, the 6DOF movements from the joystick would also move the frozen slice plane along with the volume. The user could unfreeze the slice plane with another shift key press.

We implemented the BHVC interface as described in the previous section on bimanual-asymmetric interface design.

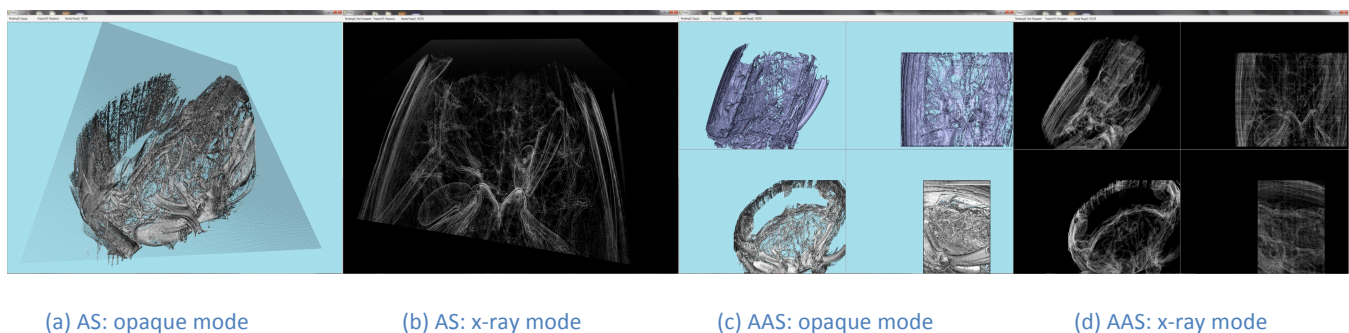


Figure 40: The arbitrary slicing (AS) and the axis-aligned slicing (AAS) interfaces.

For all interfaces, pressing the *X* key on the keyboard activated a transparent X-ray mode; pressing the *X* key again would get back the opaque rendering mode. Pressing the *R* key on the keyboard would reset the dataset, as it was at the beginning. In all interfaces, we showed a small cube as a visual feedback for the 3D hand position, and a similar cube to denote the center of rotation.

### 7.2.5 Datasets

In microscopic computed tomography ( $\mu$ -CT) x-rays are used to scan objects at micrometer-scale resolution. We used the  $\mu$ -CT scans of a ground beetle from the family Carabidae and genus *Platynus*. These beetles are of considerable scientific interest due to their dynamic tracheal behaviors (Westneat *et al.* 2003; Socha *et al.* 2008). Our dataset consisted of a stack of raw 2D images, as obtained directly from CT, MRI, PET and similar volume scanners. We used the software ImageJ (ImageJ 2014) to enhance the contrast of the images, and crop them a little, before loading them in our software.

The size of the whole dataset was larger than our software could render at an interactive frame rate. So we broke down the whole beetle dataset in five one-fifths, from its head down to its tail. We labeled the parts Part A through Part E. For every task in our study, we used only one of these five datasets (see main study tasks in Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study).

We selected a very similar sized dataset for training the participants, which was from another beetle, of the genus *Pterostichus*. The training was focused on familiarizing the participants with the interaction techniques (see Procedure). So the comparability of the data was not important. But we kept the size of the data comparable to the ones in the main study, to train them at similar frame rates of rendering.

### 7.2.6 Tasks

We closely worked with our collaborating biomechanics faculty to identify 12 tasks with the beetle dataset, which are important to his research. After trying these out ourselves with the interfaces, we rejected two tasks, as they were too difficult to fit within a one-hour study frame. We further identified five tasks for training the participants in our interfaces. The final list of tasks is in Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study.

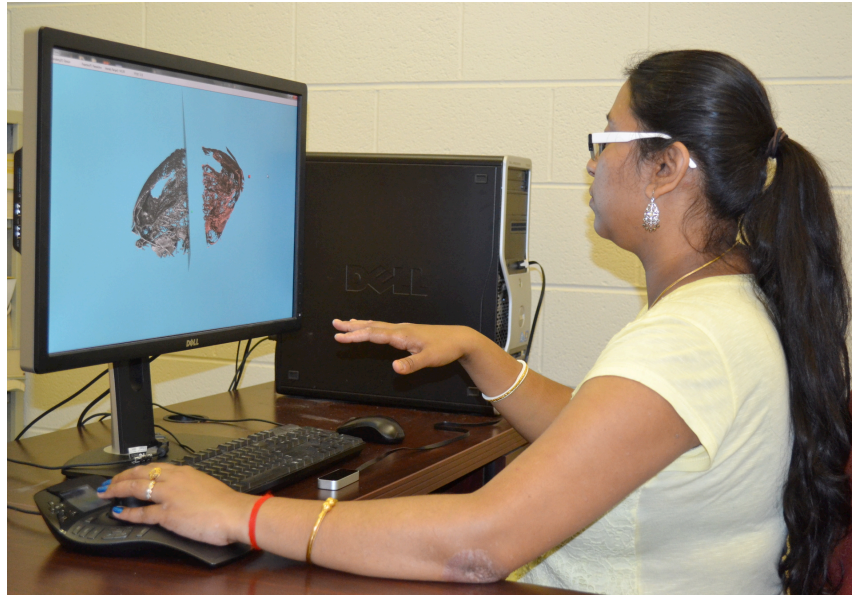


Figure 41: A participant using the asymmetric BHVC interface

### 7.2.7 Participants

We recruited 14 participants for our study. Two of them were pilot participants and had no background in biomechanics. The remaining 12 participants were our main study participants. They were either undergraduate or graduate students, with strong background in analyzing the kind of beetles we had in our study. They were 20 to 31 years old with an average age of 24.8 years. We distributed them randomly between the three groups in our study, giving us four expert participants per group.

### 7.2.8 Procedure

The Institutional Review Board of our university approved the study. Upon arrival, the participant signed an informed consent form informing them of their rights to leave at any point during the experiment. They filled out a background questionnaire that captured their demographic details. The experimenter gave a brief summary of the experimental protocol to the participant. The experimenter then introduced the participant to the hardware and software used in the study, which included a detailed explanation of the different functionalities of the 3D mouse and the Leap, as relevant to the study.

The experimenter then gave a demo of any one of the three interaction techniques, based on the participant's study group. This included system control illustrations with the 3D mouse, the Leap, and the keyboard.

The 3D mouse demonstrations showed how tilt and push on the joystick map to the rotations and translations of the dataset. We also included *unintended mistakes* where translation and rotation get triggered simultaneously.

The demo of the Leap included a brief look at the Leap motion visualizer interface (Leap-Motion 2012). It showed how the hand and finger movements are detected with high accuracy and low latency. The experimenter gave a brief explanation of the FOV of the Leap (like an inverted cone), and showed the approximate boundary of the tracked space using his hands. The

experimenter demonstrated the uses of the Leap specific to the interaction technique of the participant's group (see Implementation).

The experimenter then outlined the best practices with the Leap to avoid the tracking and occlusion problems, as much as possible. The participant was recommended to avoid getting the dominant hand within four to five inches of the leap, and from going outside the boundaries of its FOV. The experimenter also suggested the participant to stretch out the fingers (to avoid occlusion problems and easier tracking), and to sit upright as much as possible.

The participant then performed the training tasks, took an optional break, and completed the main tasks (see Tasks and Figure 41). The main tasks were ordered in increasing order of difficulty (see Table 30), as decided by our collaborating biomechanics faculty.

For each main task, the experimenter recorded the response of the participant, which was evaluated offline using a grading rubric prepared by our collaborating biomechanics faculty. There was a maximum allowable time for each task, which the participant was given before beginning a task. The experimenter recorded the time that the participant took to perform each task, using a stopwatch. But the participant was asked to focus on the accuracy of the response rather than on the time.

After completing the tasks, the participant filled out the post-questionnaire. The experimenter carried out a free-form interview to answer any questions the participant had, before concluding the study. The whole study lasted for about an hour, per participant. The experimenter's demo session lasted around five minutes and the participant's training session lasted about 20 minutes, for each technique.

The hand and body posture recommendations and the mistake demonstrations helped the participants to overcome any awkward moments with poise and confidence. This helped reduce surprises and distractions during the study.

#### **7.2.9 Results**

The grade for each task (calculated between zero and one), and the quantitative metrics from the background questionnaire and the post questionnaire (all on seven-point scales) had ordinal data, while the recorded time was of the continuous type. We employed ordinal logistic regression based on a Chi-square statistic for the ordinal variables, and a one-way analysis of variance (ANOVA) for the time metric. When we found significant difference between the techniques in any of the ordinal data metrics, we further employed the non-parametric Wilcoxon each-pair test for post-hoc analysis. The Wilcoxon test often showed higher p-values than the original regression tests, as it was more conservative than the Ordinal logistic regression model.

To get an overall performance measure for each participant, we calculated a weighted overall grade, which took into account the difficulty level of each task (Table 30) and the grade received for each task by each participant.

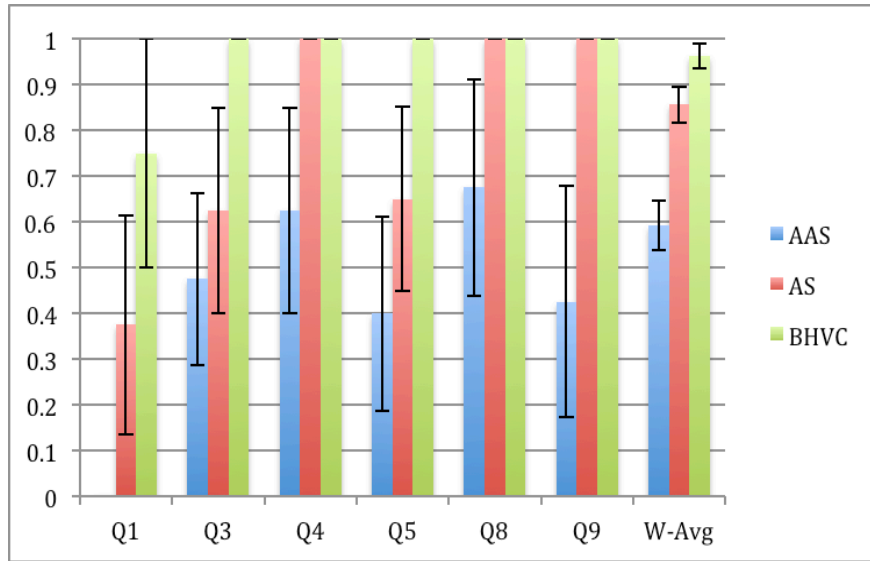


Figure 42: Average grades (with standard error bars) secured by the participants with the three interaction techniques

#### 7.2.9.1 Grade metric: The accuracy of task performance

We observed several significant and close-to-significant effects of interaction technique on the accuracy of tasks performance. We summarize these in Table 29 and Figure 42.

Table 29: Effects of interaction technique on grades.

Task: source	$\chi^2$	DF	p-value	Techniques sorted by Mean scores ("">>>" implies significantly better)
Q1	7.1941	2	0.0274*	BHVC>AS>AAS
Q3	5.9593	2	0.0508	BHVC>AS>AAS
Q4	5.2683	2	0.0718	BHVC=AS>AAS
Q5	6.7914	2	0.0335*	BHVC>AS>AAS
Q8	5.2683	2	0.0718	BHVC=AS>AAS
Q9	8.9973	2	0.0111*	BHVC=AS>AAS
Weighted Average	20.207	2	<.0001*	BHVC>>AS; BHVC>>AAS

Post hoc tests show that BHVC was more accurate than AAS at  $p=0.0603$  for Q1, BHVC was more accurate than AAS at  $p=0.0668$  for Q3, BHVC was more accurate than AAS at  $p=0.0668$  for Q5, BHVC was more accurate than AAS at  $p=0.0668$  for Q9 and AS was more accurate than AAS at  $p=0.0668$  for Q9. We also found that for the weighted total grade, BHVC was

significantly more accurate than AAS ( $p=0.0294^*$ ) and AS was significantly more accurate than AAS ( $p=0.0304^*$ ).

#### *7.2.9.2 Time metric: The speed of task performance*

We didn't find any significant difference between the techniques for the speed of completion of any of the tasks. On an average, the BHVC participants took longer than the others. For Q10, we found that the effect of technique on the speed of task performance was close to significant at  $p=0.0654$ , with BHVC being the slowest technique.

#### *7.2.9.3 Background questionnaire metrics*

We ran pairwise correlation tests (Spearman's  $\rho$ ) between the participants' weighted overall grade and their total time of task performance, to their self-reported ratings on their tiredness, expertise with computers, frequency of computer use (for work and for fun), experience with volume data analysis (in general and from biomechanics), experience with the technique they used, experience with video games or 3D movies at theater or at home, experience with Nintendo Wiimote, Sony Move or Microsoft Kinect, and experience with 3D interaction for volume data analysis. Experience with video games, or 3D movies at theater or at home showed a significant inverse correlation ( $\rho=-0.681$ ;  $p=0.0148$ ) with the total time taken for task performance.

#### *7.2.9.4 Post Questionnaire metrics*

From the post questionnaire responses we found that the overall experience of the participants differed significantly ( $\chi^2_{df=2}=14.6325$ ,  $p=0.0007^*$ ) between the interaction techniques. Post-hoc tests indicate that the participants liked BHVC significantly more than AAS ( $p=0.0202^*$ ), and also liked BHVC more than AS ( $p=0.0603$ ). There was no significant difference between the techniques for perceived usefulness or ease of use or learning. 75% of the participants, in each of the AS and the BHVC techniques, thought that some form of auto-locking feature of the cutting plane, such as along the X, Y or Z-axis will be helpful. Half of the AAS participants reported fatigue in their arms.

### **7.3 Discussion**

Revisiting the research questions, we find that we have strong evidence supporting our hypothesis H1.a. For the overall weighted average scores, BHVC was significantly better than AAS and better (not significantly) than AS. For all the tasks in which we found significant difference in the accuracy of task performance, BHVC participants consistently received the highest scores (see Figure 42). There was no significant difference between the techniques for the speed of task performance. These put together show that BHVC was the best of the three techniques for the tasks in our study.

Overall, BHVC was the slowest (not significantly) of the three techniques. In the task Q10, which was the hardest, the BHVC participants were close to being significantly slower than the others. This happened because the participants took time to precisely crack the dataset multiple times, hide some of the sub-volumes, and isolate the small portion of the volume that had a *spiracle* (see Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study) they wanted to analyze. This supports our hypothesis H1.b.

Analyzing the results, grouped by task type (see Table 30), we find some consistency in the results within task groups, supporting our hypothesis H2. For example, the only naïve search task



in our study (Q1) showed significant difference between the techniques for accuracy of task performance, but none of the counting-based search tasks (Q2, Q6 and Q8) showed any significant difference between techniques. Both the shape description tasks (Q3 and Q4) showed close to significant difference in task performance, with the BHVC participants scoring higher than the rest.

For the spatial judgment tasks (Q5, Q7, Q9 and Q10), the results were mixed. Two of the tasks (Q5 and Q9) showed significant difference between the techniques for accuracy of task performance; while for the other two (Q7 and Q10), there was no significant difference between the techniques. Analyzing closely, we found that Q7 might be different than the other three spatial judgment tasks because Q7 was open-ended search for left-right connections in the dataset, while the other three had specific “anchors” around which the tasks were designed (see Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study). Q10 is the hardest of the lot, and could probably fit in a class of its own. So there might be some consistency in the results of the spatial judgment task type as well, supporting our hypothesis H2.

For answering the third research question, we looked at the post-questionnaire measures. The participants liked BHVC more than AS and significantly more than AAS, but didn’t think that the techniques were different for ease of use, ease of learning and usefulness. This supports hypothesis H3.

The scores of the AS and the BHVC participants were comparable in some of the tasks (Q4, Q8 and Q9), and significantly (Q9), or close-to-significantly (Q4 and Q8) better than the scores of AAS participants. Also, the BHVC participants didn’t score significantly more than the AS participants for any of the tasks, although for the overall weighted score, BHVC was almost significantly better than AS (at  $p=0.0515$ ), but both BHVC and AS were significantly better than AAS overall. Both BHVC and AS being 3D interaction techniques, this strongly show that 3D interaction could significantly improve task performance in visual analysis of volume datasets.

AS could perhaps be seen as embedded in BHVC, with some extra features. In BHVC we do not discard any part of the volume (like we do in AS), we have the ability to crack the volume in any 3D orientation, and we have many more functionalities that are not present in AS. This might explain the improvement in performance observed from AS to BHVC.

We also found that the participants with higher experience in video games or 3D movies were faster in their task completion, which probably suggests that some proficiency in using 3D input devices might be helpful for using our interfaces. From our observation, and the post-questionnaire comments we found that the participants with more experience in video games thought switching the direction of some of axes in the 3D mouse might be useful. We think this resulted from their perceptual mismatch between the egocentric viewing in games and the exocentric viewing in our design.

The recurring suggestion of restricting the movement of the cutting plane from the AS and BHVC participants suggest that they found the freedom of movement interfering with their ability to focus on the real task in hand. Some of the participants in the AS and the AAS also thought that having a translucent or “ghost” version of the sliced part of the dataset might be helpful. We believe this points to the probable loss of important context in certain tasks, due to slicing.

From our observations, we believe the problems with AAS in general were caused by the random alignment of the internal structures in the data and the gaps between them, which was also not axis-aligned. This was also directly reflected in the post-questionnaire comments of one AAS participant. AS allowed the participants to slice along the direction needed to isolate or closely monitor a structure from any desired angle. But the loss in context in AS (as well as in AAS) due to the sliced part of the data interfered with their analysis. BHVC preserved this lost context, while also allowed them to hide the parts when they desired. This might have led to the superior performance of the BHVC.

Table 30: Tasks with abstract categories and difficulty levels

Task#	Abstract Task Category	Difficulty (1-5 scale; 5-Hardest)
Q1	Search-naïve	1
Q2	Search-count	1
Q3	Shape description	2
Q4	Shape description	3
Q5	Spatial judgment-connection	3
Q6	Search-count	3
Q7	Spatial judgment-connection	4
Q8	Search-count	4
Q9	Spatial judgment-connection	4
Q10	Spatial judgment-connection	5

## 7.4 Conclusion and Future Work

We demonstrated the design of a bare-hand version of the Volume Cracker technique (Laha and Bowman 2013b), and provide evidence that the interface has significant benefits over baseline 2D and 3D interaction techniques for certain categories of generic visual analysis tasks with volume visualization. Our design contributions include the final BHVC technique (bimanual-asymmetric), knowledge acquired through the iterations (including the bimanual-symmetric version of BHVC), and guidelines for designing bare-hand interactions. Our evaluation study with experts and real-world datasets provide evidence for the benefits of 3D interaction techniques over standard 2D interaction techniques, and in particular, for improvements in task performance with the BHVC. We finally contribute to a broader goal of mapping 3D interaction techniques to generic task categories for improving performance in volume data analysis. The next chapter describes our work in creating a generic visual analysis task taxonomy for analysis of volume visualizations.



## Chapter 8. Task Taxonomy

### 8.1 Introduction

Scientific and medical data generated through modalities such as magnetic resonance imaging (MRI), computed tomography (CT), confocal microscopy, and positron emission tomography (PET) are in 3D volumetric format. Visualization of volume data is an invaluable resource in various domains such as medical imaging, geophysics, materials science, and astrophysics (Kaufman 1996). Scientists and researchers traditionally analyze such data in non-immersive desktop systems, with mouse and keyboard interfaces. Typical analysis requires understanding structures inside the 3D volume, which are occluded from view by surrounding material (e.g., isolating a tumor inside the human brain). Spatial understanding in 3D also requires assimilation of views of the same space (internal to the volume) from multiple angles.

Visualization researchers have come up with innovative approaches like Focus+Context techniques for mitigating such problems. GPU researchers are significantly improving the rendering speed of volume data. But volume visualization still remains an active area of research, which underlines the fact that the challenges in the field are deep-rooted.

In addressing these challenges, researchers in the visualization, virtual reality (VR), and 3D user interface (3DUI) communities run empirical studies with volume datasets. They typically use volume data and tasks from one or two scientific domains. The findings from many of these studies are significant and deep. But the results lack portability to other scientific domains, primarily due to the lack of a mapping between the visual analysis tasks performed in various scientific domains. This is noted as the first and foremost scientific visualization research challenge by Keefe and Isenberg (Keefe and Isenberg 2013), and also identified as a major limitation of our own prior studies (Laha and Bowman 2012).

