Stereo and Motion Cues Effect on Depth Perception in Volumetric Data

(Technical Report: CVC-UNCC-13-12,
http://viscenter.uncc.edu/publications)

Isaac Cho, Zachary Wartell, Wenwen Dou, Xiaoyu Wang and William Ribarsky

Stereo and Motion Cues Effect on Depth Perception of Volumetric Data

Isaac Cho, Zachary Wartell, Wenwen Dou, Xiaoyu Wang and William Ribarsky

Charlotte Visualization Center, University of North Carolina at Charlotte, Charlotte, NC, USA 28213-0001

ABSTRACT

Displays supporting stereoscopic and head-coupled motion parallax can enhance human perception of containing 3D surfaces and 3D networks but less for so volumetric data. Volumetric data is characterized by a heavy presence of transparency, occlusion and highly ambiguous spatial structure. There are many different rendering and visualization algorithms and interactive techniques that enhance perception of volume data and these techniques' effectiveness have been evaluated. However, how VR display technologies affect perception of volume data is less well studied. Therefore, we conduct two formal experiments on how various display conditions affect a participant's depth perception accuracy of a volumetric dataset. Our results show effects of VR displays for human depth perception accuracy for volumetric data. We discuss the implications of these finding for designing volumetric data visualization tools that use VR displays. In addition, we compare our result to previous works on 3D networks and discuss possible reasons for and implications of the different results.

Keywords: stereoscopic display, head-tracking, depth perception, virtual reality, volumetric data, depth cues

1. INTRODUCTION

Previous research has demonstrated the utility of computer displays that provide stereoscopy and head-coupled motion parallax for enhancing human perception of complex three-dimensional (3D) datasets. This includes fully immersive displays such the Cave Automatic Virtual Environment (CAVE) and Head Mounted Display (HMD) and semi-immersive displays such as desktop VR and the virtual workbench [1]. For example, studies by Ware et al. examine the effect of stereoscopic display and kinetic depth on understanding 3D networks which are represented by tubes or lines. User performance at finding paths in a complex 3D network improves when using stereoscopy and structure-from-motion [2,3,4].

A significant number of visual analytic domains, however, also heavily use 3D volumetric data. Examples include medical imaging, weather and environmental simulations, and fluid flow. Volumetric data is characterized by large amounts of transparency, occlusion and ambiguous spatial structure. There has been a fair amount of evaluation of perception of volumetric data under different rendering algorithms and different parameterizations such as modifying transfer functions. For a simple volume dataset one might find a combination of transfer function settings, rendering algorithm and rendering augmentations (such as edge enhancement) for which adding VR type display capabilities does little to further improve depth and shape perception. However, for more complex volumetric dataset, even with perceptually optimized transfer functions and rendering parameters, VR display capabilities may further improve depth and shape perceptions.

For surface and 3D networks, this type of result is typical. When the 3D geometry is relatively simple, well-chosen rendering parameters can maximize shape perception causing VR display technology to have less additional benefit. However, with more realistic complex geometry, VR technology shows significant positive effects on shape and depth perception.

One would expect similar results for volumetric data. Further, the addition of VR display technology could be especially important with time-varying volumetric dataset that are viewed in real-time where extensive preprocessing for optimized transfer functions and volume rendering parameters is not possible. An example would be real-time, streaming Doppler weather radar data [5]. With the increasing affordability of semi-immersive VR displays and GPUs capable of advanced volume rendering, there is a need to quantify the effectiveness of stereopsis and structure-from-motion on volumetric data and also to quantify how these display parameters interact with other volumetric rendering conditions. The large number of display hardware and rendering algorithm variations make such evaluations challenging. In this paper, we

take a first step towards this by using a fixed set of volume rendering parameters-chosen through pilot studies-while varying the VR display hardware conditions.

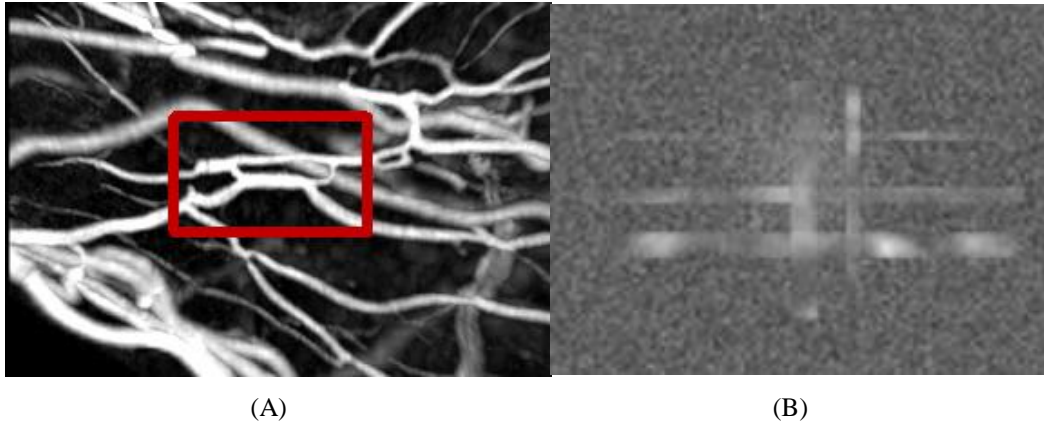(A)   (B)

Figure 1: Similarity comparison between our artificial dataset and actual MRI blood vessel scan. (A) Maximum Intensity Projection rendering of blood vessels [6]. (B) Our artificial dataset.