It is also important to note that a major challenge in running empirical studies with domain scientists is the limited accessibility to well trained study participants. The reasons range from very limited availability of expert domain scientists, to possible confounds in measuring task performance due to differences in prior level of knowledge of the study participants. A viable solution could be mapping the tasks performed by domain scientists to more abstract tasks not requiring expert knowledge. This will preserve the nuances of the tasks themselves (as important to the domain scientists), but make them fully understandable to non-experts (Laha and Bowman 2012). Such a categorization of tasks would allow visualization researchers to tap the pool of novice participants with comparable (or absolute zero) level of domain knowledge, and the results will be applicable to specific scientific domains as well.

Finally, traditional empirical studies in visualization, VR and 3D interaction (3DI) have been studying task performance with basic interaction tasks such as rotation, translation, and scaling. But such results do not help the VR system designers seeking more user-centric approach for system design (which will need closer ties to domain tasks), nor do they help the domain scientists trying to relate to such findings. This is becoming very important with the increasing interdisciplinary nature of our research.

A generic framework describing visual analysis tasks with volume data therefore would address three goals: portability of empirical results across domains, understandability of abstract

scientific tasks by novices, and meaningfulness of results to domain scientists. The framework we describe in this chapter has a basic task categorization of visual analysis tasks, a set of supporting tasks, and a list of dataset characteristics that we believe can influence the performance measures on the basic tasks.

Our design differs significantly from prior work in task taxonomy, which has mostly focused on tasks related to information visualization (Shneiderman 1996; Zhou and Feiner 1998). Morse et al. proposed a domain-independent taxonomy of visual tasks (Morse *et al.* 2000), for testing of visualization in isolation from the rest of the system. Amar and Stasko presented a task-based framework or task taxonomy (Amar and Stasko 2004) for bridging analytic gaps in information visualization. Tory and Moller presented a high-level visualization task taxonomy (Tory and Moller 2004), for categorization of visualization algorithms, rather than data. Valiati et al. proposed a visualization task taxonomy (Valiati *et al.* 2006), based on previous designs, and observation of tasks performed by users with two visualization techniques in a multidimensional dataset. In spite of all this work on task categorization, the volume visualization community does not yet have a comprehensive taxonomy of tasks focusing purely on 3D spatial data analysis (Keefe and Isenberg 2013).

In the following sections, we describe our methodology: coming up with the primary set of tasks through interviewing various scientists and researchers from multiple domains, and then reaching out to a broader population through a survey to evaluate our initial design. We summarize the results from this survey, which supported the design of our task taxonomy, and gave insights to make it more comprehensive.

## 8.2 Designing the Visual Analysis Task Taxonomy

The idea of an abstract visual analysis task taxonomy came from our free-form interview sessions with domain scientists from medical biology, geophysics, materials science, biomechanics, and paleontology. We also interviewed a professional imaging analyst, who works on a daily basis with a variety of MRI and CT scans of humans and animals. We held these interviews to come up with tasks for our empirical research on the effects of VR system fidelity on analysis of volume datasets. During these sessions, the researchers gave examples of a variety of visual analysis tasks they typically perform with the volume datasets that they use or encounter in their daily research.

Although the kind of tasks each of them described were specific to their own domain, we observed an interesting trend in their task descriptions—many of the tasks were comparable between the domains at a fundamental level. For example, the geophysicist follows cracks through the plates in the Earth’s mantle, the biomechanics scientist tracks tracheal tubes inside a beetle, the medical biologist tracks blood vessels inside the limb of a mouse, and the paleontologist tracks the jagged boundary between the cells inside a microscopic fossil. The medical biologist searches for floating soft tissues inside a dense mass of structures, while the paleontologist searches for intracellular bodies (such as nuclei) inside the cells. The biomechanics researcher is interested in the morphology of the tracheal system inside a beetle, while the paleontologist wants to know the overall structure of a fossil.

Also, as we ran multiple VR experiments producing empirical findings in disparate scientific domains (see Chapters 3, 4 and 5), we did not want to restrict any significant result just to the domain in which one experiment was run. Associating the results with the abstract categories

allowed us to have results that could be linked and generalized across multiple domains. Thus, we identified a few abstract task categories that we could use to map the tasks from these multiple VR experiments. These included *search*, *spatial understanding*, *path following*, *quantitative estimation*, *shape description* and *pattern recognition*.

We discussed these task groups with our colleagues at the workshops, and in general, at the IEEE VR and IEEE Vis conferences. This led us to include a set of supporting factors in the design (see 8.2.1.2 below), which made the taxonomy more comprehensive. We also found that our definitions of *path following* and *spatial understanding* categories had close connection. After further discussions with domain scientists, we observed that *path following* would refer to an *action* we would take to carry out a *search* or a *spatial understanding* task, but it wouldn't constitute a category all by itself; it thus got subsumed into the other categories, and our proposed task taxonomy then had five categories.

### 8.2.1 Proposed task taxonomy

#### 8.2.1.1 Basic Task categorization

Here we provide the basic grouping of the abstract task categories. We also provide our definition of each group, along with an example in each sub-group.

1. Search: We define a *search task* as one in which the user is searching for a feature in the dataset.
  - a. Searching for presence/non-presence of a particular feature/object in a dataset. e.g., “Search for distinct bone segments in this sample.”
  - b. Counting objects/features, given they are present. e.g., “Count the total number of tracheal tubes and categorize them by diameter.”
2. Pattern Recognition: We define a pattern as a characteristic that is repeated or a trend that we see as we look through the dataset.
  - a. Looking for a trend. e.g., “Blood vessels on one side of the dataset are skinny and they get progressively bigger as we go through the dataset. Identify on which side of the dataset we notice this trend.”
  - b. Does a pattern repeat itself in different parts? How many times? e.g., “Looking through the sample longitudinally, count the number of fibers curving through the sample, and the number of straight fibers.”
3. Spatial Understanding: These are tasks where the user makes a judgment in a 3D spatial context about the position and/or orientation of a feature on an absolute or relative sense. e.g., “Which sets of tubes are located closest to the front of the dataset?”
  - a. Absolute - Locating a feature/object which is the highest, lowest, farthest, or closest to/from a viewpoint, or in the overall dataset. e.g., “Which sets of tubes are located closest to the front of the dataset?”
  - b. Relative - Judging whether a feature/object ‘A’ is higher/lower/farther/closer than another feature/object ‘B’. e.g., “Which of the twisting fibers is the closest to the center of the bundle?”

- c. Judging whether two features/objects intersect or if they are distinct. e.g., “Does the left side of the tracheal system connect to the right side, throughout the body?”
- 4. Quantitative Estimation: These include tasks in which the user has to estimate some property of the dataset, or part of it, quantitatively. Examples of things we can estimate in a volume dataset are density, size, volume, distance, length, surface-area, curvature, angle, velocity and other super-scalar quantities, etc.
  - a. Absolute estimation. e.g., “What is the density of the highlighted region of the bone marrow?”
  - b. Relative estimation (binary). e.g., “Study the thickness of the borders and the joints between the cells. Are different joints and borders of comparable or variable thickness?”
  - c. Relative estimation (quantitative). e.g., “How many times denser is the boundary of the cells at the center than the outer boundary of the cluster?”
- 5. Shape Description: These include tasks where the user describes the shape of either the whole or some part of the dataset qualitatively. e.g., “What is the cross-sectional shape of each tracheal tube? Shapes tend to run from circular to flattened ellipses.”

#### 8.2.1.2 Other factors influencing visual analysis performance

Volume visualization researchers and colleagues we met at professional conferences observed that although the basic categorization might cover the tasks from different domains, the type of volume data generated could vary in different ways between domains. Also, there are various algorithms to visualize volume datasets, and different ways to explore the rendering. Based on their suggestion, we included the following dimensions or factors in our framework. These factors not only make this framework more comprehensive, but also bring in varied but important perspectives to look at empirical results tied to the above abstract task categories.

##### Based on viewing style

We can explore volume datasets in a virtual world from a first-person view (egocentric) or from a third-person view (exocentric).

- 1. Egocentric
- 2. Exocentric

##### Based on volume rendering algorithms

There are many different ways to visualize or render a volume. This list is a high-level categorization of the different volume rendering algorithms, adopted from (Marmitt *et al.* 2006):

- 1. Decomposition
- 2. Isosurface rendering
- 3. Maximum intensity projection (MIP)
- 4. Semi-transparency
- 5. X-ray rendering

Based on the temporal nature of the data

Volume data can vary based on how it is obtained. If we scan a static object in 3D, like CT scan of a human body part, we get a non-time varying data. If we scan a volume over some time, e.g. a fluid flow, or astronomical changes over a period of time, we get time-varying data.

1. Time varying
2. Static/non-time varying

Based on dimensionality of the volume data

Volume data maps material properties of the object that it is representing, such as color, density, refractive index, etc. If a dataset maps just one material property we call it single-dimensional; if it maps more than one material property we call it multi-dimensional.

1. Single-dimensional
2. Multi-dimensional

### 8.3 Evaluating the design of the taxonomy

To verify if our proposed taxonomy really made sense to the broader scientific community, particularly to researchers using volume data outside the set of people we interviewed. Mainly, we were interested to know the following:

1. Do people agree with this way of categorizing tasks?
2. Do people suggest other categories that we do not have?
3. How often do people perform tasks in each of our categories?
4. What real-world tasks align to the categories?
  - a. How similar are the tasks in each category?
  - b. Will the different tasks in a category have similar results in a study?
  - c. Do we need to further sub-divide?

To explore the above questions we designed and ran a survey targeting people from academia and industry working with volume data analysis. These included scientists and researchers from various universities, as well as imaging analysts and similar professionals from the industry. We hosted our survey on Qualtrics.com. The survey was reviewed and approved by the Institutional Review Board at our University. The estimated time to complete the survey was around 20 minutes. There was no guaranteed award for taking the survey, but we sent \$20 Amazon gift cards to ten randomly chosen survey takers, after the survey was completed. We reached out to a wide variety of people, working regularly with volume datasets, from across the globe through various professional email listservs as well as through our direct and indirect professional contacts.

#### 8.3.1 Overview of Survey

The full survey is available in Appendix F. Experiment Documents: Task Taxonomy Survey. It broadly consisted of three sub-sections, as described below. We provided a pre-determined initial framework (as opposed to a fully open-ended survey) to collect the responses in a structured fashion. The structure of the survey was finalized through inputs gathered from various

workshops and conferences over multiple years. But we also kept the window open for additional feedback of the respondents for missing tasks and/or categories in the post-questionnaire part of our survey. Finally, we think that our approach is devoid of bias towards task categorization in a single domain as the basic framework was designed based on inputs from scientists of multiple domains.

#### **8.3.1.1 Background Information**

The first part focused on the demographics of the survey-taker. The questions in this part captured details on their domain expertise, and various details about their experience with analyzing volume data.

#### **8.3.1.2 Task Categorization**

The next part of the survey had specific questions tied to each of the basic task categories. For each task category, we first provided a definition of the category, and its sub-categories (if there were any). We also provided one example task from three different domains (medical biology, paleontology and biomechanics) in each category, prepared by our collaborating scientists in those domains. We then captured their level of agreement to each category. Finally, we requested at least one task from the respondent's domain, in each category, or an example task that they felt was closest to those in that category (when they had a low level of agreement to that category).

#### **8.3.1.3 Post Questionnaire**

The final part of the survey had questions related to the respondents' daily division of time between the task categories, tasks that they thought were missing (with abstract groups that they would want to put the tasks in), and their general level of agreement with our design.

### **8.3.2 Results and Discussion**

The survey was active for around five months. During this period, 167 respondents took the survey. They ranged from 23 to 77 years of age, with a median age of 40. There were 112 males and 31 females; 24 declined to mention their gender. 107 self-reported to be scientist, medical-practitioner or radiologist working or doing research with volume datasets, 19 self-reported as professional imaging analyst and 17 self-reported as others.

The respondents came from a variety of scientific domains, including geology/geophysics (3), medical biology (10), paleontology (1), biomechanics (3), cell biology (1), engineering (26), medicine (8), radiology (9), architecture (2), animation and 3D modeling (7), and others (20). Their experience with volume data had different ranges: less than a year (9), 2-5 years (38), 5-10 years (28), and greater than 10 years (33).

Most of them (67 respondents) have worked with more than 20 different volume datasets, but there were respondents who worked with less than 20 datasets: three people have used just one dataset, 15 people have used between two and five datasets, 10 people have used between five and 10 datasets, and 13 people have used between 10 and 20 datasets. Of the data that they use, 24% is time-varying data while 76% is static or non-time-varying, and 70% is single-valued while 30% is multi-valued.



They predominantly analyze segmented data (79), but there were respondents who work with raw volume data as well (13). For segmentation, 30 reported using automatic methods and 16 reported using manual methods, but a majority reported using both automatic and manual methods (59). To explore the data, 24 people reported using egocentric method (navigating or flying through/within the data) while 63 people reported using exocentric method (viewing by spinning the data and looking from various angles from the outside).

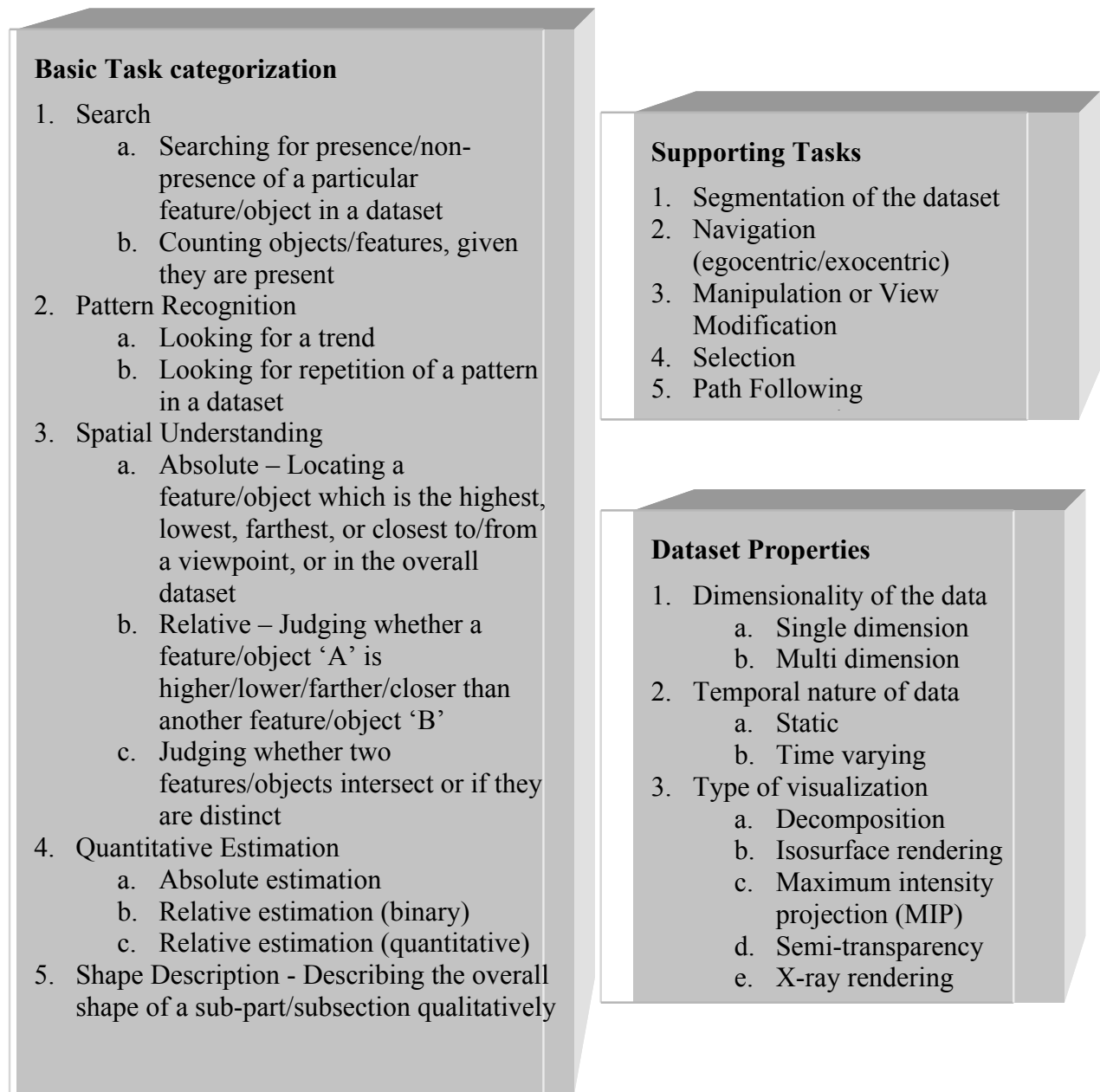


Figure 43: Current design of the task taxonomy framework



### 8.3.2.1 Agreement with classification

We asked the participants to report how much they agreed with our design of the task taxonomy, before the end of the survey. They gave their response on a 1-7 scale, with 1 being the least agreement and 7 the most. The following box plot (Figure 44) shows their agreement, and their responses tied to each task category in our task taxonomy.

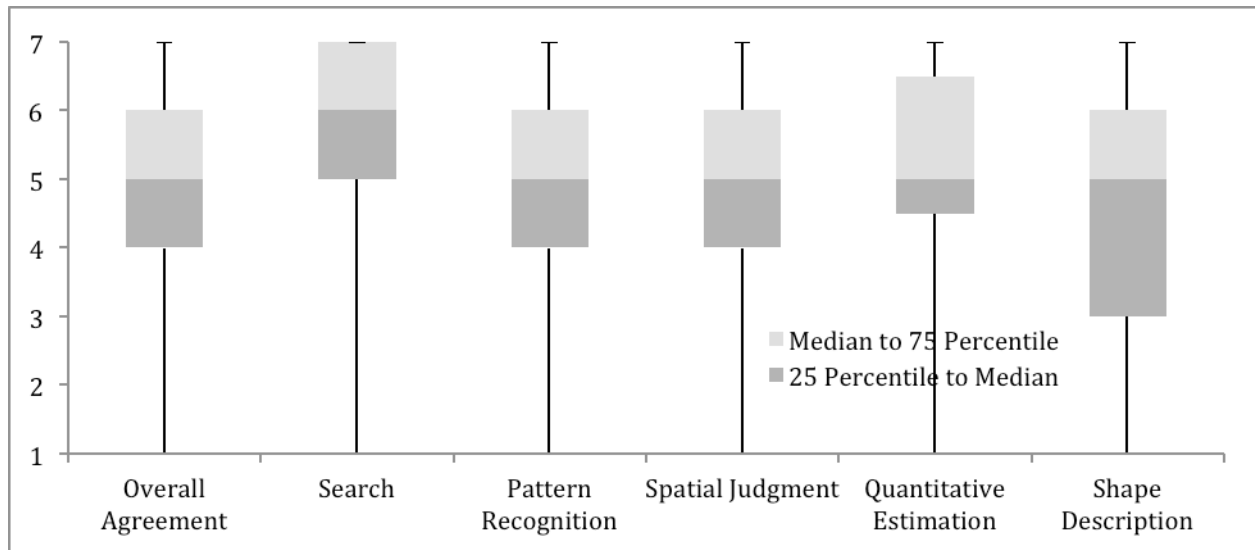


Figure 44: Box plot showing the match between the task description and the survey respondents' expectations

Although each plot above has variance ranging from 1 to 7, the 25 percentile of each one is above the scale average of 4, except for the one with shape description. The box plot on the overall agreement indicates that majority of the respondents agreed in general with our design of the task taxonomy. The other five plots indicate that the majority of the respondents agreed with the definition and the categorization of the individual task types as well, with the weakest agreement shown towards our shape description categorization. Also, at least one respondent had the least and at least one respondent had the highest agreement to our definition of every task category (the responses ranged from 1 through 7 for each task category). This points to the lack of a general consensus, although a majority population seems to agree with the categorization. This makes it hard to design a taxonomy or framework that will be satisfactory to everyone.

#### Time division of the respondents between the categories

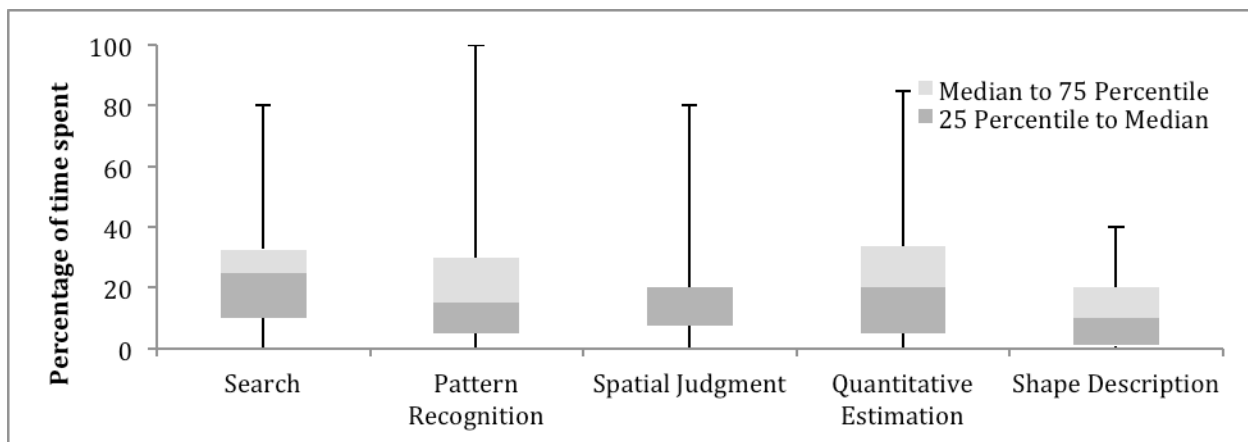


Figure 45: Box plot showing the respondents' division of time between the different tasks from our survey

The box plot in Figure 45 shows the division of time by the respondents in the different task types. There is no significant difference between the categories in the mean time spent by the respondents. However, some of the participants reported spending 80% or more of their time in each of search, pattern recognition, spatial judgment and quantitative estimation tasks.

#### Correlations between different entries in our survey

We ran non-parametric Spearman's  $\rho$  tests between the different measures in our survey, and looked for correlation at the 5% significance level. We found a significant correlation ( $\rho=0.2282$ ;  $p=0.0474$ ) between the domain of the respondents and their use of segmentation, which indicates that the need for segmentation of volume datasets may vary between scientific domains. The domains from which respondents reported using segmentation more than others are medical biology, engineering, medicine and radiology. We observed a significant correlation between the amount of time the respondents spend for shape description tasks, and their involvement with single-valued ( $\rho=0.2632$ ;  $p=0.0440$ ) and multi-valued data ( $\rho=-0.2718$ ;  $p=0.0373$ ). This indicates that the respondents using single-valued data would more likely be in need of describing shapes in volume data. We didn't find any significant correlation between their division of time between our task types and the domains they came from, except for a close to significant correlation for time spend in quantitative estimation tasks to domains ( $\rho=0.2456$ ;  $p=0.0925$ ). This indicates that quantitative estimation tasks might be more common in some domains than others, such as engineering, radiology, animation and 3D modeling.

#### Task examples given by the respondents

For each task category, we requested the respondents to give examples from their specific domains. The respondents generously provided close to 500 individual tasks from numerous domains of volume data, mentioning all of which would be outside the scope of this paper. But the respondents' mapping of this huge variety of tasks to our task categories supports our basic hypothesis that a few abstract task categories could sufficiently and comprehensively capture a vast majority of tasks performed by various scientists across multiple disparate domains.



Figure 46: Exploring the fibers in a micro-scaffold volume dataset.



Figure 47: Analyzing the tracheal tubes in volume visualized micro-CT scan of a *Platynus* beetle.

Our detailed analysis of each individual task and the respondents' mapping to our taxonomy indicate that different tasks in a category would be broadly similar. For example, describing shapes of voids between tightly packed fibers (from medical biology, Figure 46) by an imaging analyst and looking at 3D shape variation among fossils and extinct specimens by anthropologists are grouped in *shape description* tasks; quantitative characteristics of the distribution of cerebral fluid (medical biology) and estimating how big one mantle plume is relative to another (geophysics) are grouped in *quantitative estimation* tasks; whether a structure is connected to another structure in airline baggage search, evaluating distance between lesions (in stereotactic radiosurgery or SRS of brain) and determining the tracheal tubes connecting the left and right sides in a beetle (from biomechanics, Figure 47) are grouped in spatial judgment tasks. Further subdivision of the categories might help to fine tune the tasks in each category, but it doesn't appear to be necessary.

The respondents didn't suggest any new category, which probably points to the comprehensiveness of our current design. But after detailed analysis of their inputs, we found that our design would benefit from another dimension of "supporting tasks". We define a supporting task as one that is used in the process of analyzing volume data, but which is not an

analysis task by itself. Inherently, these are also not abstract tasks that we want to map across multiple scientific domains. For example, we observed that an analyst’s objective might be to compare multiple datasets, not just do the tasks in one dataset. So, we added “data comparison” as a supporting task. The full list of supporting tasks is given in Figure 43.

## 8.4 The Role of the Task Taxonomy in our VR and 3DI Evaluation studies

The generic task categories from our taxonomy helped us to generalize the findings in all our VR empirical experiments and the 3DI evaluation studies. In VR Study-III, we mapped the evaluation tasks directly to our task categories, which helped to analyze our findings based on task types (see section 5.3.2). This analysis identified that stereo alone, or head tracking in a higher field of regard system, aids in spatial judgment and searching.

We also revisited the tasks from VR Study-I, categorizing them based on our taxonomy, and did a meta-analysis of our results based on this task taxonomy (see section 5.3.3). This allowed us to get preliminary insights on how different rendering styles affect volume data analysis tasks. We found that higher field of regard lead to improvement in task performance for analyzing volume datasets with 3D texture rendering, while stereo was useful for analyzing isosurfaces.

We used some of the task categories (search and pattern recognition) from our taxonomy in our first evaluation of the Volume Cracker prototype (see section 6.2.6). This helped to produce results meaningful to domain scientists from this study. In our second evaluation of the bare-hand version of the volume cracker, we again used tasks directly mapped to this task taxonomy (see Table 30). This helped us producing generic results from this evaluation study.

We found some consistency in the significant findings in VR Study-III within our task categories, and a substantial consistency in the results within each task category in the evaluation study of the BHVC technique (see section 7.3). These highlight the strength of our framework.

There is some room for overlap between the spatial understanding and some of the other categories, such as quantitative estimation, pattern recognition, and search. This is because to identify or match a pattern, or to do a quantitative estimation, or while searching, we might need to make a spatial judgment. E.g., to answer whether a quantity is bigger or smaller than another (relative quantitative estimation), we might need to know how close or far the quantities are, as during visual analysis in perspective rendering, the distance from viewpoint affects the size. Also, for matching two patterns, we might have to understand their coiling or arrangement in 3D. In this context, it is important to point out that our taxonomy is aimed at categorizing the tasks based on the mutually exclusive nature of their goals. So although the methods involved in carrying out a task may borrow properties from another task category, we want to group them based on their final objective.

## 8.5 Conclusion

We have outlined a framework centered on an abstract task taxonomy of visual analysis tasks with volume datasets. We also described the results from a survey we ran to evaluate our design. Based on our findings, we believe this design is sufficiently nuanced to capture the tasks from a wide variety of scientific domains.

This work will serve to provide a framework to share and generalize the empirical findings from studies run in visualization, VR and 3DI communities across multiple scientific domains. This

theoretical framework is aimed at providing wider and richer impact of our fundamental findings to the broader scientific community working with volume datasets.

The most obvious use of this taxonomy is in the context of choosing tasks while designing an experiment. For example, in a study to compare multiple visualization techniques, the designer can choose tasks from one or more of the categories in our classification (see Figure 43). This will allow the findings from the study to have broader impacts beyond the domains from where the volume dataset(s) originated.

Researchers can also use this taxonomy to revisit and categorize the tasks from prior empirical studies. Mapping the findings from previous studies to the generic task categories will make the results more generalizable. We can also carry out a meta-analysis of multiple studies to find out if the results are consistent within each task category. This will help to increase the reliability of the task categories, or will serve to identify missing features and help to make our framework more comprehensive.

Finally, while designing empirical studies to analyze task performance, we can take into account the supporting tasks and dataset properties relevant to the scenario, to reduce confound in our results, as well as gather evidence on how they affect task performance. For example, in our recent VR study (Laha *et al.* 2014), we found some preliminary evidence suggesting that the effects of the VR system components may vary based on the visualization technique used to render volume data.

This classification is a first of its kind in the volume visualization community, and can surely benefit from positive critiques on how to make it more comprehensive and useful to empirical researchers in visualization, VR and 3DI communities. We invite researchers to use it, expand it, and evolve it into a rich framework of abstract task categorization, which we can reliably use and base our results upon in the future.

## Chapter 9. Conclusion and Future Work

In this dissertation, we have presented scientific evidence suggesting that immersive virtual reality and 3D interaction techniques can significantly improve analysis of volumetric datasets. Our empirical results tie the improvements in task performance to the components of a VR system, instead of the whole system. Our preliminary validation of the MR simulation approach has presented a possibility that the effects of the system components could be consistent across different VR platforms. We showed that carefully designed 3D interaction techniques could significantly improve task performance over existing 2D and 3D interaction techniques, for a variety of tasks. We also presented a framework of abstract task taxonomy that comprehensively covers visual analysis tasks with volume visualizations from multiple scientific domains.

### 9.1 Primary Contributions

Our main contributions could be grouped in three linked categories.

#### 9.1.1 Effects of VR System Fidelity

Answering research question 1 (section 1.4.1), we have presented the results of three empirical experiments investigating the effects of three components of system fidelity on analysis of volume datasets: field of regard (FOR), stereoscopy (ST) and head-based rendering or head tracking (HT). We have explored the effects of these components on microscopic computed tomographic ( $\mu$ -CT) datasets from medical biology, paleontology, and biomechanics domains, rendered as 3D textures or as isosurfaces. We found that FOR (at higher fidelity levels) improved analysis of volume data rendered as 3D textures, while ST improved analysis of isosurfaces, for a good variety of abstract task types. FOR (higher than 270°) and HT, together, significantly improved task performance for spatially complex search tasks, and ST or HT individually improved analysis of spatially complex structures in 3D texture rendering. In isosurface rendering, ST individually improved task performance for search and spatial judgment tasks.

We also found that users significantly preferred certain individual and combinations of system components to others. With 3D texture rendering, users preferred ST or HT. They felt that the ease of getting the required view and exploring the dataset in general was higher with HT, while ease of understanding the features in a dataset was higher with ST. Users felt search tasks were significantly less difficult with ST, while quantitatively, ST didn't improve their task performance in search tasks with 3D texture rendering. With isosurface rendering, users felt significantly less dizzy with HT in general, and more confident with their responses in a search task. They felt higher ease in getting desired viewpoints and also significantly less dizzy in conditions with both FOR and HT at fidelity levels matching (like high-high) rather than clashing (like high-low).

In our attempt to validate MR simulation (see VR Study-II), it appeared that the effects of the simulator platform could not be fully discounted. But the MR simulator platforms in our study differed significantly in important ways (see Table 10). The group of participants between the two platforms differed substantially in their spatial abilities as well (see section 4.1.6). We recommend that future studies evaluating the generality of MR simulation consider having more similar platforms and participant groups with comparable spatial abilities.



In general, we found that different components of system fidelity are helpful in different cases, but in general, higher fidelity is a good thing. We present below a set of design guidelines from our studies, which can inform design of effective systems for visual analysis of volume data.

#### **9.1.1.1 Implications for Designing Effective VR Systems for Volume Data Analysis**

1. Use stereo when analyzing isosurfaces, but have high field of regard for dense visualizations like 3D texture based rendering.
2. Use head-based rendering along with high field of regard ( $>270^\circ$ ) for spatial judgment tasks.
3. For analysis of complex 3D structures, use systems with head-based rendering or stereo.

Based on our findings, we believe a stereoscopic display with  $270^\circ$  field of regard, and head-based rendering would produce a good VR display in most scenarios. It would be prudent to have the ability to switch stereo and head-based rendering on and off, as the interaction between these two components is not completely clear at this point.

#### **9.1.2 Design of 3D Interaction Techniques**

Answering research question 2 (section 1.4.2), we presented our design iterations, demonstrations of three working prototypes, and two evaluation studies showing the benefits of using the Volume Cracker technique. We presented a list of Usability Criteria for designing effective 3D interaction techniques, based on our many interviews with scientists working with volume data in multiple domains. In general, our results indicate that well-designed 3DI techniques could be better than existing 3D and traditional 2D interaction techniques. More specifically, we found that the Volume Cracker technique and the Arbitrary Slicing technique significantly outperformed the Axis-aligned Slicing technique for a variety of abstract task types.

The VC technique, based on the real world metaphor of cracking or opening a book and turning through the pages, proved to be intuitive and useful for certain kinds of tasks. This indicates that design of 3D interaction techniques based on natural interaction style could be useful. Also, working closely with domain scientists and identifying a list of usability criteria for design proved to be important as well.

#### **9.1.3 Task Taxonomy for Visual Analysis of Volume Data**

Answering research question 3 (see section 1.4.3), we presented an abstract categorization of visual analysis tasks in the form of a taxonomy. We developed this taxonomy in consultation with domain scientists from multiple scientific domains, and evaluated the final design by running a survey reaching out to a broader population of scientists from around the world. We used abstract task categories from this taxonomy at various points in our research for running the VR empirical experiments and the evaluation studies of our Volume Cracker prototypes (see section 8.4), which serves to generalize our results beyond the specific domains from where the volume datasets came in our studies. Our abstract task categorization (see Figure 43) provides a framework that VR, 3DI and visualization researchers can directly use to produce generalizable empirical findings cutting across multiple scientific domains, have novice study participants in evaluation studies, and produce results more meaningful to domain experts.



#### 9.1.4 Discussion

In our studies with the three components of system fidelity (FOR, HT and ST), we observed that with two components at higher fidelity levels, task performance improved more than when all three together were at higher levels of fidelity. We also observed that in some tasks the participants had trouble analyzing the datasets in our studies at the highest level of system fidelity (with all three components at higher levels). This might appear counter-intuitive because in real world we regularly operate in highest levels of display fidelity. But it is important to note that the participants did not have as much experience analyzing the datasets in the fidelity of the real world, as they may have had in a desktop environment. We believe that the analysis in an environment combining higher fidelity levels of the three display components might have demanded extra cognitive and attentional resources. For example, the presence of the extra depth perception produced by stereo and the motion parallax produced by the head tracking, in the space surrounding them (due to the higher field of regard), might have strained their pool of attentional resources needed for analysis, thus causing some reduction in their task performance.

In several cases, we found that the conditions produced by matching fidelity levels of FOR and HT (e.g., high-high, or low-low) produced better task performance than the conditions in which their fidelity levels mismatched (e.g., high-low, or low-high). We believe that the condition with both FOR and HT at higher fidelity was closest to the real world, and the condition with both at lower fidelity was closest to a desktop environment. As such, these two conditions were more familiar to the participants than the other conditions, which might have helped in leaving more cognitive resources available for the actual task analysis. But it is also important to note that in low FOR, we can get ST cancellation problem at the edges, which worsens with HT working. On the other hand, with HT missing in a wide FOR condition, we could have perception issues with ST. These might have lowered the performance in the conditions where the fidelity levels of FOR and HT were mismatching.

The two evaluation studies of our VC prototypes (see 6.2 and 7.2) used systems differing significantly in the fidelity of the VR components. We had stereo and head tracking in our first evaluation; but both of these were missing in the second. A closer look at the design details of the original VC (see 6.1.2.3) and the asymmetric BHVC prototypes (see 7.1.5) will indicate that BHVC had far less 3D interaction than the original VC. For example, in BHVC, the slicing plane orientation was set to be always parallel to the viewing axis, the unimanual selection of the sub-volumes ignored the depth of the hand, and the cracking and joining back were automated. The lack of stereo and head tracking in BHVC increased the perceptual gap between the display and the interaction space for the user, which probably resulted in the reduction of the 3D interaction features during our design iteration process.

This points to a possible mapping between the fidelity of the system components and the amount of 3D interaction we can use in a system. This mapping probably depends on how well the user perceives the 3<sup>rd</sup> dimension in the display. The 3<sup>rd</sup> dimension of interaction and of display can probably improve cognitive perception of the depth, in both absolute and relative sense, which could be crucial for performing certain analysis tasks. We believe that such conformity is more important for analysis and visualization of 3D spatial data than that of 2D abstract data. The structures and gaps in 3D spatial data often run arbitrarily in any 3D direction, and often demand hand-eye co-ordination of the user in a 3D space, much more than 2D tasks.

## 9.2 Limitations

We tried to be circumspect and comprehensive in our research, but we acknowledge the possible limitations of our approach and findings. Although domain experts guided our VR empirical studies, the study participants were always novices. This may introduce possible confounds primarily due to the differences between experts and novices in strategies for visual data analysis, and interaction with the data. To partly compensate for this, we recommended expert strategies to our novice participants before each task. We also strictly monitored that the understandability of each task was clear to the novices, by using simpler scientific terms in the definitions, and using further clarifications wherever necessary.

The visualizations of data that we used in our studies may vary greatly between volume visualization research labs. To account for this, we included a factor in our taxonomy for visualization or rendering style. In our VR Study-III, we did a meta-analysis and compared two widely used rendering styles (3D texture and isosurfaces). But we understand that this comparison is very minor compared to the hugely variety of visualization approaches and the nuances currently existing in the volume visualization community.

Our evaluations of the 3D interaction prototypes compared our designs to just two other most popular or widely used techniques. But there are many more 2D and 3D interaction techniques currently in use in different domains. We understand that our results will have more relevance, if compared against other popular interaction designs.

Finally, our task taxonomy was designed in consultation with just a handful of scientific domain experts, and only around 175 scientists from around the world responded to our evaluation survey. This is just a very small subset if we consider the total number of people analyzing scientific data currently in different domains. We still tried to be broad and more representative by having people, from as many disparate domains as possible, review and comment on our design of the task taxonomy.

## 9.3 Impact on Theory, Practice and Design

This dissertation strengthens the foundation of the field of VR by providing empirical evidence of the fundamental effects of the components of VR systems on scientific visualization, and evidence towards validating MR simulation. It contributes directly towards consolidating scientific domains, in this era of big data, by providing an abstract task taxonomy for visual analysis of volume data.

For more than two decades, VR researchers have been running studies and gathering empirical results for finding the effects of VR system fidelity. Researchers have explored the effects of full VR systems on scientific visualizations. But not much work went into finding empirical evidence of the benefits of immersive VR and 3D interaction, and more importantly, understanding what features of VR and 3DI are important for scientific visualization tasks. This dissertation is aimed at bridging that gap and bringing back the interest and focus in using immersive VR for facilitating fundamental discoveries in the basic sciences. We believe that VR and 3DI has the potential of making visualization a very effective tool and transform the process of scientific discoveries by leveraging the benefits of the 3<sup>rd</sup> dimension for interaction. We have also provided a task taxonomy aimed at providing a framework that volume visualization, VR and 3DI researchers can use for generalizing the results from their empirical studies to multiple scientific domains.

This dissertation contributes towards defining a design space for free-hand 3D interaction (Bowman *et al.* 2005) to improve the effectiveness of scientific visualization. It provides the usability criteria for designing 3DI techniques to overcome problems faced by volume data analysts. We provide multiple prototypes of the Volume Cracker technique, tested with simulated and real-world datasets, and with novices and experts. The design iterations for creating these prototypes, and their evaluation studies provide the inspiration and methodology for future designs of effective 3D interaction techniques for scientific data analysis.

## 9.4 Future Work

This dissertation work can be extended in several important directions. We have explored the effects of just three components (FOR, ST, HT) of a VR system on task performance with volume data analysis. As researchers have suggested (Slater 2003; Bowman and McMahan 2007), there are many more components of a system, with measurable levels of fidelity. In our controlled studies, we kept all of these other components at fixed levels of fidelity. We want to know how each of these other variables individually affects, if at all, task performance and user preference for volume data analysis. Further, and probably more importantly, we want to know how all these components interact with each other and how that affects analysis of volume data. For example, we kept the display resolution of our VR displays constant in our studies. But the interplay between the resolution of the display and the data could be very important for analysis of certain kinds of volume datasets. A very high resolution of the system would be able to disclose all the smaller features in a dataset, but does the increased amount of visual information increase the clutter or the cognitive load for quick analysis of the data?

Our experiments show that different components, and their combinations, affect different tasks to different degrees. A fully immersive system (like a CAVE) may not be practical from a cost-benefit ratio, and considering the office space constraints for researchers and analysts. To be able to predict the designs of immersive systems that would work sufficiently well for different task types across multiple domains, VR researchers need to have a much more comprehensive mapping of the system components to the abstract tasks performed with volume datasets. If we consider there are around nine visual components of a VR system (Bowman and McMahan 2007), and around seven abstract task types (see Chapter 8), we can see that this dissertation scratches the surface of what still remains a largely unexplored territory of VR research.