We present results of two experiments on the benefits of stereoscopy and head-tracking for a person's correct perception of depth ordering of volumetric objects. The experiment is motivated by dataset such as the MRI scan of blood vessels shown in Figure 1111A reproduced from [6]. As is typical of volumetric data, this dataset is has a heavy presence of transparency, occlusion and a highly ambiguous spatial structure. In Figure 1111A, it is particularly challenging to determine the depth order of the blood vessel inside the red square. As discussed in [7], the volume rendering technique used here makes it appear that the square-shaped loop vessel is in front of the diagonal one. However, in fact the diagonal one is in front of the square-shaped loop.

We mimic this type of ambiguity by generating controlled experimental dataset such as Figure 1111B where the user's task is to determine the depth ordering of various occluding transparent cylinders. The subjects view the datasets under a variety of display conditions including combinations of stereoscopic display, head-tracking, and small object rotations. We present a set of cylinders of various size, opacities and depth orderings to mimic datasets such Figure 1111A but in an experimentally controlled manner. To isolate the effect of semi-immersive display conditions the user is not allowed to alter that transfer function or other volumetric rendering parameters such as lighting or switching between rendering methods. The two experiments include a depth-ordering task, in which participants must understand the full depth ordering between six volumetric cylinders and a depth discrimination task, in which participants must distinguish the relative order of just two cylinders within a limited exposure time (2 sec). In addition, the experiments are also designed to differentiate the differences between experienced and less-experienced users with respect to experience with 3D games and VR related technology. Results from both groups show an overall benefit for stereoscopy with head-tracking in enhancing depth perception of volumetric data. In addition, we compare our finding with previous Ware et al.'s works [2,3,4] on non-volumetric 3D network data and discuss possible reasons of different results. More interestingly, our study also suggests that familiarity with 3D games and VR related technology significantly affects the user's depth perception accuracy.

2. PREVIOUS WORK

Display supporting stereoscopy and head-coupled motion parallax (such as head-tracking) can enhance human perception and understanding of complex 3D dataset. Structure-from-motion is sub-classified into motion parallax due to observer motion and the Kinect-depth-effect due to object rotation [4]. Computer graphic displays provide motion parallax by appropriate manipulation of the virtual camera parameters. One approach is for the user to control the camera using a hand-help spatial input device or tracked hand motion. Another is to use a display that when observed from different physical vantage points show the observer a different corresponding perspective view of the virtual objects. We refer to this as head-cou7pled motion parallax. A common approach is to employ head-tracking. Volumetric displays, full parallax holography or other multi-view-point displays can also achieve such head position dependent perspective without head-tracking. The perceptual benefits of head-coupled motion parallax provided by all these systems are

equivalent to those due to head-tracked systems. While our study uses a head-tracked system, our results concerning the benefits of head-coupled motion parallax generalize to these other non-tracked systems.

Prior experiments with surface and 3D network datasets show that stereoscopic display can aid depth perception when either the visual stimuli lacks other depth cues, as can occur in teleoperator environments and remote sensing or in less sophisticated computer graphic presentations, or when the visual stimuli contains a high depth complexity as measured by many occlusions [8]. As a specific recent example, Ware et al. examine the effect of stereoscopy and the kinetic depth effect on a person's understanding of a 3D network which is represented by 3D tubes or lines [2,3,4]. They demonstrate great benefit for stereoscopic and kinetic depth when a user must find paths in between nodes in a complex 3D network. Grossman et al. introduce some volumetric aspects, although the focus is on comparing types of stereoscopic and head-coupled motion parallax display technologies instead of volumetric rendering algorithms. They perform three experiments to evaluate the depth perception of 3D scenes with 3 different types of displays: a volumetric display, a stereoscopic display, a stereoscopic display with head-tracking [9]. The participants performed 3 tasks: ranking the depth of a sphere, tracing the path in a 3D network, and judging whether an object will collide with another object. Their results show the volumetric display has significantly better user performance than the others, but their results also show that stereoscopic display with head tracking is better than stereoscopy alone.