The current list of display components (Bowman and McMahan 2007) may not be able to capture completely the problems arising due to the discontinuity of a VR display at the seams and at the corners of a VR system. These problems may create gaps in the overall system fidelity due to stereo cancellation and distortion of images at the edges, leading to glaring gaps in the human perception of visuals created using such displays. In our VR empirical studies, we found some evidence suggesting that the condition with ST and HT at higher fidelity, but with FOR at lower fidelity, was inferior to other conditions. We would want to expand the current list of display components or find a way to incorporate ways to define such characteristics of a VR system using the current components, as we find more evidence of how display continuity or discontinuity affects task performance.

If we go into the details, our studies raised some intriguing questions, such as from VR Study-III: Why do we see the pair of significant interactions that are almost mirror images of each other (Figure 26 and Figure 27)—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. In VR Study-I: Why did stereo lower grades in F1 and F2 and slow down task completion in M4? Why did higher FOR increase the perceived difficulty in F5? These questions indicate that higher fidelity of VR system components may not always improve task performance, as is widely believed.

We have summarized in section 8.4 the different situations where we leveraged our abstract task taxonomy to generalize the results within our abstract task categories. In VR Study-III, we were struggling to find consistency of the results within the task groups (see section 5.3.2). But we did find more consistency within some of the task categories in our evaluation study for the second VC prototype (see section 7.3). It remains to be seen how the community of VR, 3DI and visualization researchers use our task categorization to generalize their empirical findings.

We ran just one MR simulation experiment, in which the platforms differed quite a bit in important ways. Although we believe that a VR system can be broken down in measurable components, we need a lot more evidence to strongly claim that we can use a pre-defined level of each component to recreate a VR system of our choice. This points to the need for validation experiments for the MR simulation concept. A strong claim in support of MR simulation can be very important to our community, as we discover the connections between the fidelity levels of each component and task performance in multiple scientific domains.

Our approach of controlled experimentation, supported by MR simulation, for evaluating the effects of VR components can be extended beyond just visual components. We can use a very similar approach to find the effects of interaction, haptic, aural, olfactory and proprioceptive components of a VR system. As the VR community develop VR systems capable of extending our other senses (beyond just visual and audio) we may want to find the usefulness of such systems for scientific discovery.

This dissertation has demonstrated successful design of a 3D interaction technique (Volume Cracker), which is shown to be useful for only a few abstract task types. The usability criteria outlined in this dissertation can be used to explore the 3DI design space to prototype and evaluate 3DI techniques that address problems specific to one or more abstract task types. We want to know the individual mapping of 3DI techniques to the various task types, as well as how we can combine techniques and tools in a single package, which can comprehensively cover a set of tasks for any particular domain. For example, 3DI designers can develop a toolset for coarse interaction segmentation, which can replace the traditional pipeline of manual or automatic segmentation. This can potentially have a huge impact on medical biology, medicine, and several other domains.

Since VR gives us the possibility of creating *magic* interactions (going beyond real-world interactions), we can explore the mapping between natural and magic interaction styles to the different task types. We can also explore how bare-hand interaction works in combination with direct touch, such as table-top interaction (Klein *et al.* 2012), for analysing scientific datasets. We also observed that our first prototype of VC used stereoscopic and head-based rendering, while for our bare-hand version, we used a regular monoscopic desktop monitor. We believe that our bimanual-symmetric design of the bare-hand version didn't perform so well partly due to the mismatch between the interaction and the visual space. We want to explore the mapping between interaction techniques and components of a VR system. This can evolve our current desktop into a more effective workstation, catering to the nuances of tasks performed by scientists, by

selectively offering combinations of immersive display and matching interaction styles, such as in the HybridSpace (Bogdan *et al.* 2014).

We presented two prototypes of the Volume Cracker technique, one symmetric and the other asymmetric. We found the benefits of both of these, although for different task scenarios. The bimanual literature suggests preference of symmetric designs over asymmetric (Balakrishnan and Kurtenbach 1999), and vice versa (Balakrishnan and Hinckley 2000) for different task types, and factors affecting task performance. But we need more evidence for performance, and preference of one design over the other, for predicting what interaction style would work best, and when.

## References

- 3D-Connexion. (2014). "<http://www.3dconnexion.com/products/spacemousepro.html>."
- 3D-Slicer. (2012). "<http://www.slicer.org/>."
- Agresti, A. (2010). Analysis of Ordinal Categorical Data, John Wiley & Sons.
- Amar, R. and J. Stasko (2004). A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. Proceedings of the IEEE Symposium on Information Visualization. Washington, DC, USA, IEEE Computer Society: 143–150.
- Amira. (2012). "<http://www.amira.com/>."
- Arns, L., C. Cruz-Neira and D. Cook (1999). The Benefits of Statistical Visualization in an Immersive Environment. Proceedings of the IEEE Virtual Reality, IEEE Computer Society.
- Arthur, K. (2000). Effects of Field of View on Performance with Head-Mounted Displays. Chapel Hill, NC, University of North Carolina. **PhD**.
- Arthur, K. W., K. S. Booth and C. Ware (1993). "Evaluating 3D task performance for fish tank virtual worlds." ACM Transactions on Information Systems **11**(3): 239–265.
- Avizo. (2012). "<http://www.vsg3d.com/avizo/overview>."
- Bacim, F., M. Nabiyouni and D. Bowman (2014). Slice-n-Swipe: A Free-Hand Gesture User Interface for 3D Point Cloud Annotation. Proceedings of IEEE 9th Symposium on 3D User Interfaces (3DUI), Minneapolis, MN.
- Balakrishnan, R. and K. Hinckley (1999). The role of kinesthetic reference frames in two-handed input performance. Proceedings of the 12th annual ACM symposium on User interface software and technology. New York, NY, USA, ACM: 171–178.
- Balakrishnan, R. and K. Hinckley (2000). Symmetric bimanual interaction. Proceedings of the SIGCHI conference on Human Factors in Computing Systems, ACM.
- Balakrishnan, R. and G. Kurtenbach (1999). Exploring bimanual camera control and object manipulation in 3D graphics interfaces. ACM CHI conference.
- Ball, R. and C. North (2005). Effects of tiled high-resolution display on basic visualization and navigation tasks. CHI'05 extended abstracts on Human factors in computing systems: 1196–1199.
- Barfield, W., C. Hendrix and K. Bystrom (1997). Visualizing the structure of virtual objects using head tracked stereoscopic displays. IEEE Virtual Reality Annual International Symposium.
- Billen, M. I., O. Kreylos, B. Hamann, M. A. Jadamec, L. H. Kellogg, O. Staadt and D. Y. Sumner (2008). "A geoscience perspective on immersive 3D gridded data visualization." Computers & Geosciences **34**(9): 1056-1072.

- Bogdan, N., T. Grossman and G. Fitzmaurice (2014). HybridSpace: Integrating 3D freehand input and stereo viewing into traditional desktop applications. IEEE Symposium on 3D User Interfaces: 51-58.
- Bowman, D. and D. Raja (2004). "A method for quantifying the benefits of immersion using the cave." Presence-Connect 4(2).
- Bowman, D. A., E. Kruijff, J. J. LaViola and I. Poupyrev (2005). 3D user interfaces: theory and practice, Addison-Wesley Reading, MA.
- Bowman, D. A. and R. P. McMahan (2007). "Virtual Reality: How Much Immersion Is Enough?" Computer 40(7): 36-43.
- Bowman, D. A., C. Stinson, E. D. Ragan, S. Scerbo, T. Höllerer, C. Lee, R. P. McMahan and R. Kopper (2012). Evaluating effectiveness in virtual environments with MR simulation. Interservice/Industry Training, Simulation, and Education Conference.
- Bruckner, S. and M. E. Groller (2005). VolumeShop: an interactive system for direct volume illustration. IEEE Visualization.
- Bruckner, S. and M. E. Groller (2006). "Exploded Views for Volume Data." IEEE Transactions on Visualization and Computer Graphics 12(5): 1077-1084.
- Bruyns, C. D. and K. Montgomery (2002). "Generalized interactions using virtual tools within the spring framework: Cutting." Medicine Meets Virtual Reality 02/10: Digital Upgrades, Applying Moore's Law to Health 85: 79-85.
- Bryson, S. (1996). Virtual reality in scientific visualization. Communications of the ACM. 39: 62–71.
- Buxton, W. and B. Myers (1986). "A study in two-handed input." SIGCHI Bull. 17(4): 321–326.
- Chapman, R. F. (1998). The Insects - Structure and Function., Cambridge University Press.
- Chen, J., H. Cai, A. P. Auchus and D. H. Laidlaw (2012). "Effects of Stereo and Screen Size on the Legibility of Three-Dimensional Streamtube Visualization." IEEE Transactions on Visualization and Computer Graphics 18(12): 2130 -2139.
- Cignoni, P., C. Montani and R. Scopigno (1994). "Magicsphere: an insight tool for 3D data visualization." Computer Graphics Forum 13(3): 317–328.
- Cruz Neira, C., D. J. Sandin and T. A. DeFanti (1993). Surround-screen projection-based virtual reality: the design and implementation of the CAVE. Proceedings of the 20th annual conference on Computer graphics and interactive techniques, Anaheim, CA.
- Cutler, L. D., B. Fröhlich and P. Hanrahan (1997). Two-handed direct manipulation on the responsive workbench. Proceedings of the 1997 symposium on Interactive 3D graphics. New York, NY, USA, ACM: 107–114.
- Datey, A. (2002). Experiments in the use of immersion for information visualization. Computer Science, Virginia Tech. **Master of Science**.
- Demiralp, C., C. D. Jackson, D. B. Karelitz, S. Zhang and D. H. Laidlaw (2006). "CAVE and Fishtank Virtual-Reality Displays: A Qualitative and Quantitative Comparison." IEEE Transactions on Visualization and Computer Graphics 12(3): 323-330.



- Ekstrom, R. B., J. W. French and H. H. Harman (1979). "Cognitive factors: Their identification and replication." Multivariate Behavioral Research Monographs.
- Forsberg, A., J. Chen and D. H. Laidlaw (2009). "Comparing 3D vector field visualization methods: A user study." IEEE Transactions of Visualization and Computer Graphics **15**(6): 1219-1226.
- Gandy, M., R. Catrambone, B. MacIntyre, C. Alvarez, E. Eiriksdottir, M. Hilimire, B. Davidson and A. C. McLaughlin (2010). Experiences with an AR evaluation test bed: Presence, performance, and physiological measurement. 2010 9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR): 127 -136.
- GE-Healthcare. (2012). "[http://www.gehealthcare.com/euen/ct/products/products\\_technologies/750\\_HD/discover\\_y\\_ct750hd.html](http://www.gehealthcare.com/euen/ct/products/products_technologies/750_HD/discover_y_ct750hd.html)."
- GE-Measurement-and-Control-Solutions. (2012). "<http://www.phoenix-xray.com/>."
- Gruchalla, K. (2004). Immersive well-path editing: investigating the added value of immersion. Proceedings of the IEEE Virtual Reality.
- Guiard, Y. (1987). "Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model." Journal of motor behavior **19**: 486-517.
- Hauptmann, A. G. (1989). "Speech and gestures for graphic image manipulation." SIGCHI Bull. **20**(SI): 241–245.
- Hinckley, K., R. Pausch, J. C. Goble and N. F. Kassell (1994). Passive real-world interface props for neurosurgical visualization. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM.
- Hinckley, K., R. Pausch and D. Proffitt (1997). Attention and visual feedback: the bimanual frame of reference. Proceedings of the 1997 symposium on Interactive 3D graphics. New York, NY, USA, ACM: 121–126.
- Huff, R., C. A. Dietrich, L. P. Nedel, C. M. D. S. Freitas, J. L. D. Comba and S. D. Olabarriaga (2006). Erasing, digging and clipping in volumetric datasets with one or two hands. ACM international conference on Virtual reality continuum and its applications, ACM.
- ImageJ. (2014). "<http://imagej.nih.gov/ij/>."
- Jackson, B., T. Y. Lau, D. Schroeder, K. C. T. Jr. and D. F. Keefe (2013). "A Lightweight Tangible 3D Interface for Interactive Visualization of Thin Fiber Structures." IEEE Transactions of Visualization and Computer Graphics **19**(12): 2802-2809.
- Juan, M. C. and D. Pérez (2009). "Comparison of the Levels of Presence and Anxiety in an Acrophobic Environment Viewed via HMD or CAVE." Presence: Teleoperators and Virtual Environments **18**(3): 232-248.
- Kabbash, P., W. Buxton and A. Sellen (1994). Two-handed input in a compound task. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA, ACM: 417–423.
- Kaufman, A. E. (1990). "Volume visualization." The Visual Computer **6**(1): 1-1.
- Kaufman, A. E. (1996). "Volume visualization." ACM Computing Surveys **28**(1): 165-167.

- Keefe, D. F. and T. Isenberg (2013). "Reimagining the Scientific Visualization Interaction Paradigm." IEEE Computer **46**(5): 51 - 57.
- Kelso, J., S. G. Satterfield, L. E. Arsenault, P. M. Ketchan and R. D. Kriz (2003). "DIVERSE: A Framework for Building Extensible and Reconfigurable Device-Independent Virtual Environments and Distributed Asynchronous Simulations." Presence: Teleoperators and Virtual Environments **12**(1): 19-36.
- Kelso, J., D. L. Southard and D. Goodman (1979). "On the coordination of two-handed movements." Journal of Experimental Psychology: Human Perception and Performance **5**(2): 229-238.
- Klein, T., F. Guéniat, L. Pastur, F. Vernier and T. Isenberg (2012). "A Design Study of Direct-Touch Interaction for Exploratory 3D Scientific Visualization." Computer Graphics Forum **31**(3): 1225-1234.
- Konrad Verse, O., B. Preim and A. Littmann (2004). Virtual resection with a deformable cutting plane. Proceedings of simulation und visualisierung.
- Kreylos, O. (2008). Environment-Independent VR Development. Advances in Visual Computing. 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, Springer Berlin / Heidelberg. **5358**: 901-912.
- Laha, B. and D. A. Bowman (2012). Identifying the Benefits of Immersion in Virtual Reality for Volume Data Visualization. Immersive Visualization Revisited Workshop of the IEEE VR conference.
- Laha, B. and D. A. Bowman (2013a). Interactive Coarse Segmentation and Analysis of Volume Data with a Suite of 3D Interaction Tools. IEEE VR Workshop on Immersive Volume Interaction.
- Laha, B. and D. A. Bowman (2013b). Volume Cracker: A Bimanual 3D Interaction Technique for Analysis of Raw Volumetric Data. ACM Symposium on Spatial User Interaction (SUI), Los Angeles, CA.
- Laha, B., D. A. Bowman and J. D. Schiffbauer (2013). "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 **19**(4): 529-538.
- Laha, B., D. A. Bowman and J. J. Socha (2014). "Effects of VR System Fidelity on Analyzing Isosurface Visualization of Volume Datasets." IEEE Transactions of Visualization and Computer Graphics **20**(4): 513-522.
- Laha, B., D. A. Bowman (2013). Volume Cracker: A Bimanual 3D Interaction Technique for Volume Data Analysis (Submitted). IEEE Symposium of 3D User Interfaces (3DUI) 2013.
- Laha, B., K. Sensharma, J. D. Schiffbauer and D. A. Bowman (2012). "Effects of Immersion on Visual Analysis of Volume Data." IEEE Transactions on Visualization and Computer Graphics **18**(4): 597-606.

- LaMar, E., B. Hamann and K. I. Joy (2001). A magnification lens for interactive volume visualization. Ninth Pacific Conference on Computer Graphics and Applications, 2001. Proceedings.
- Leap-Motion. (2012). "<https://leapmotion.com/>."
- Lee, C., S. Bonebrake, T. Höllerer and D. A. Bowman (2010a). The Role of Latency in the Validity of AR Simulation. Proceedings of IEEE Virtual Reality.
- Lee, C., S. Gauglitz, T. Hollerer and D. A. Bowman (2012). Examining the equivalence of simulated and real AR on a visual following and identification task. IEEE Virtual Reality Conference. Los Alamos, CA, USA, IEEE Computer Society: 77-78.
- Lee, S. J., K. Sensharma, E. A. Fox and G. Wang (2010b). Micro-CT scanner training in a 3D virtual world: Second Life aided training and education (SLATE). BMES Annual Meeting. Austin, TX.
- Leganchuk, A., S. Zhai and W. Buxton (1998). "Manual and cognitive benefits of two-handed input: an experimental study." ACM Transactions on Computer-Human Interaction **5**(4): 326-359.
- Mapes, D. P. and J. M. Moshell (1995). "A two-handed interface for object manipulation in virtual environments." Presence: Teleoperators and Virtual Environments **4**(4): 403–416.
- Marmitt, G., H. Friedrich and P. Slusallek (2006). "Interactive volume rendering with ray tracing." Eurographics State of the Art Reports: 115–136.
- McGuffin, M. J., L. Tancu and R. Balakrishnan (2003). Using deformations for browsing volumetric data. IEEE Visualization.
- McMahan, R. P., D. A. Bowman, D. Zielinski and R. Brady (2012). "Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game." IEEE Transactions on Visualization and Computer Graphics **18**(4): 626-633.
- McMahan, R. P., D. Gorton, J. Gresock, W. McConnell and D. A. Bowman (2006). Separating the effects of level of immersion and 3D interaction techniques. Proceedings of the ACM symposium on Virtual reality software and technology, ACM.
- Meehan, M., B. Insko, M. Whitton and F. Brooks (2002). "Physiological measures of presence in stressful virtual environments." ACM Trans. Graph. **21**(3): 645–652.
- Midttun, M., R. Helland and E. Finnstrom (2000). "Virtual reality--adding value to exploration and production." The Leading Edge **19**(5): 538-544.
- Milgram, P. and F. Kishino (1994). "A Taxonomy of Mixed Reality Visual Displays." IECE Transactions on Information and Systems **18**(4): 1321-1329.
- Mishra, S., K. Sensharma, S. J. Lee, E. A. Fox and G. Wang (2012). "SLATE: Virtualizing multiscale CT training " Journal of X-Ray Science and Technology **20**(2): 239-248.
- MITO-DICOM-viewer. (2012). "<http://sourceforge.net/projects/mito/>."
- Mlyniec, P., J. Jerald, A. Yoganandan, F. J. Seagull, F. Toledo and U. Schultheis (2011). iMedic: a two-handed immersive medical environment for distributed interactive consultation. Medicine Meets Virtual Reality 18: NextMed, IOS Press. **163**: 372-378.

- Morse, E., M. Lewis and K. A. Olsen (2000). "Evaluating visualizations: using a taxonomic guide." International Journal of Human-Computer Studies **53**(5): 637-662.
- Narayan, M., L. Waugh, X. Zhang, P. Bafna and D. Bowman (2005). Quantifying the benefits of immersion for collaboration in virtual environments. Proceedings of the ACM symposium on Virtual reality software and technology: 81.
- NGRAIN-constructor-SDK. (2014). "<http://www.ngrain.com/portfolio/constructor-sdk/>."
- Ni, T., D. Bowman and J. Chen (2006). Increased display size and resolution improve task performance in Information-Rich Virtual Environments. Proceedings of Graphics Interface, Canadian Information Processing Society.
- Owen, R., G. Kurtenbach, G. Fitzmaurice, T. Baudel and W. Buxton (2005). When it gets more difficult, use both hands: exploring bimanual curve manipulation. Proceedings of Graphics Interface 2005. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, Canadian Human-Computer Communications Society: 17–24.
- Pausch, R., D. Proffitt and G. Williams (1997). Quantifying immersion in virtual reality. Proceedings of the 24th annual conference on Computer graphics and interactive techniques: 13-18.
- Pausch, R., M. A. Shackelford and D. Proffitt (1993). A user study comparing head-mounted and stationary displays. Proceedings of the IEEE 1993 Symposium on Research Frontiers in Virtual Reality: 41–45.
- Pflesser, B., U. Tiede and K. Höhne (1995). Towards realistic visualization for surgery rehearsal. Computer Vision, Virtual Reality and Robotics in Medicine. N. Ayache, Springer Berlin / Heidelberg. **905**: 487-491.
- Philips-CT-Scan. (2012). "[http://www.healthcare.philips.com/us\\_en/products/ct/products/scanners/index.wpd](http://www.healthcare.philips.com/us_en/products/ct/products/scanners/index.wpd)."
- Polys, N. F., S. Kim and D. A. Bowman (2007). "Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments." Computer Animation and Virtual Worlds **18**(1): 19-38.
- Prabhat, A. Forsberg, M. Katzourin, K. Wharton and M. Slater (2008). "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 **14**(3): 551-563.
- Ragan, E. D., R. Kopper, P. Schuchardt and D. A. Bowman (2013). "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 **19**(5): 886-896.
- Ritman, E. L. (2004). "Micro-computed tomography-current status and developments." Annual Review of Biomedical Engineering **6**: 185-208.
- Robertson, G., M. Czerwinski and M. Van Dantzich (1997). Immersion in desktop virtual reality. Proceedings of the 10th annual ACM symposium on User interface software and technology: 11–19.

- Ruddle, R. A., S. J. Payne and D. M. Jones (1999). "Navigating large-scale virtual environments: what differences occur between helmet-mounted and desktop displays?" Presence: Teleoperators & Virtual Environments **8**(2): 157–168.
- Schiffbauer, J. D., S. Xiao, K. Sensharma and G. Wang (2012). "The origin of intracellular structures in Ediacaran metazoan embryos." Geology.
- Schuchardt, P. and D. A. Bowman (2007). The benefits of immersion for spatial understanding of complex underground cave systems. Proceedings of the 2007 ACM symposium on Virtual reality software and technology, ACM.
- Sensharma, K., T. Andric, W. J. Freeman, C. L. Wyatt and G. Wang (2009). Micro-CT for osteon-like scaffolds. 11th Annual Conference of the North Carolina Tissue Engineering and Regenerative Medicine Society.
- Sensharma, K., 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 (Oct 2011). Novel Biomedical and Biological Applications using Lab-based Multi-scale CT System. Annual Meeting of the Biomedical Engineering Society.
- Sensharma, K., 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 (2011). Novel Biomedical and Biological Applications using Lab-based Multi-scale CT System. Annual Meeting of the Biomedical Engineering Society.
- Sherman, W. R. and A. B. Craig (2003). Understanding Virtual Reality-Interface, Application, and Design, Morgan-Kaufman publishers.
- Shiratuddin, M. F., W. Thabet and D. Bowman (2004). "Evaluating the effectiveness of virtual environment displays for reviewing construction 3D models." CONVR 2004: 87–98.
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proceedings of the 1996 IEEE Symposium on Visual Languages. Washington, DC, USA, IEEE Computer Society: 336.
- Siemens-Healthcare. (2012). "<http://healthcare.siemens.com/computed-tomography>."
- Slater, M. (2003). "A note on presence terminology." Presence **3**(3).
- Socha, J. J. and F. De Carlo (2008). Use of synchrotron tomography to image naturalistic anatomy in insects. Developments in X-Ray Tomography VI: 2008, SPIE, San Diego, CA.
- Socha, J. J., T. D. Förster and K. J. Greenlee (2010). "Issues of convection in insect respiration: Insights from synchrotron X-ray imaging and beyond." Respiratory Physiology & Neurobiology **173**(Supplement 1): S65-S73.
- Socha, J. J., W. K. Lee, J. F. Harrison, J. S. Waters, K. Fezzaa and M. W. Westneat (2008). "Correlated patterns of tracheal compression and convective gas exchange in a carabid beetle." Journal of Experimental Biology **211**(21): 3409-3420.
- Stock, S. R. (2008). Microcomputed tomography: Methodology and applications, CRC Press.
- Sutherland, I. (1968). A head-mounted three-dimensional display. Fall Joint Computer Conference, Arlington, VA.

- Tan, D. S., D. Gergle, P. Scupelli and R. Pausch (2003). With similar visual angles, larger displays improve spatial performance. Proceedings of the SIGCHI conference on Human factors in computing systems: 224.
- Tory, M. and T. Moller (2004). Rethinking Visualization: A High-Level Taxonomy. IEEE Symposium on Information Visualization, 2004. INFOVIS 2004: 151 -158.
- Toshiba-Medical-Systems. (2012). "<http://www.medical.toshiba.com/products/vl/index.php>."
- Ulinski, A. C. (2008). Taxonomy and experimental evaluation of two-handed selection techniques for volumetric data. Department of Computer Science, The University of North Carolina.
- Valiati, E. R. A., M. S. Pimenta and C. M. D. S. Freitas (2006). A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization. New York, NY, USA, ACM: 1-6.
- Viola, I., A. Kanitsar and M. E. Groller (2004). Importance-Driven Volume Rendering. Proceedings of the conference on Visualization, IEEE Computer Society.
- Wang, L., Y. Zhao, K. Mueller and A. E. Kaufman (2005). The magic volume lens: an interactive focus+context technique for volume rendering. IEEE Visualization.
- Wang, S. W. and A. E. Kaufman (1995). Volume Sculpting. ACM Symp. on Interactive 3D Graphics.
- Wang, Y. S., T. Y. Lee and C. L. Tai (2008). "Focus+Context Visualization with Distortion Minimization." IEEE Transactions on Visualization and Computer Graphics **14**(6): 1731 -1738.
- Ware, C. and G. Franck (1996). "Evaluating stereo and motion cues for visualizing information nets in three dimensions." ACM Transactions on Graphics **15**: 121-140.
- Weiskopf, D., K. Engel and T. Ertl (2003). "Interactive clipping techniques for texture-based volume visualization and volume shading." IEEE Transactions on Visualization and Computer Graphics **9**(3): 298-312.
- Westneat, M. W., O. Betz, R. W. Blob, K. Fezzaa, W. J. Cooper and W. K. Lee (2003). "Tracheal respiration in insects visualized with synchrotron X-ray imaging." Science **299**(5606): 558-560.
- Westneat, M. W., J. J. Socha and W. K. Lee (2008). Advances in biological structure, function, and physiology using synchrotron x-ray imaging. Annual Review of Physiology.
- Xradia-Solutions. (2012). "<http://www.xradia.com/solutions/index.php>."
- Zhang, S., 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 (2001). An Immersive Virtual Environment for DT-MRI Volume Visualization Applications: A Case Study. Proceedings of IEEE Visualization, IEEE Computer Society.
- Zhang, S., C. Demiralp and D. H. Laidlaw (2003). "Visualizing Diffusion Tensor MR Images Using Streamtubes and Streamsheets." IEEE Transactions on Visualization and Computer Graphics **9**(4): 454-462.

- Zhao, X., W. Zeng, X. D. Gu, A. E. Kaufman, W. Xu and K. Mueller (2012). "Conformal Magnifier: A Focus+Context Technique with Local Shape Preservation." IEEE Transactions on Visualization and Computer Graphics **18**(11): 1928-1941.
- Zhou, J., M. Hinz and K. D. Tonnies (2002). Focal region-guided feature-based volume rendering. First International Symposium on 3D Data Processing Visualization and Transmission.
- Zhou, M. X. and S. K. Feiner (1998). Visual task characterization for automated visual discourse synthesis. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA, ACM Press/Addison-Wesley Publishing Co.: 392–399.
- Zhou, Z., A. D. Cheok, J. Pan and Y. Li (2004). Magic Story Cube: an interactive tangible interface for storytelling. Proceedings of the ACM SIGCHI International Conference on Advances in computer entertainment technology, ACM.



## Appendix A. Experiment Documents: VR Study-I



VirginiaTech

Office of Research Compliance  
Institutional Review Board  
2000 Kraft Drive, Suite 2000 (0497)  
Blacksburg, Virginia 24060  
540/231-4606 Fax 540/231-0959  
e-mail [irb@vt.edu](mailto:irb@vt.edu)  
Website: [www.irb.vt.edu](http://www.irb.vt.edu)

### MEMORANDUM

**DATE:** March 16, 2011

**TO:** Doug A. Bowman, Bireswar Laha

**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires October 26, 2013)

**PROTOCOL TITLE:** Benefits of Immersion in 3D Micro-CT Data Analysis

**IRB NUMBER:** 11-226

Effective March 16, 2011, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the new protocol for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at <http://www.irb.vt.edu/pages/responsibilities.htm> (please review before the commencement of your research).

### PROTOCOL INFORMATION:

Approved as: **Expedited, under 45 CFR 46.110 category(ies) 5, 7**

Protocol Approval Date: **3/16/2011**

Protocol Expiration Date: **3/15/2012**

Continuing Review Due Date\*: **3/1/2012**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

\*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office ([irbadmin@vt.edu](mailto:irbadmin@vt.edu)) immediately.

cc: File

**Informed Consent for Participant of Investigative Project**  
Virginia Polytechnic Institute and State University

Title of Project: **Benefits of immersion in 3D micro-CT data analysis**

Investigators: Dr. Doug Bowman, Bireswar Laha.

**I. THE PURPOSE OF THIS RESEARCH/PROJECT**

You are invited to participate in a study to investigate the benefits of exploring 3D microscopic computed tomography data in a 3D immersive virtual environment. This research aims to quantify possible benefits for analysis of the 3D datasets at a higher level of immersion, and aims to find out which components of immersion contribute to the perceived benefits.

**II. PROCEDURES**

Upon arrival, you will be asked to read and sign this informed consent form. After that you will be given an introduction of our experiment background, facilities to be used, and study procedures. Then you will be asked to take a spatial ability test. Following the test, you will perform 5-6 training tasks with volume visualization of a scaffold dataset (<http://www.imaging.sbes.vt.edu/research/sam-ct/library/joseph-freeman/>) on the Viscube, in which you will be introduced to the 3D interface of the hardware, as well as be trained on how to analyze a 3D volume visualization of micro-CT data. Each task will involve answering a question pertinent to the dataset.

You will rest for couple of minutes at this point, as the experimenter loads the first micro-CT dataset, which is volume visualization from micro-CT data of a mouse-limb (<http://www.imaging.sbes.vt.edu/research/sam-ct/library/chad-markert/>) on the Viscube. Then you will perform 4 tasks on the dataset, each of which will involve answering a question pertinent to the dataset. Your performance will be recorded by tracking the time taken by you to complete the task, and an offline evaluation of the accuracy your answer by the experimenters. Each task will be followed by evaluating your subjective rating of the task difficulty and your subjective level of confidence in your answer, both on a 1-7 Likert's scale on paper.

After completion of the 4 tasks, you will be asked to take a short break of around 5 mins, while the experimenter will load the last dataset on the Viscube, which is a parapandorina dataset (<http://www.imaging.sbes.vt.edu/research/sam-ct/library/shuhai-xiao/>). After resuming from rest, you will perform a series of 14 tasks with the parapandorina dataset, each of which is similar to the ones for the mouse limb dataset. You will be asked to rest after the 7th task for 5 mins. Through the entire process of task evaluation on these two datasets, the experimenter will record your subjective answers with respect to different questions involving the analysis of the datasets, which will help us better understand how you performed on the tasks. Following this, you will complete a post-questionnaire. A free form interview will be conducted if you have any additional comments not addressed by the questionnaire.

**III. RISKS**

There will not be more than minimal risks involved in our study.

**IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us to quantify the benefits of immersion for analysis of 3D visualized micro-CT data. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

**V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The

information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

#### **VI. COMPENSATION**

Your participation is voluntary and unpaid.

#### **VII. FREEDOM TO WITHDRAW**

You are free to withdraw from this study at any time for any reason.

#### **VIII. APPROVAL OF RESEARCH**

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: MN-WXYZ

#### **IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION**

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (please print)

\_\_\_\_\_  
Contact: phone or address or

\_\_\_\_\_  
Email address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Computer Science Department (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
Graduate student, Computer Science Department  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

Participant #:

## Background Questionnaire: Phase-I

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                      Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired                      somewhat tired                      a little tired                      not tired at all

Rate your expertise with computers: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

...for fun? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

Rate your experience with 3D stereo movies, at theaters and on TV: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

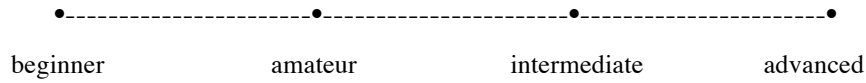
Comments:

Rate your experience with 3D immersive virtual environments, like the CAVE: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

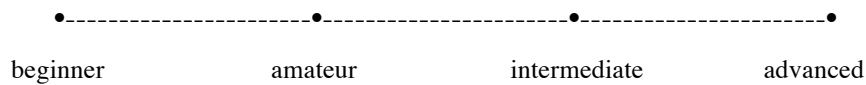
Comments:

Rate your experience with analyzing computed tomography (CT) data: (circle one)



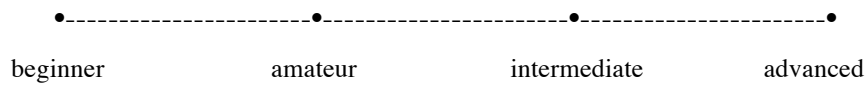
Comments:

Rate your experience with microscopic CT datasets: (circle one)



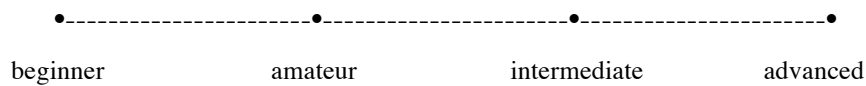
Comments:

If you had significant experience with micro-CT data, then rate your experience with the mouse limb dataset, otherwise please ignore: (circle one)



Comments:

If you had significant experience with micro-CT data, then rate your experience with the Parapandorina dataset, otherwise please ignore: (circle one)



Comments:



---

## Post-Experiment Questionnaire: Phase-I

Please complete the following questions.

1. Were you more comfortable in the low immersion or high immersion condition? Why?
2. Was it easier to get the view that you wanted in the low immersion or high immersion condition? Why?
3. Was it easier to explore the dataset in low immersion or high immersion condition? Why?
4. Was it easier to understand what you were looking at in the low or high immersion condition? Why?
5. Do you think stereo display was of any help as opposed to mono display? Why?

6. Do you think having display on 3 walls of the CAVE was better than on just 1 wall? Why?
7. At the high display fidelity condition, as you moved your head, the data was rendered from your head position. Do you think having a head-tracked display helped you? Why?
8. What was your strategy for exploring the dataset with the high display fidelity?
9. What was your strategy for exploring the dataset with the low display fidelity?
10. Which dataset was easier to explore and why?
11. Please feel free to add any comments, suggestions, and concerns.

Participant #:

## Background Questionnaire: Phase-II

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                      Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired              somewhat tired              a little tired              not tired at all

Rate your expertise with computers: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

...for fun? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

Rate your experience with 3D stereo movies, at theaters and on TV: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

Comments:

Rate your experience with 3D immersive virtual environments, like the CAVE: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with analyzing computed tomography (CT) data: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with microscopic CT datasets: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Please rate your experience with the mouse limb dataset from ICTAS. If you do not recognize the name, please mark beginner: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Please rate your experience with the parapandorina (fossil) dataset from ICTAS. If you do not recognize the name, please mark beginner: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Name \_\_\_\_\_

# CUBE COMPARISONS TEST -- S-2 (Rev.)

Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

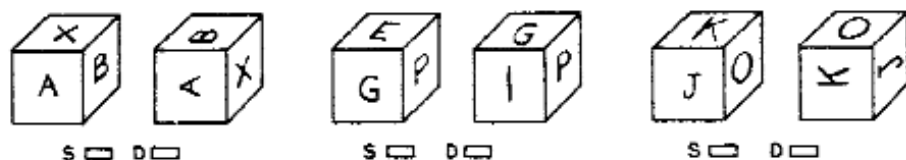


The first pair is marked D because they must be drawings of different cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right-hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

Note: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, any letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left-hand drawing and at the base of the A on the right-hand drawing. The second pair is "different" because P has its side next to G on the left-hand cube but its top next to G on the right-hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.



Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.



You will have 3 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.



DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.



Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.



## Part 1 (3 minutes)



1.    
S ☐ D ☐

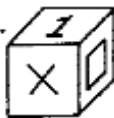

2.    
S ☐ D ☐

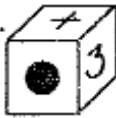

3.    
S ☐ D ☐



4.    
S ☐ D ☐

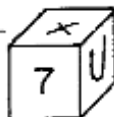
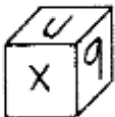
5.    
S ☐ D ☐

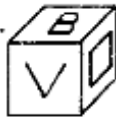

6.    
S ☐ D ☐

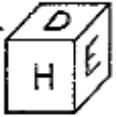

7.    
S ☐ D ☐

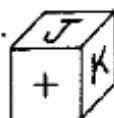
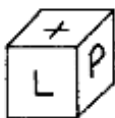
8.    
S ☐ D ☐


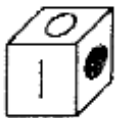
9.    
S ☐ D ☐

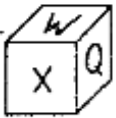

10.    
S ☐ D ☐



11.    
S ☐ D ☐

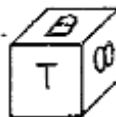

12.    
S ☐ D ☐



13.    
S ☐ D ☐



14.    
S ☐ D ☐

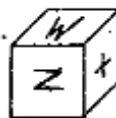

15.    
S ☐ D ☐

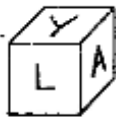

16.    
S ☐ D ☐

17.    
S ☐ D ☐

18.    
S ☐ D ☐

19.    
S ☐ D ☐

20.    
S ☐ D ☐

21.    
S ☐ D ☐

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.



Participant #:

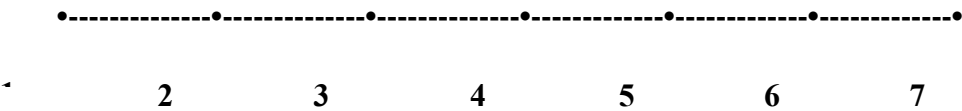
—

Post-Experiment Questionnaire: Phase-II

Please complete the following questions.

1. Rate each condition you experienced, in terms of your comfort level:

With head tracking (1 being very low, 7 being very high):



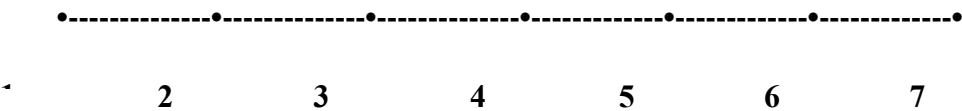
Without head tracking (1 being very low, 7 being very high):



Comments:

2. Rate each condition you experienced, in terms of ease of getting the view you wanted, and exploring the dataset in general:

With head tracking (1 being very low, 7 being very high):



Without head tracking (1 being very low, 7 being very high):



Comments:

3. Rate each condition you experienced, in terms of ease of understanding the features of a dataset and doing the different tasks with the dataset:

With head tracking (1 being very low, 7 being very high):



Without head tracking (1 being very low, 7 being very high):



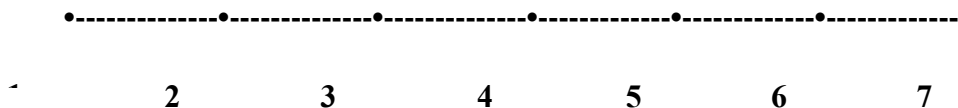
Comments:

4. Rate your strategies for exploring the dataset in the condition with head tracking:

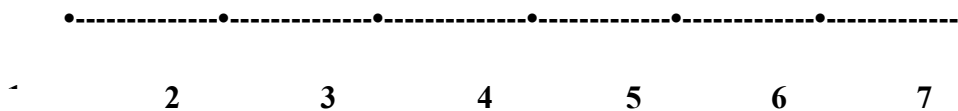
- a. Changed viewpoint by rotating or grabbing the dataset (1 being very infrequently, 7 being very frequently):



- b. Sliced the dataset (1 being very infrequently, 7 being very frequently):



- c. Walked around the dataset to look from different point (1 being very infrequently, 7 being very frequently):



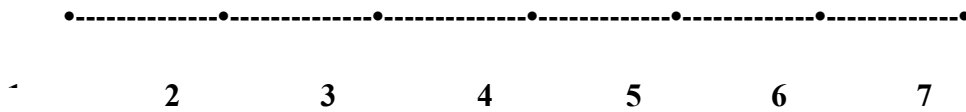
Please describe any other strategies that you used:

5. Rate your strategies for exploring the dataset in the condition without head tracking:

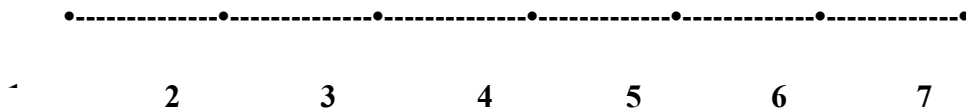
- a. Changed viewpoint by rotating or grabbing the dataset (1 being very infrequently, 7 being very frequently):



- b. Sliced the dataset (1 being very infrequently, 7 being very frequently):



- c. Walked around the dataset to look from different point (1 being very infrequently, 7 being very frequently):



Please describe any other strategies that you used:

6. Which dataset was easier to explore and why?

7. Please feel free to add any comments, suggestions, and concerns.

## Tasks for VR Study-I

### Tasks with the mouse limb dataset:

*T1. 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.

*T2. 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.

*T3. 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.

*T4. 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?

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.