The most detailed evaluations of how to improve a user's perception and spatial understanding of volumetric data focus on comparing different rendering techniques and/or different transfer functions. A. Steward introduces a shading model named "vicinity shading" which provides shadowing to enhance the perception of surfaces within a volume [10]. Svakhine et al. introduce the 2D outline illustration technique which can be merged with a 3D volumetric medical dataset to improve the depth perception [11]. Bruckner et al. present a volumetric halo drawing technique to emphasize depth information of a volumetric medical dataset [12]. Boucheny et al. present perceptual studies that examine how transparency affects depth perception in Direct Volume Rendering (DVR) [13]. The participant determines the depth ordering of two cylinders with semi-transparency, two luminance conditions of cylinders (left cylinder dark or left cylinder bright) and 3 different background images matching to the intersection's brightness. Overall performance of participants was relatively weak, but above chance level.

A few authors have studied the effectiveness of stereoscopic display for volumetric data. Maciejewski et al. introduce the Interactive Volume System which implements stereoscopic and haptic rendering with interactive transfer function to provide visual and haptic feedback to users [7]. They evaluate depth cues and accuracy of the user's docking of two proteins using their system. The result shows the benefit of stereoscopy for finding the best docking configuration. However, the haptic transfer function did not provide any benefit. Hancock evaluates the depth perception of a volumetric dataset rendered using DVR on a stereoscopic display [14]. In these experiments, observers determine the relative depths of three smaller non-overlapping spheres which are either: presented directly; embedded in a large transparent in volumetric sphere; or embedded in a larger transparent sphere enhanced by modulating its opacity based on surface gradient. Observers viewed the objects using both a stereo and monoscopic display. Observers were most accurate under the transparent sphere condition while stereoscopic display improved accuracy only in the 'presented directly' condition. Overall, this indicated a dominance of aerial perspective over stereopsis for their scenes. Our experiment complements these works by focusing more on low-level occlusion and depth ordering perception with the addition of structure-from-motion as an independent variable.

Some work examines how stereoscopy and structure-from-motion interact to effect perception of volumetric data. Kersten et al. show the effectiveness of stereopsis and simulated aerial perspective for the depth perception of a volume dataset of cylinders. The data are digitally reconstructed radiographs (DRR) [15]. Their results show that the stereoscopic display and stimulated aerial perspective provide depth cue better than opacity and spatial frequency for recognizing the rotation direction of the cylinder. B. Mora et al. show the effectiveness of order-independent direct volume rendering (Maximum Intensity Projection (MIP) and X-ray projection rendering techniques) [6]. They showed advantages of stereoscopy and transfer functions for enhancing the depth perception of a volume object rendered by MIP and X-ray projection renderings that lack other spatial depth cues. While compelling, their observations regarding stereoscopy's effectiveness are anecdotal. Finally, Hubbold et al. present a technique named "tunneling" which allows users to see internal features and details of volumetric dataset rendered via DVR [16]. In their experiments, participants assess patterns of small blobs inside a volumetric brain. The experimental conditions include combinations of head-tracking, kinetic depth and stereoscopy. The results show that stereoscopy improves the users' reported understanding of the depth structure and the combination of head-tracking with stereo is most preferred. Our experiment complements this

work by using a more precise, quantitative measure—user reported depth ordering—rather than having the experimenters assign scores to the user's verbal descriptions of the perceived image.

3. ENVIRONMENT

Our study tests the effectiveness of a semi-immersive VR display on depth perception of volumetric data. We examine the effects of display environment on two tasks: a depth ordering task, in which participants determine the general depth ordering among six volumetric cylinders with no time limit; and a depth discrimination task, in which participants must distinguish the depths of a pair of cylinders within a short time limit (2 sec).

Figure 2: Environment system: This picture shows our environment system. A 3D display with Nvidia 3D vision shutter glasses, Polhemus tracker and a chin-rest.

As shown in Figure 2222, both tasks use a desktop VR setup which consists of a stereo display (Samsung Sync Master 2233RZ) and a tracked Nvidia 3D vision glasses. Head-tracking is available by a Polhemus Fastrak. The display rate is 120 Hz time-multiplexed to 60 Hz per eye. The subjects sit roughly 60 cm in front of the screen. The independent experimental variable is the display condition which a combination of stereoscopy, head-tracking and/or a small object rotation (±10º roll) for a kinetic-depth-effect.

We informally experimented with using various freely available volumetric dataset including an engine block, an orange, a tree and medical data. We considered having participants identify a variety of different shape characteristics. Base on pilot test for a variety of scenarios, we narrowed our experimental design to depth discrimination and ordering tasks. Further, we chose to create an artificially generated dataset loosely based on the types of depth ambiguities found in the blood vessel example (Figure 1111A).