## Appendix B. Experiment Documents: VR Study-II



VirginiaTech

Office of Research Compliance  
Institutional Review Board  
2000 Kraft Drive, Suite 2000 (0497)  
Blacksburg, Virginia 24060  
540/231-4606 Fax 540/231-0959  
e-mail [irb@vt.edu](mailto:irb@vt.edu)  
Website: [www.irb.vt.edu](http://www.irb.vt.edu)

### MEMORANDUM

**DATE:** April 2, 2012

**TO:** Doug A. Bowman, Bireswar Laha

**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)

**PROTOCOL TITLE:** Effects of Immersion in Analysis of Micro-CT Datasets With a Head-Mounted Display

**IRB NUMBER:** 12-057

Effective April 2, 2012, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at <http://www.irb.vt.edu/pages/responsibilities.htm> (please review before the commencement of your research).

### PROTOCOL INFORMATION:

Approved as: **Expedited, under 45 CFR 46.110 category(ies) 7**

Protocol Approval Date: **1/25/2012**

Protocol Expiration Date: **1/24/2013**

Continuing Review Due Date\*: **1/10/2013**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*



Date*	OSP Number	Sponsor	Grant Comparison Conducted?

\*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office ([irbadmin@vt.edu](mailto:irbadmin@vt.edu)) immediately.

cc: File

**Informed Consent for Participant of Investigative Project**  
Virginia Polytechnic Institute and State University

Title of Project: **Effects of immersion in analysis of micro-CT datasets with a head-mounted display**

Investigators: Dr. Doug Bowman, Bireswar Laha.

**I. THE PURPOSE OF THIS RESEARCH/PROJECT**

You are invited to participate in a study to investigate the effects of immersion in exploring 3D microscopic computed tomography data in a 3D immersive virtual environment. This research aims to quantify possible benefits for analysis of the 3D datasets at a higher level of immersion, and aims to find out which components of immersion contribute to the perceived benefits.

**II. PROCEDURES**

Upon arrival, you will be asked to read and sign this informed consent form. After that you will be given an introduction of our experiment background, facilities to be used, and study procedures. Then you will be asked to take a spatial ability test. Following the test, you will be introduced to the Head-Mounted Display (HMD) system. You will perform 5-6 training tasks with volume visualization of a scaffold dataset on the HMD, in which you will be introduced to the 3D interface of the hardware, as well as will be trained on analyzing a 3D volume visualization of micro-CT data. Each task will involve answering a question pertinent to the dataset. You will rest for couple of minutes at this point, while the experimenter loads the first micro-CT dataset on the HMD, which is the volume visualization from micro-CT data of a mouse-limb. You will perform 4 tasks with the dataset, each of which will involve answering a question pertinent to the dataset. Your performance will be recorded by tracking the time taken to complete the task, and an offline evaluation of the accuracy of your answer by the experimenters. Each task will be followed by evaluating your subjective rating of the task difficulty and your subjective level of confidence in your answer, both on 1-7 Likert's scale on paper. After completion of the 4 tasks, you will be asked to take a short break of around 5 mins, while the experimenter loads the last dataset on the HMD, which is the parapandorina dataset.

After resuming from rest, you will perform a series of 7 tasks with the parapandorina dataset, each of which is similar to the ones for the mouse limb dataset. Through the entire process of task evaluation on these two datasets, the experimenter will record your subjective answers with respect to different questions involving the analysis of the datasets, which will help us better understand how you performed on the tasks. Following this, you will complete a post-questionnaire. A free form interview will be conducted if you have any additional comments not addressed by the questionnaire.

**III. RISKS**

There will not be more than minimal risks involved in our study.

**IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us to quantify the benefits of immersion for analysis of 3D visualized volume data. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

**V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

**VI. COMPENSATION**

Your participation is voluntary and unpaid.

## VII. FREEDOM TO WITHDRAW

You are free to withdraw from this study at any time for any reason.

## VIII. APPROVAL OF RESEARCH

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: 12-057

## IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (please print)

\_\_\_\_\_  
Contact: phone or address or

\_\_\_\_\_  
Email address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Computer Science Department (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
Graduate student, Computer Science Department  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

Participant #:

## Background Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                  Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired                  somewhat tired                  a little tired                  not tired at all

Rate your expertise with computers: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

...for fun? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

Rate your experience with 3D stereo movies, at theaters and on TV: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

Comments:

Rate your experience with 3D immersive virtual environments, like a Head-Mounted Display or the CAVE: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with analyzing computed tomography (CT) data: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with microscopic CT datasets: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Please rate your experience with the mouse limb dataset from ICTAS. If you do not recognize the name, please mark beginner: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Please rate your experience with the parapandorina (fossil) dataset from ICTAS. If you do not recognize the name, please mark beginner: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Name \_\_\_\_\_

CUBE COMPARISONS TEST -- S-2 (Rev.)

Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

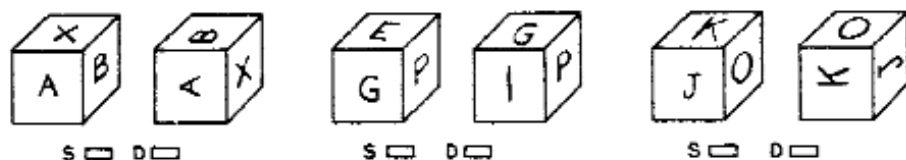


The first pair is marked D because they must be drawings of different cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right-hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

Note: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, any letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left-hand drawing and at the base of the A on the right-hand drawing. The second pair is "different" because P has its side next to G on the left-hand cube but its top next to G on the right-hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.

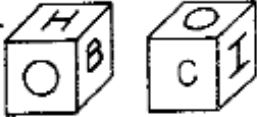
Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.


You will have 3 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.


DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.


Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.

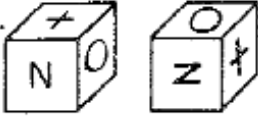
## Part 1 (3 minutes)

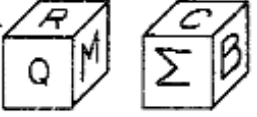
1.   
S ☐ D ☐

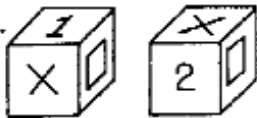
2.   
S ☐ D ☐

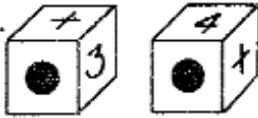
3.   
S ☐ D ☐

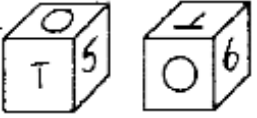
4.   
S ☐ D ☐

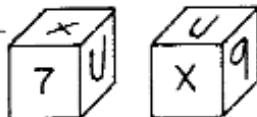
5.   
S ☐ D ☐

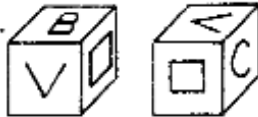
6.   
S ☐ D ☐

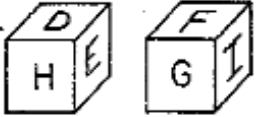
7.   
S ☐ D ☐

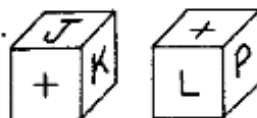
8.   
S ☐ D ☐

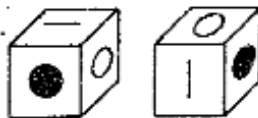
9.   
S ☐ D ☐

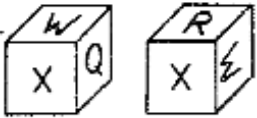
10.   
S ☐ D ☐


11.   
S ☐ D ☐

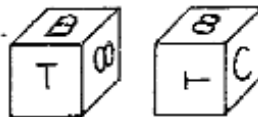
12.   
S ☐ D ☐

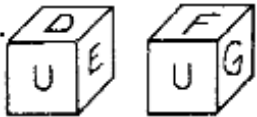
13.   
S ☐ D ☐


14.   
S ☐ D ☐

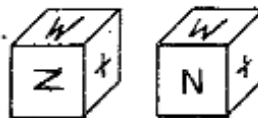
15.   
S ☐ D ☐

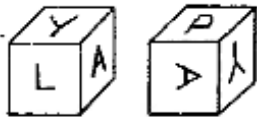
16.   
S ☐ D ☐

17.   
S ☐ D ☐

18.   
S ☐ D ☐

19.   
S ☐ D ☐

20.   
S ☐ D ☐

21.   
S ☐ D ☐

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.



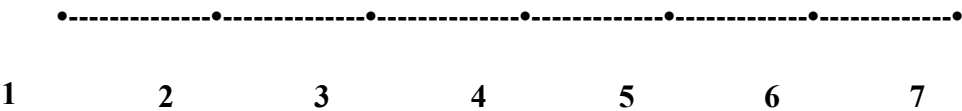
Participant #:  
—

Post-Experiment Questionnaire

Please complete the following questions.

1. Rate each condition you experienced, in terms of your comfort level:

With head tracking (1 being very low, 7 being very high):



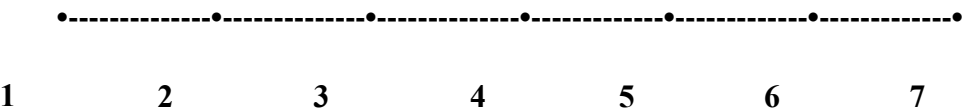
Without head tracking (1 being very low, 7 being very high):



Comments:

2. Rate each condition you experienced, in terms of ease of getting the view you wanted, and exploring the dataset in general:

With head tracking (1 being very low, 7 being very high):



Without head tracking (1 being very low, 7 being very high):



Comments:

3. Rate each condition you experienced, in terms of ease of understanding the features of a dataset and doing the different tasks with the dataset:

With head tracking (1 being very low, 7 being very high):



Without head tracking (1 being very low, 7 being very high):



Comments:

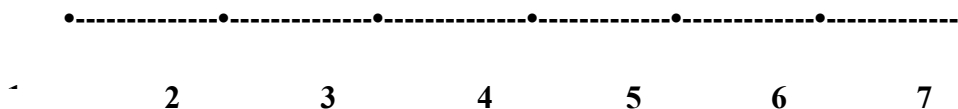
4. Rate your strategies for exploring the dataset in the condition with head tracking:
- d. Changed viewpoint by rotating or grabbing the dataset (1 being very infrequently, 7 being very frequently):



- e. Sliced the dataset (1 being very infrequently, 7 being very frequently):



- f. Walked around the dataset to look from different point (1 being very infrequently, 7 being very frequently):

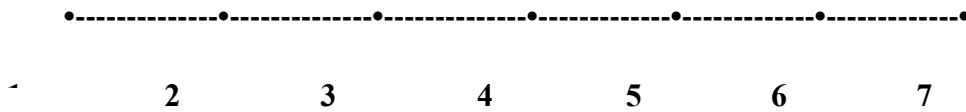


Please describe any other strategies that you used:

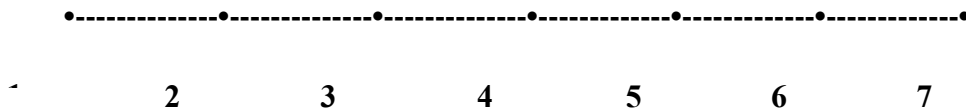
5. Rate your strategies for exploring the dataset in the condition without head tracking:
- d. Changed viewpoint by rotating or grabbing the dataset (1 being very infrequently, 7 being very frequently):



- e. Sliced the dataset (1 being very infrequently, 7 being very frequently):



- f. Walked around the dataset to look from different point (1 being very infrequently, 7 being very frequently):



Please describe any other strategies that you used:

6. Which dataset was easier to explore and why?

7. Please feel free to add any comments, suggestions, and concerns.

## Tasks for VR Study-II

### **Tasks with the mouse limb dataset:**

*T1. 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.

*T2. For your next task, you are allowed a maximum time of 1 minute. This is a bone of the mouse limb. Please study the inner core of the bone for a minute. After that I will ask you to choose an answer to describe it. Your time starts now.*

Response options - A. Network of fibers/strands. B. Fluffy like cotton. C. Spongy. D. Single block; no texture. E. Collection of many small pieces.

**(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.

*T3. 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.

*T4. 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 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.*

Response options:

- a. Half of a nearly spherical object, with distinct internal chambers. Outside resembles a soccer ball.

- b. A completely spherical object, with a cloud-like internal structure. Outside resembles a basketball.
- c. A spherical object cut in half, with a grainy texture inside. Outside resembles a soccer ball.
- d. A completely spherical object with no distinct internal structure.
- e. Half of an oblong object, like an American football, with distinct internal chambers.

**(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 various 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?

Response options:

	Thickness	Texture/Close Inspection	Brightness		Junctions
A	Comparable	Variable; predominant sheet-like	Darker	than	X-shaped surrounding material
B	Variable	Similar fabric; clotted, similar to bunches of grapes	Brighter	than	X-shaped surrounding
C	Variable	Similar fabric; clotted, similar to bunches of grapes	Brighter	than	Y-shaped surrounding
d	Comparable	Variable; clotted, similar to bunches of grapes	Darker	than	Y-shaped surrounding material
e	Variable	Similar fabric; predominant sheet-like	Brighter	than	X-shaped surrounding

**(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).

Response options:

- a. Always floating within the cell not attached to boundaries.
- b. Attached to the outer surface of the fossil structure.
- c. Contained within cell boundaries only, no part of the intracellular structure juts into the interior of the cell.
- d. When close to the cell boundaries, they always merge with the boundary. Otherwise, they are completely detached and floating within the cell.
- e. Always touching cell boundaries.

(**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?

Response options - A. Cubic, solid, and variable in size. B. Pill-shaped (spherocylindrical), solid, and variable in size. C. Spherical or moon-shaped, typically hollow rather than solid, and comparable in size. D. Spherical or moon-shaped, typically solid rather than hollow, and comparable in size. E. Pill-shaped, typically hollow rather than solid, and comparable in size.

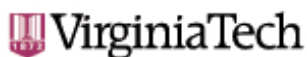
(**General description**) – The task was about describing the structures of the intracellular bodies in the cells.

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.

## Appendix C. Experiment Documents: VR Study-III



Office of Research Compliance  
Institutional Review Board  
2000 Kraft Drive, Suite 2000 (0497)  
Blacksburg, VA 24060  
540/231-4606 Fax 540/231-0959  
email [irb@vt.edu](mailto:irb@vt.edu)  
website <http://www.irb.vt.edu>

### MEMORANDUM

**DATE:** February 15, 2013  
**TO:** Doug A Bowman, Bireswar Laha  
**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)  
**PROTOCOL TITLE:** Validation of MR Simulation for the Effects of Display Fidelity in Analysis of Tracheal Systems of beetles from the *Platynus* and *Pterostichus* genera  
**IRB NUMBER:** 13-137

Effective February 14, 2013, the Virginia Tech Institutional Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

### PROTOCOL INFORMATION:

Approved As: **Expedited, under 45 CFR 46.110 category(ies) 7**  
Protocol Approval Date: **February 14, 2013**  
Protocol Expiration Date: **February 13, 2014**  
Continuing Review Due Date\*: **January 30, 2014**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

\* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.



## **Informed Consent for Participant of Investigative Project**

Virginia Polytechnic Institute and State University

Title of Project: **Validation of MR Simulation for the Effects of Display Fidelity in Analysis of Tracheal Systems of beetles from the Platynus and Pterostichus genera: Phase-I**

Investigators: Dr. Doug Bowman, Bireswar Laha.

### **I. THE PURPOSE OF THIS RESEARCH/PROJECT**

- You are invited to participate in a study investigating the effects of different properties of virtual reality (VR) display systems on the analysis of 3D scientific data.
- In this study, you will analyze the tracheal systems of a beetle, scanned, segmented in isosurface format, and visualized on a state-of-the-art spatially immersive display system - the four-sided CAVE.
- Your participation will help us gather quantitative and qualitative insights into, and generalize possible benefits for the analysis of 3D volume datasets at higher levels of display fidelity (visual immersion).

### **II. PROCEDURES**

Upon arrival, you will be asked to read, and sign this informed consent form. After that you will be given an introduction of our experiment background, facilities to be used, and study procedures. Then you will be asked to take a spatial ability test. Following the test, you will be introduced to the CAVE system. You will perform 5-6 training tasks with isosurfaces generated from the scans of a beetle of the Pterostichus genus. During the training, you will be introduced to the 3D interface of the hardware, as well as will be trained on the different strategies, and interactions for performing tasks very similar to those you will face during the main part of the study next. After the training, you will take a break for few minutes, while the experimenter loads the main dataset, which are the isosurfaces generated from the scans of a beetle of the Platynus genus. After that you will perform about 12 to 15 tasks, one after another, with the Platynus dataset. Each task will consist of listening to a question, analyzing the dataset for the answer, and responding the answer back to the experimenter. Your performance will be recorded by the experimenter by tracking the time taken to complete each task, and by an offline evaluation of the accuracy of your answer by the experimenters. Each task will be followed by evaluating your subjective rating of the task difficulty, and your subjective level of confidence in your answer, both on 1-7 Likert's scale on paper. Through the entire process of task evaluation, the experimenter will record your subjective answers with respect to different questions involving the analysis of the datasets, which will help us better understand how you performed on the tasks. Following this, you will complete a post-questionnaire. A free form interview will be conducted if you have any additional comments not addressed by the questionnaire. The entire study will last for about an hour, and 15 minutes.

### **III. RISKS**

There will not be more than minimal risks involved in our study.

### **IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us to quantify the effects of higher levels of display fidelity for analysis of 3D visualized volume data. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

### **V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

### **VI. COMPENSATION**

Virginia Tech Institutional Review Board Project No. 13-137  
Approved February 14, 2013 to February 13, 2014

Your participation is voluntary and unpaid.

## **VII. FREEDOM TO WITHDRAW**

You are free to withdraw from this study at any time for any reason.

## **VIII. APPROVAL OF RESEARCH**

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: 13-137

## **IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION**

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (please print)

\_\_\_\_\_  
Contact: phone or address or

\_\_\_\_\_  
Email address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Department of Computer Science (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
PhD Candidate, Department of Computer Science  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

Participant #:

## Background Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                      Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired                      somewhat tired                      a little tired                      not tired at all

Rate your expertise with computers: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

...for fun? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

Rate your experience with video games like Xbox, Nintendo, etc.: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

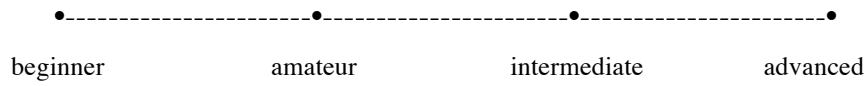
Comments:

Rate your experience with 3D stereo movies, at theaters and on TV: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

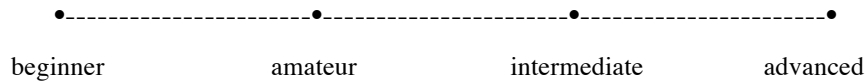
Comments:

Rate your experience with 3D immersive virtual environments, like a Head-Mounted Display or the CAVE: (circle one)



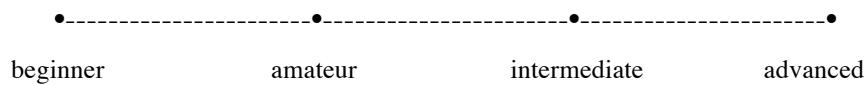
Comments:

Rate your experience with analyzing volume datasets, from any domain: (circle one)



Comments:

Rate your experience with analyzing isosurfaces of insects from biomechanics: (circle one)



Comments:

Name \_\_\_\_\_

# CUBE COMPARISONS TEST -- S-2 (Rev.)

Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

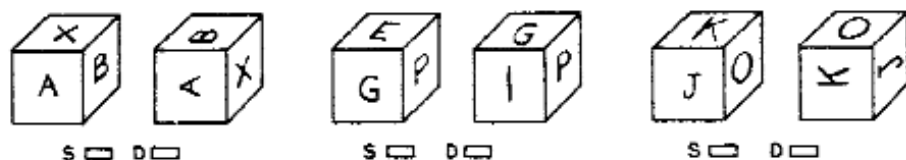


The first pair is marked D because they must be drawings of different cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right-hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

Note: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, any letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left-hand drawing and at the base of the A on the right-hand drawing. The second pair is "different" because P has its side next to G on the left-hand cube but its top next to G on the right-hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.



Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.



You will have 3 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.



DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.



Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.



## Part 1 (3 minutes)



1.    
S ☐ D ☐

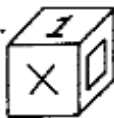

2.    
S ☐ D ☐

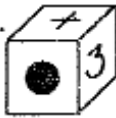

3.    
S ☐ D ☐



4.    
S ☐ D ☐

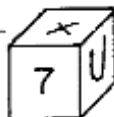
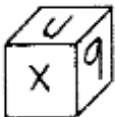
5.    
S ☐ D ☐

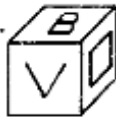

6.    
S ☐ D ☐

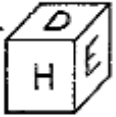

7.    
S ☐ D ☐

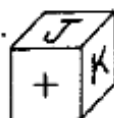
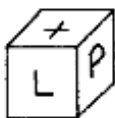
8.    
S ☐ D ☐


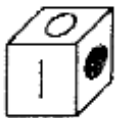
9.    
S ☐ D ☐

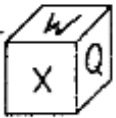

10.    
S ☐ D ☐



11.    
S ☐ D ☐

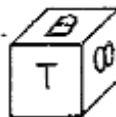

12.    
S ☐ D ☐



13.    
S ☐ D ☐



14.    
S ☐ D ☐

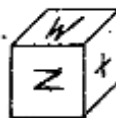

15.    
S ☐ D ☐

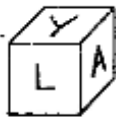

16.    
S ☐ D ☐

17.    
S ☐ D ☐

18.    
S ☐ D ☐

19.    
S ☐ D ☐

20.    
S ☐ D ☐

21.    
S ☐ D ☐

DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Copyright © 1962, 1976 by Educational Testing Service. All rights reserved.

Participant #:  
—

## Post-Experiment Questionnaire

Please answer the following questions.

1. How much do you agree with the statement - “I could easily get the different viewpoints I wanted to explore the dataset in general, and for the various analysis tasks I performed”?



Comments:

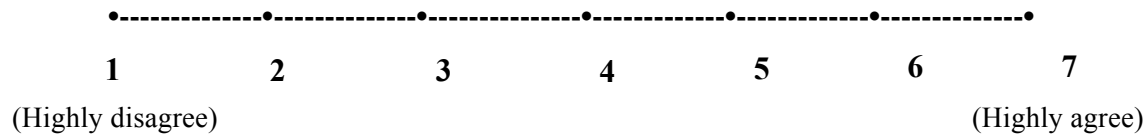
2. How much do you agree with the statement - “I could easily understand the different features of the dataset, and perform the various analysis tasks given to me”?



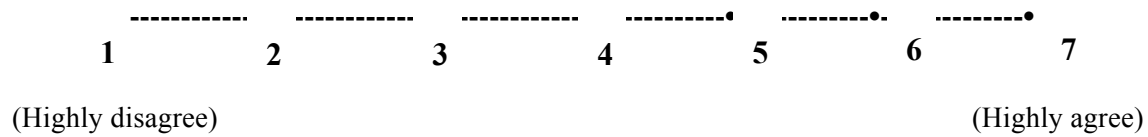
Comments:

3. For exploring/analyzing the dataset:
- a. How much do you agree with the statement - “I frequently changed the viewpoint by rotating or grabbing the dataset”?



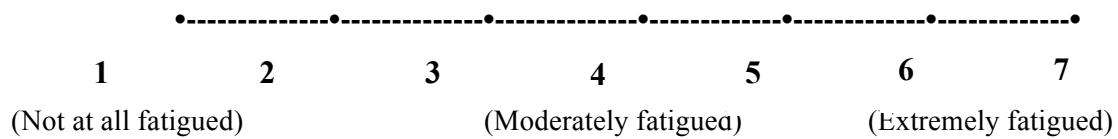


- b. How much do you agree with the statement - "I frequently walked around the dataset to look from different viewpoints"?



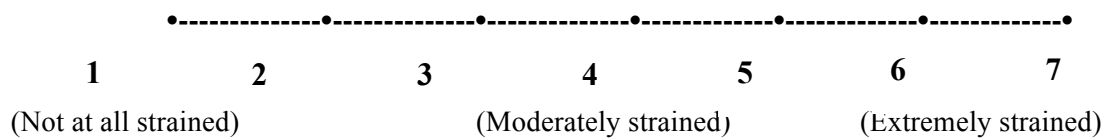
Please describe any other strategies that you used:

4. Rate your level of fatigue after using the system to analyze the dataset.



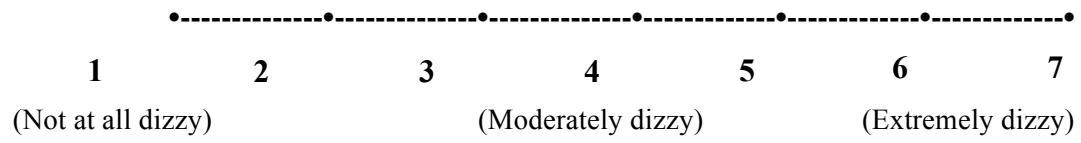
Comments:

5. Rate your level of eye strain after using the system to analyze the dataset.



Comments:

6. Rate your level of dizziness after using the system to analyze the dataset.



Comments:

7. Please feel free to add any comments, suggestions, and concerns.

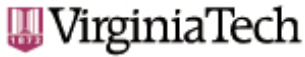
## Tasks for VR Study-III

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**)

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**)

## Appendix D. Experiment Documents: Volume Cracker Study



Office of Research Compliance  
Institutional Review Board  
2000 Kraft Drive, Suite 2000 (0497)  
Blacksburg, VA 24060  
540/231-4606 Fax 540/231-0959  
email [irb@vt.edu](mailto:irb@vt.edu)  
website <http://www.irb.vt.edu>

### MEMORANDUM

**DATE:** October 30, 2012  
**TO:** Doug A Bowman, Bireswar Laha  
**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)  
**PROTOCOL TITLE:** Empirical evaluation of Volume Cracker  
**IRB NUMBER:** 12-927

Effective October 30, 2012, the Virginia Tech Institution Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

### PROTOCOL INFORMATION:

Approved As: **Exempt, under 45 CFR 46.110 category(ies) 2**  
Protocol Approval Date: **October 30, 2012**  
Protocol Expiration Date: **N/A**  
Continuing Review Due Date\*: **N/A**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

\* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

**Informed Consent for Participant of Investigative Project**  
Virginia Polytechnic Institute and State University

Title of Project: **Volume Cracker: A Novel Bimanual 3D Interaction tool for analyzing volume data**

Investigators: Dr. Doug Bowman, Bireswar Laha.

**I. THE PURPOSE OF THIS RESEARCH/PROJECT**

You are invited to participate in a study to compare a novel bimanual 3D interaction tool, called the *Volume Cracker*, to two other standard 3D interaction tools used for slicing a volume data.

**II. PROCEDURES**

Upon arrival, you will be asked to read and sign this informed consent form. After that you will be given an introduction of our experiment background, facilities to be used, and study procedures. You will then be introduced to the Vizwall in the Blacklab. This is a rear projected stereo display wall. We shall use just the right wall of the VizWall system.

You will then begin the tasks one by one. You will perform one task from any four task categories, selected by the experimenter. For task categories, please ask the experimenter to show the document titled scoresheet.docx. You will perform each of the four tasks, with each of the three interaction styles - Volume Cracker (VC), Unconstrained Slicing (US), and Axis-aligned slicing (AAS). The Axis-aligned slicing technique uses a mouse for interaction kept on a table in front of the VizWalls. The Unconstrained slicing technique uses the intersense wand and the intersense wired head tracker in our lab (<http://www.intersense.com/pages/20/14>). The volume cracker uses the head tracker, the wand in the right hand, and a wired 5DT data glove in the left hand (<http://www.5dt.com/products/pdataglove5u.html>). For every task, you will face a simulated volumetric dataset, created by the experimenter. The datasets will either have random letters of the English alphabet scattered inside the volume, or simulated pipes running inside the volume, in pattern consistent with the task category selected.

Before beginning each task, you will perform a training task with the same interaction style (VC, US, or AAS) and a similar dataset, which you will be facing immediately after. For every task, you will get a maximum of three minutes to complete the task. During the actual task, the experimenter will record the time taken by you to complete the task, and also record a score, either 1 or 0 (correct and complete or incorrect and incomplete), in a scoresheet.

After the tasks are completed, you will fill out a post-questionnaire. A free form interview will be conducted if you have any additional comments not addressed by the questionnaire.

The entire experiment will last about an hour, with a maximum of one hour 15 minutes.

**III. RISKS**

There will not be more than minimal risks involved in our study.

**IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us understand the benefits of using the *Volume Cracker* for analysis of 3D volume visualized data, as opposed to traditional slicing techniques. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

**V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

## VI. COMPENSATION

Your participation is voluntary and unpaid.

## VII. FREEDOM TO WITHDRAW

You are free to withdraw from this study at any time for any reason.

## VIII. APPROVAL OF RESEARCH

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: 12-927

## IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (please print)

\_\_\_\_\_  
Contact: phone or address or

\_\_\_\_\_  
Email address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Computer Science Department (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
Graduate student, Computer Science Department  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

Participant #:

## Background Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                      Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired                      somewhat tired                      a little tired                      not tired at all



Rate your expertise with computers: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

...for fun? (circle the best answer)

a. not at all

b. once a month

c. once a week

d. several times a week

e. daily

Rate your experience with 3D stereo movies, at theaters and on TV: (circle one)

•-----•-----•-----•  
Never                      Sometimes                      Often                      Everyday

Comments:

Rate your experience with display walls or power walls: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with 3D immersive virtual environments, like a Head-Mounted Display or the CAVE: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Rate your experience with analyzing volume data, e.g. CT/MRI/ultrasound scans: (circle one)

•-----•-----•-----•  
beginner                      amateur                      intermediate                      advanced

Comments:

Participant #:  
—

Post-Experiment Questionnaire

Please complete the following questions.

1. How easy to use were the three tools that you experienced?

Volume Cracker (1 being very hard to use, 7 being very easy to use):

•-----•-----•-----•-----•-----•-----•-----•

1                    2                    3                    4                    5                    6                    7

Unconstrained Slicing (1 being very hard to use, 7 being very easy to use):

•-----•-----•-----•-----•-----•-----•-----•

1                    2                    3                    4                    5                    6                    7

Axis-Aligned Slicing (1 being very hard to use, 7 being very easy to use):

•-----•-----•-----•-----•-----•-----•-----•

1                    2                    3                    4                    5                    6                    7

Comments:

2. How easy was it to learn the techniques for using the three tools?

Volume Cracker (1 being very hard to learn, 7 being very easy to learn):

•-----•-----•-----•-----•-----•-----•-----•

1                    2                    3                    4                    5                    6                    7

Unconstrained Slicing (1 being very hard to learn, 7 being very easy to learn):

•-----•-----•-----•-----•-----•-----•-----•

1                    2                    3                    4                    5                    6                    7

Axis-Aligned Slicing (1 being very hard to learn, 7 being very easy to learn):

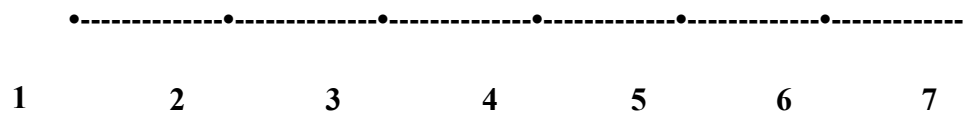


Comments:

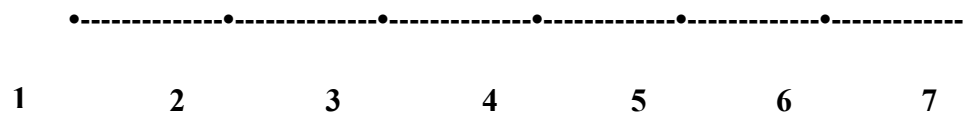
3. How much did you use and/or prefer to use the cracking preview in the Volume Cracker to look for objects inside the dataset?

4. How much did you like each of the three techniques?

Volume Cracker (1 being liked the least, 7 being liked the most):



Unconstrained Slicing (1 being liked the least, 7 being liked the most):



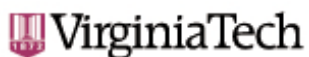
Axis-Aligned Slicing (1 being liked the least, 7 being liked the most):



Comments:

5. What were your main strategies while using the volume cracker?
6. What did you think the volume cracker was missing? For which task(s) did you think so?
7. What did you wish worked differently in the volume cracker? For which task(s)?
8. Please feel free to add any comments, suggestions, and concerns.

## Appendix E. Experiment Documents: Bare-Hand Volume Cracker Study



**Office of Research Compliance**  
Institutional Review Board  
North End Center, Suite 4120, Virginia Tech  
300 Turner Street NW  
Blacksburg, Virginia 24061  
540/231-4606 Fax 540/231-0959  
email [irb@vt.edu](mailto:irb@vt.edu)  
website <http://www.irb.vt.edu>

### MEMORANDUM

**DATE:** November 6, 2013  
**TO:** Doug A Bowman, Bireswar Laha  
**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires April 25, 2018)  
**PROTOCOL TITLE:** Evaluation of a free-hand gesture based interaction technique for analyzing micro-CT data  
**IRB NUMBER:** 13-1002

Effective November 6, 2013, the Virginia Tech Institution Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

### PROTOCOL INFORMATION:

Approved As: **Expedited, under 45 CFR 46.110 category(ies) 7**  
Protocol Approval Date: **November 6, 2013**  
Protocol Expiration Date: **November 5, 2014**  
Continuing Review Due Date\*: **October 22, 2014**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*

Date*	OSP Number	Sponsor	Grant Comparison Conducted?
11/04/2013	13131412	National Science Foundation	Compared on 11/06/2013

\* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

## **Informed Consent for Participant of Investigative Project**

Virginia Polytechnic Institute and State University

Title of Project: **Evaluation of a free-hand gesture based interaction technique for analyzing micro-CT data**

Investigators: Dr. Doug Bowman, Bireswar Laha.

### **I. THE PURPOSE OF THIS RESEARCH/PROJECT**

You are invited to participate in a study to compare a 3D interaction tool, called the *Volume Cracker*, to two other standard interaction techniques used for analyzing volume datasets.

### **II. PROCEDURES**

Upon arrival, you will be asked to read and sign this informed consent form. After that you will be given an introduction of our experiment background, facilities to be used, and study procedures. You will then be introduced to the display system we would use, which would be either a regular desktop computer workstation in the Socha lab of Norris Hall or the Vizwall in the Blacklab, at CRC.

For interaction with the hardware, we shall use the Leap Motion device (<http://leapmotion.com/>) for tracking free-hand gestures of the participants. We might use a mouse during the axis aligned slicing. You will go through a training session, during which you will get training on the Volume Cracker technique, the arbitrary slicing technique, and the axis-aligned technique as shown here - <http://www.youtube.com/watch?v=6YIMKyQP0Fw&feature=youtu.be>. In this study we shall use the leap motion device instead of the tracked 5DT gloves shown in the video. We shall also use real datasets in this study, unlike the simulated datasets shown in the video.

The experimenter will record your responses to questions involving analysis of volume datasets with semi-transparent or x-ray rendering ([http://en.wikipedia.org/wiki/Volume\\_rendering](http://en.wikipedia.org/wiki/Volume_rendering)). The responses will be evaluated offline by the experimenter, who will come up with a score as a performance measure for the study. The experimenter will also record the time taken to perform the tasks during the study.

After the tasks are completed, you will fill out a post-questionnaire. A free form interview will be conducted if you have any additional comments not addressed by the questionnaire.

The entire experiment will last about an hour, with a maximum of one hour 15 minutes.

### **III. RISKS**

There will not be more than minimal risks involved in our study.

### **IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us understand the benefits of using the *Volume Cracker* for analysis of 3D volume visualized data, as opposed to traditional interaction techniques. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

### **V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.



## VI. COMPENSATION

Your participation is voluntary and unpaid.

## VII. FREEDOM TO WITHDRAW

You are free to withdraw from this study at any time for any reason.

## VIII. APPROVAL OF RESEARCH

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: 13-1002.

## IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (please print)

\_\_\_\_\_  
Contact: phone or address or

\_\_\_\_\_  
Email address (OPTIONAL)

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Computer Science Department (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
Graduate student, Computer Science Department  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

Participant #:

## Background Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one):      Male                  Female

Age: \_\_\_\_\_

Do you wear glasses or contact lenses (circle one)?

No      Glasses      Contact Lenses

Are you (circle one)

Right-handed      Left-handed      Ambidextrous

Occupation (if student, indicate graduate or undergraduate):

\_\_\_\_\_

Major / Area of specialization (if student): \_\_\_\_\_

Rate your tiredness level today: (circle one)

•-----•-----•-----•  
very tired                  somewhat tired                  a little tired                  not tired at all

Rate your expertise with computers: (circle one)

.....  
beginner                      amateur                      intermediate                      advanced

How often do you use computers...

...for work? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

...for fun? (circle the best answer)

- a. not at all
- b. once a month
- c. once a week
- d. several times a week
- e. daily

Rate your experience with analyzing volume data:

.....  
never                      amateur                      intermediate                      expert

Comments:

Rate your experience with analyzing volume data from biomechanics:

.....  
never                      amateur                      intermediate                      expert

Comments:

Rate your experience with the technique:

.....

never

amateur

intermediate

expert

Comments:

Rate your experience with video games, or 3D movies at theater or home:

•-----•-----•-----•  
never sometimes often everyday

Comments:

Rate your experience with Nintendo Wiimote, Sony Move, or Microsoft Kinect:

•-----•-----•-----•  
never sometimes often everyday

Comments:

Rate your experience with using 3D interaction for analyzing volume data:

•-----•-----•-----•  
never amateur intermediate expert

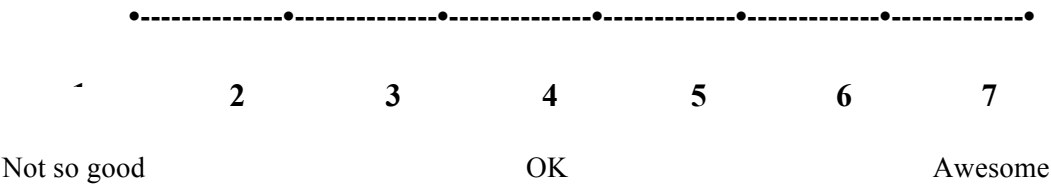
Comments:

Participant #:  
—

## Post-Experiment Questionnaire

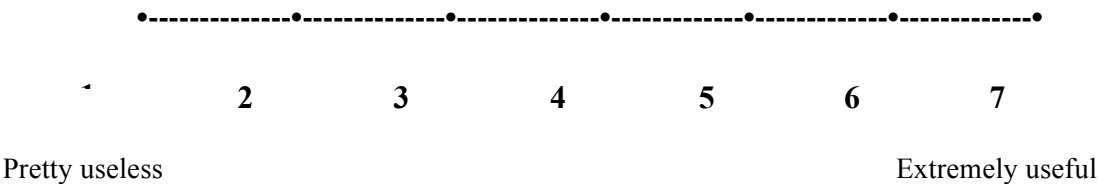
Please complete the following questions.

1. How would you rate your overall experience with (or your liking of) the interaction technique that you used?



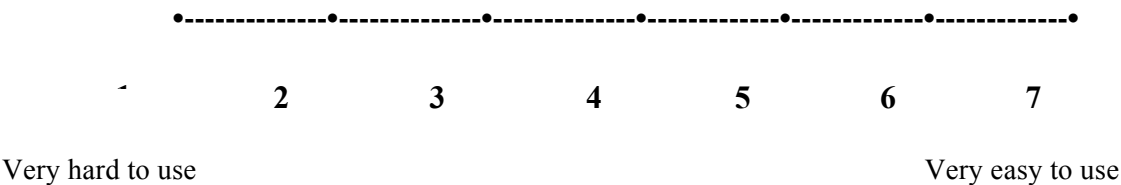
Comments:

2. How useful do you think the interaction technique is for analyzing tracheal tubes in beetles?



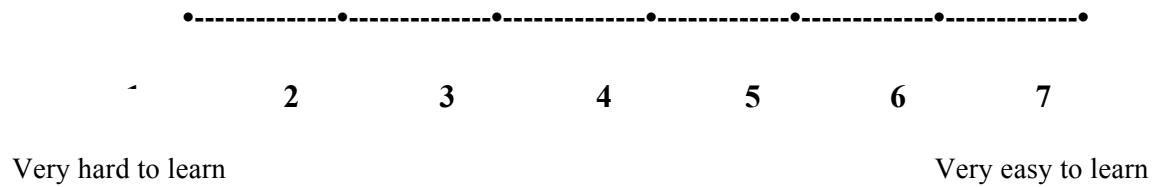
Comments:

3. How easy to use was the technique that you were using?



Comments:

4. How easy was it to learn the technique that you used?



Comments:

5. Have you used this technique, or some modified version of it before (encircle one)?

YES

NO

6. What were your main strategies (to complete a task) while using the technique?

7. What did you like and what you didn't like about the interaction technique?

8. What did you think the technique was missing? For which task(s) did you think so?

9. What did you wish worked differently in the interaction technique? For which task(s)?

10. Please feel free to add any comments, suggestions, and concerns.

## Bare-Hand Volume Cracker Evaluation Study Tasks

### Training Tasks

T1. This task is for your training in 3D rotations with the 3D mouse pad. Without translating or moving the dataset in any direction,

- a. Please complete one full rotation about the x-axis
- b. Please complete one full rotation about the y-axis
- c. Please complete one full rotation about the z-axis

T2. This task is for your training with translations using the 3D mouse pad. Without rotating the dataset,

- a. Please move the dataset right and left
- b. Please move the dataset up and down
- c. [VC or AS only] Please move the dataset forward and backward
- d. [AAS only] Please zoom into and zoom out from the data

T3. This task is for you to acquire a desired viewpoint around the dataset, using the skills you just learned in the previous two training tasks.

- a. Get a viewpoint from the top-left corner (from the default position).
- b. Get a viewpoint from the bottom-right corner (from the default position).
- c. Get a viewpoint from the top-rear (from the default position).

The next training task will allow you to get used to the Leap Motion controller.

[AAS only]

T4. Isolate one fourth of the dataset, using the following steps:

- a. Select the top-right quadrant. Slice the dataset approximately to its middle.
- b. Select the bottom-left quadrant. Slice the dataset approximately to its middle.
- c. Select the bottom-right quadrant. Slice the dataset approximately to its middle.
- d. View the remaining dataset from at least four different viewpoints, using the 3D mouse pad.
- e. Switch the planes along each of the axes, and see how the view changes in the corresponding quadrant.

[AAS only]

[AS only]



T4. Use the slice plane to slice halfway through the dataset from any direction.

- a. Freeze the plane using the shift key on the 3D mouse pad.
- b. Rotate/translate the dataset using the 3D mouse pad, to view it from at least four different viewpoints.
- c. Unfreeze the slice plane.

[AS only]

[VC only]

T4. Crack the dataset in roughly two equal halves, in any direction.

- a. Use the turn-the-pages feature to have approximately a  $1/4^{\text{th}}$  -  $3/4^{\text{th}}$  division between the halves.
- b. Rotate/translate the dataset using the 3D mouse pad, to view it from at least four different viewpoints.
- c. Crack one of the sub-volumes.
- d. Hide any of the sub-volumes. Unhide it.
- e. Join back the two cracked halves.

[VC only]

T5. Practice with the mode changes based on key presses in the keyboard

- a. View the dataset in the transparent mode by pressing the X key in the keyboard.
- b. View it back in the opaque mode by pressing the X key again.
- c. Reset the dataset using the R key in the keyboard.

## **Main Study Tasks**

Q1. Part-C: Are there tubes in the exoskeleton at the top of the animal? If yes, show all the ones that you find.

Q2. Part-A: The major tubes are the largest ones, and they're roughly equal in diameter. Find all the major tubes in the head. Is there an equal number of major tubes in the left vs. the right half of the head?

Q3. Part-C: For the largest tube in the right bottom leg, is the cross-section circular or elliptical?

Q4. Part-A: What is the cross-sectional shape of the main tube in the right-bottom leg (in the initial position)? Note: Point out that the cut is non-perpendicular.

Q5. Part-B: There are four major vessels running through the body. The bottom vessels branch into how many tubes as they go towards the head?

Q6. Part-B: In the first major segment of the right front leg (point out), how many tubes are there in the middle of the segment?

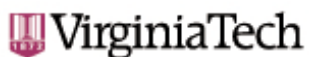
Q7. Part-C: Is there a tube that connects the left side to the right side of the body? If yes, point to a connection. Can you find another connection?

Q8. Part-D: Identify as many spiracles as you can find. Point at them and count the total number that you found.

Q9. Part-D: Are there 2 tubes connected to the manifold of this spiracle (near the cavity of the exoskeleton) or are there more than two tubes? Please point at them.

Q10. Part-E: Looking from the back of the animal, select the middle spiracle on the right. How many tubes connect to its manifold? In other words, how many tubes feed into the spiracle?

## Appendix F. Experiment Documents: Task Taxonomy Survey



Office of Research Compliance  
Institutional Review Board  
2000 Kraft Drive, Suite 2000 (0497)  
Blacksburg, VA 24060  
540/231-4606 Fax 540/231-0959  
email [irb@vt.edu](mailto:irb@vt.edu)  
website <http://www.irb.vt.edu>

### MEMORANDUM

**DATE:** March 13, 2013  
**TO:** Doug A Bowman, Bireswar Laha  
**FROM:** Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)  
**PROTOCOL TITLE:** Survey of Volume Analysis Tasks  
**IRB NUMBER:** 13-138

Effective March 12, 2013, the Virginia Tech Institution Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

<http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

### PROTOCOL INFORMATION:

Approved As: **Exempt, under 45 CFR 46.110 category(ies) 2**  
Protocol Approval Date: **March 12, 2013**  
Protocol Expiration Date: **N/A**  
Continuing Review Due Date\*: **N/A**

\*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

### FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

*Invent the Future*

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY  
*An equal opportunity, affirmative action institution*

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

\* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

**Informed Consent for Participant of Investigative Project**  
Virginia Polytechnic Institute and State University

Title of Project: Volume Analysis Task Taxonomy Survey

Investigators: Dr. Doug Bowman, Bireswar Laha.

**I. THE PURPOSE OF THIS RESEARCH/PROJECT**

You are invited to complete a survey for creating comprehensive and abstract categorizations of the tasks performed by scientists working with volume datasets from different scientific domains.

**II. PROCEDURES**

You are requested to complete the survey, filling in the responses to each question to the best of your knowledge. The average time of completion of this survey is estimated to be around 20 minutes.

**III. RISKS**

There will not be more than minimal risks involved in our study.

**IV. BENEFITS OF THIS PROJECT**

Your participation in this project will provide information that will help us to quantify the effects of higher levels of display fidelity for analysis of 3D visualized volume data. No guarantee of benefits has been made to encourage you to participate. You may receive a synopsis summarizing this research when completed.

**V. EXTENT OF ANONYMITY AND CONFIDENTIALITY**

The results of this study will be kept confidential. The information you provide will have your name removed and only a participant number will identify you during analyses and any written reports of the research.

**VI. COMPENSATION**

As a token of appreciation of your participation in the survey, on completing the survey, you will be entered into a drawing to receive one of ten \$20 gift cards from [amazon.com](https://www.amazon.com). No gift is guaranteed.

**VII. FREEDOM TO WITHDRAW**

You are free to withdraw from this study at any time for any reason.

**VIII. APPROVAL OF RESEARCH**

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.  
Virginia Tech IRB Approval number: 13-138

**IX. SUBJECT'S RESPONSIBILITIES AND PERMISSION**

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

Should I have any questions about this research or its conduct, I may contact:

Investigators: Dr. Doug A. Bowman Phone (540) 231-2058  
Professor, Department of Computer Science (231-6931)  
email: [bowman@vt.edu](mailto:bowman@vt.edu)

Bireswar Laha.  
PhD Candidate, Department of Computer Science  
email: [blaha@vt.edu](mailto:blaha@vt.edu)

Review Board: Dr. David Moore Phone (540) 231-4991  
Chair, Virginia Tech Institutional Review Board  
For the Protection of Human Subjects  
Email: [moored@vt.edu](mailto:moored@vt.edu)

cc: the participant, Dr. Bowman, Bireswar Laha.

The survey from Qualtrics:

#### Default Question Block

##### Introduction

Thank you for taking part in this survey! We are interested in understanding the tasks that users of volume visualizations perform when analyzing their datasets. Based on our experiences and interviews, we have developed an initial draft of a set of tasks. In this survey, we want to see how closely this list of tasks matches with your experiences. Your responses will be used to inform our research on the use of advanced displays and interaction techniques for volume data analysis. Your responses will be kept anonymous, but we may use quotations from your responses in our publications. Primarily, the responses will only be reported in aggregate form. If you have any questions about the survey, or wish to know more about our research, please contact Bireswar Laha ([blaha@vt.edu](mailto:blaha@vt.edu)).

We expect the survey to take you approximately 20 minutes.

##### Background Information

This survey is approved by the Institutional Review Board (IRB) of Virginia Tech. Please read the informed consent here - [Informedconsent](#)

Please select one of the following options to continue.

- ☐ I am more than 18 years old. I have read the informed consent and want to take the survey.
- ☐ I want to exit the survey now.

##### Background Information

Gender

- ☐ Male
- ☐ Female

How old are you?

Would you classify yourself as

- ☐ A researcher/scientist/medical practitioner/radiologist who works with volume datasets directly
- ☐ A professional imaging analyst - someone working in an imaging lab/institute/organization, who scans/analyzes volume datasets, as a service to researchers/doctors/faculty/others
- ☐ Other

In which scientific domain do you categorize yourself?

- ☐ Geophysics/geology
- ☐ Medical biology
- ☐ Paleontology
- ☐ Biomechanics
- ☐ Cell biology
- ☐ Engineering
- ☐ Archaeology
- ☐ Medicine
- ☐ Radiology
- ☐ Architecture
- ☐ Animation and 3D modeling
- ☐ Other

Please give a one-line description of the type of data you typically look at in your domain, and why (e.g., I look at micro-CT scans of fossils, to analyze their internal structures, and understand whether they were living or non-living):



How many years of experience do you have analyzing volume datasets?

- ☐ < 1 year
- ☐ 2 - 5 years
- ☐ 5 - 10 years
- ☐ more than 10 years

How many different volume datasets have you worked with so far?

- ☐ 1
- ☐ 2 - 5
- ☐ 5 - 10
- ☐ 10 - 20
- ☐ more than 20

By *segmentation* we mean the process of marking or specifying parts of the dataset as belonging to particular structures or features. Do you segment your datasets?

- ☐ Yes
- ☐ No

Select the methods you use for segmentation:

- ☐ Automatic
- ☐ Manual
- ☐ Some combination of Automatic and Manual

What percentage of the volume datasets that you analyze falls in the following categories? Please add the percentage beside each category below (e.g., 5% Time-varying, 95% Static):

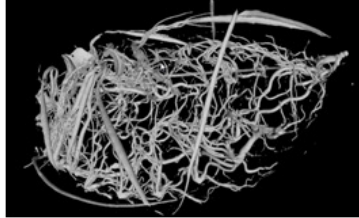
- ☐ Time varying
- ☐ Static or non-time varying

What percentage of the volume datasets that you analyze falls in the following categories? Please add the percentage beside each category below (e.g., 90% Single-Valued, 10% Multi-Valued):

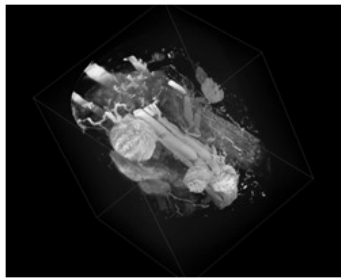
- ☐ Single-valued - the voxels in the volume represent just one material property, through a single scalar valued function
- ☐ Multi-valued - the voxels in the volume represent, and map multiple material properties, through a multi-valued function

In what form do you mostly analyze your datasets? Please add the percentage beside each category below:

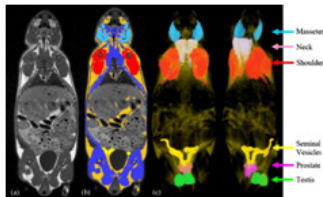
- ☐ 3D rendering of segmented data, such as isosurface rendering, based on a chosen density value



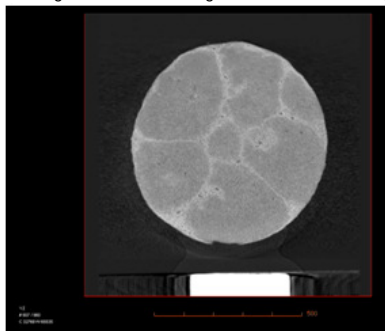
- ☐ 3D rendering of unsegmented data - 3D texture, ray-casting or maximum intensity projection, x-ray rendering, semi-transparency rendering, or a similar technique, using some transfer function



- ☐ 2D image slices of segmented data



- ☐ 2D image slices of raw unsegmented data - with axial, sagittal, and/or coronal views



- ☐ Other

Please name all the different software that you currently use to analyze volume datasets:

Does your software show

- ☐ just 2D image slices
- ☐ 3D perspective rendering of the volume
- ☐ both
- ☐ other

Do you mostly view the datasets from the inside the dataset, by navigating or flying through/within it (egocentric), or view them mostly by spinning it around and looking from various angles from the outside (exocentric)?

- ☐ Egocentric
- ☐ Exocentric

### Instructions for the remainder of the survey

You will be shown six abstract task groups one after another, along with their sub-categories. To give you a clear understanding of the type of tasks in each group, you will be shown two to three examples from medical biology, paleontology, and/or biomechanics domains. All of the task groups may or may not make sense in your scientific domain. For each task group, you will be asked a few questions, before moving on to the next group.

### Search Tasks

Definition: We define a *search task* as one in which the user is searching for a feature in the dataset. Our current list has two types of *search* tasks:

#	Sub-Categories	Example from medical biology	Example from paleontology	Example from biomechanics

1	Searching for presence/absence of a particular feature in a dataset	Search for distinct bone segments in this sample.	Within each cell there may or may not be intracellular bodies. In how many of the cells can you identify intracellular bodies?	Does the insect possess air sacs? Search for them.
2	Counting objects/features, given they are present	Count the number of fibers that you see in the dataset.	The fossil consists of multiple bounded compartments, each of which is potentially a cell. You can see these compartments forming in 3D as you use a cutting plane to slice through the dataset in a single direction. Count the number of compartments that you can identify.	Count the total number of tracheal tubes and categorize them by diameter

How much do you agree with the statement, "Some of the tasks that I perform in my research activities involve *search* tasks like those described above"?

highly disagree      2      3      neither agree nor disagree      5      6      highly agree

☐      ☐      ☐      ☐      ☐      ☐      ☐

Please provide at least one example of a *search* task from your domain, similar to the examples above.

Please give one task from your domain that you think is most similar to *search* tasks.

### Pattern Recognition Tasks

Definition: We can define a *pattern* as a characteristic that is repeated or a trend that we see as we look through the dataset. This task category includes all tasks for recognizing *patterns*.

#	Sub-Categories	Example from medical biology	Example from paleontology	Example from biomechanics
1	Looking for a trend	Blood vessels on one side of the dataset are skinny and they get progressively bigger as we go through the dataset. Identify on which side of the dataset we notice this trend.		Are the majority of tracheal tubes directly adjacent to muscles?
2	Looking for repetition of a pattern in a dataset	Looking through the sample longitudinally, count the number of fibers curving through the sample, and the number of straight fibers.	For the position of the intracellular bodies within the cells, choose one from below: a. Always floating within the cell not attached to boundaries. b. Attached to the outer surface of the fossil structure. c. Contained within cell boundaries only, no part of the intracellular structure juts into the interior of the cell. d. When close to the cell boundaries, they always merge with the boundary. Otherwise, they are completely detached and floating within the cell. e. Always touching cell boundaries.	

How much do you agree with the statement, "Some of the tasks that I perform in my research activities involve *pattern recognition* tasks like those described above"?

highly disagree      2      3      neither agree nor disagree      5      6      highly agree

☐      ☐      ☐      ☐      ☐      ☐      ☐

Please provide at least one example of a *pattern recognition* task from your domain, similar to the examples above.

Please give one task from your domain that you think is most similar to *pattern recognition* tasks.

### Spatial Judgment Tasks

Definition: These are tasks where the user makes a judgment in a 3D spatial context about the position and/or orientation of a feature on an absolute or relative sense.

#	Sub-Categories	Example from medical biology	Example from paleontology	Example from biomechanics
1	Absolute – Locating a feature/object which is the highest, lowest, farthest, or closest to/from a viewpoint, or in the overall dataset	From this point of view, which blood vessel extends the farthest inside these chambers?		Which sets of tubes are located closest to the front of the dataset?

2	Relative – Judging whether a feature/object 'A' is higher/lower/farther/closer than another feature/object 'B'	Which of the twisting fibers is the closest to the center of the bundle?		Which tracheal tubes are adjacent to the gut, and how close are they? Are they attached?
3	Judging whether two features/objects intersect or if they are distinct	Are the three blood vessels visible from the top distinct or do they branch out from just one blood vessel?		Does the left side of the tracheal system connect to the right side, throughout the body?

How much do you agree with the statement, "Some of the tasks that I perform in my research activities involve *spatial judgment* tasks like those described above"?

highly disagree      2      3      neither agree nor disagree      5      6      highly agree

☐      ☐      ☐      ☐      ☐      ☐      ☐

Please provide at least one example of a *spatial judgment* task from your domain, similar to the examples above.



Please give one task from your domain that you think is most similar to *spatial judgment* tasks.

### Quantitative Estimation Tasks

Definition: These include tasks in which the user has to estimate some property of the dataset, or part of it, *quantitatively*. Examples of things we can estimate in a volume dataset are density, size, volume, distance, length, surface-area, curvature, angle, velocity and other super-scalar quantities, etc.

Our current list has three types of *Quantitative Estimation* tasks:

#	Sub-Categories	Example from medical biology	Example from paleontology	Example from biomechanics
1	Absolute estimation	Which is the density of the highlighted region of the bone marrow?		What is the density (or volume fraction) of tracheal tubes in a particular region of the body? What is the density throughout the entire body?
2	Relative estimation (binary)	Which parts of the blood vessel has flow faster than that at the center?	Study the thickness of the borders and the joints between the cells. Are different joints and borders of comparable or variable thickness?	
3	Relative estimation (quantitative)		How many times denser is the boundary of the cells at the center than the outer boundary of the cluster?	How many times bigger is the biggest tracheal tube than the smallest one? Your choices are 5, 15, 30, 50, and 60.

How much do you agree with the statement, "Some of the tasks that I perform in my research activities involve *quantitative estimation* tasks like those described above"?

highly disagree      2      3      neither agree nor disagree      5      6      highly agree

☐      ☐      ☐      ☐      ☐      ☐      ☐

Please provide at least one example of a *quantitative estimation* task from your domain, similar to the examples above.

Please give one task from your domain that you think is most similar to *quantitative estimation* tasks.

### Shape Description Tasks

Definition: These include tasks where the user describes the shape of either the whole or some part of the dataset qualitatively.

#	Sub-Categories	Example from medical biology	Example from paleontology	Example from biomechanics
1	Describing the overall shape of a sub-part/sub-section qualitatively	Please explain the alignment of the material in the major bundles (e.g., are they arranged in concentric	Describe the size and shape of the intercellular bodies, using these options: a. Cubic, solid, and variable in size.	What is the cross-sectional shape of each tracheal tube? Shapes tend to run from circular to flattened ellipses.

		circles/spirals?).	b. Pill-shaped (spherocylindrical), solid, and variable in size. c. Spherical or moon-shaped, typically hollow rather than solid, and comparable in size. d. Spherical or moon-shaped, typically solid rather than hollow, and comparable in size. e. Pill-shaped, typically hollow rather than solid, and comparable in size.	How does cross-sectional shape vary with size?
--	--	--------------------	---	--

How much do you agree with the statement, "Some of the tasks that I perform in my research activities involve *shape description* tasks like those described above"?

highly disagree

2

3

neither agree nor disagree

5

6

highly agree

Please provide at least one example of a *shape description* task from your domain, similar to the examples above.

Please give one task from your domain that you think is most similar to *shape description* tasks.

### Post Questionnaire

What percentage of your time analyzing volume data do you think is spent doing ..

- ☐ Search
- ☐ Pattern Recognition
- ☐ Spatial Judgment
- ☐ Quantitative Estimation
- ☐ Shape Description

Do you think there are tasks that you perform in your research that do not fall in any of the task categories discussed above?

- ☐ Yes
- ☐ No

Please give a list of tasks that you carry out with volume datasets, that didn't fit in any of the above categories.

For each of the tasks you mentioned in the previous question, what abstract group(s) would you put those tasks in?

How much do you agree with this statement: "The list of tasks for volume data analysis described in this survey is how I would describe or categorize the tasks that I carry out with volume datasets for my work"?

highly disagree

2

3

neither agree nor disagree

5

6

highly agree

Please tell us why you thought so.

Any other comments?

If you would like to be entered into the drawing for one of ten \$20 gift cards, please enter your email address below. You do not need to enter your email address if you do not wish to take part in the drawing.  
We will not use your email address for any purpose other than to notify you if you have been selected to receive a gift card.