These decisions are motivated as follows. First, there is a much larger range of rendering algorithms and rendering parameters for volumetric data than there are for surfaces and 3D networks. The choice of volume rendering technique or transfer function can make a huge difference to an observer's perceptual understanding of a volumetric dataset's spatial structure. This complicates creating a controlled experiment on shape or depth perception of volumetric data. Second, after testing various datasets with various volume rendering techniques and rendering parameters and after considering various perceptual questions, we concluded it was necessary to artificially generate a volumetric dataset where we could completely control these many factors and in order to present different volumetric datasets on each trial to avoid learning effects. At the same time, we wanted the created dataset to roughly mimic perceptual ambiguities found in a real-world volumetric dataset. Therefore, we developed a dataset that contains depth ambiguities between volumetric tubes inspired by 3D medical scans of networks of blood vessels.

Figure 3: This figure shows one of our artificial volumetric datasets. It has six cylinders (three horizontal and three vertical). Perlin Noise modulates the 3D texture. In the actual experiment, the voxels are far more transparent (see Figure 1111B).

As shown in Figure 3Figure 3, our synthetic volumetric dataset contains six overlapping cylinders of varying diameters and transparency. In this image, we decreased the transparency levels from that used in the experiment for expository purposes. Three cylinders are vertical and three and horizontal. The voxel resolution is $512 \times 256 \times 256$. Perlin noise is used for the internal texture of the cylinders and the texture of a large background polygon. In the experiment, an additional background polygon approximates the visual effect of having the cylinders embedded in a more complex volumetric environment.

To minimize a participant's knowledge gained from performing perceptual tasks on the same dataset over many repeat trials, the synthetic dataset present varies from trial to trial. Each cylinder in a trial is assigned a random depth location, a unique noise texture, a direction (vertical or horizontal), and a random cylinder size (thin, medium and large). A non-randomized dataset might allow a participant to consciously or non-consciously pick up and respond to coincidental relationships between these parameters such as "*the horizontal cylinders are always further away*" or "*the thinnest cylinder is always the one farthest away.*" Avoiding such conflating factors is a key reason we use an artificial dataset rather than a single real-world data set.

3.1 Rendering Techniques

We choose a high quality GPU-based ray-casting rendering techniques [17,18]. Compared to other rendering techniques, such as per pixel lighting and MIP rendering [13,19], the GPU based ray casting technique yields more accurate depth cues [18]. The renderer is available as an OpenSceneGraph plugin [20].

We add a black polygon with a square hole in front of the volumetric data to act as a window to hide the ends of the cylinders. This was necessary because being able to see the cylinder ends made the depth ordering task trivial. In real-world datasets such as the blood vessel example (Figure 1111A, the complex intertwined paths of the tubes typically tends to obscure the tube endpoints. We also tested scaling up the voxel volume's rendered size (to extend the cylinders' ends off screen) but this failed because aliasing artifacts were too visible. Attempting to increase the volumetric resolution beyond $512 \times 256 \times 256$ to counter aliasing problems exceeded the renderer's memory limitations.

Following a previous study [6], we fix the data parameters, such as the alpha gradient and transparency, to represent a reasonably clear outline of each semi-transparent cylinder (alpha=0.9, transparency=0.2 and density=0.025). Note that, to isolate the effects of stereo and structure-from-motion, we do not allow users to interactively adjust the transfer function in our study even though many previous studies demonstrated its utility in depth perception [21,22].

4. EXPERIMENT DESIGN

Our two experiments examine the effect of stereoscopy and/or head-tracking on the perception of volumetric data. Experiment 1 has four display conditions and Experiment 2 has six. We use a within-subject design with repeated measure. Each subject is randomly assigned a sequence of display conditions using Latin squares. The measures in our experiment are answering time and error rate. Before each experiment, the participant provides demographic information such as gender, academic major and degree being sought. A questionnaire inquired regarding their familiarity with

stereoscopic display, VR technology and gaming. Questions include: how often do you play games on a computer or game console with/without motion capture devices such as Xbox Kinect and Sony PlayStation Move; how often do you watch 3D movies in the theatre; how often do you sue 3D displays for movies or games. After the experiment, each participant fills out a post-questionnaire regarding their confidence in their answers to the task's spatial questions and their opinions on various visual aspects of the volumetric dataset such as transparency, the noise background, etc.

We recruited twenty-eight participants, twelve for Experiment 1 and sixteen for Experiment 2. Sixteen out of twenty eight are undergraduate students and twelve are graduate students. Fourteen participants' major in Computer Science and fourteen participants are of other majors including psychology, nursing history and fashion design. All participants have (corrected) 20/20 or higher vision. We provide a tutorial to familiarize the participants with the stereo display and head-tracking hardware. We designed two experiments. Experiment 1 examines the effect of stereoscopy and head-tracking on a depth discrimination task. In this task, subjects are exposed to the volumetric dataset for a short amount of time (2 sec) so that they do not have time to cognitively reason about the depth order based on factors such as transparency, window size, etc. (In many psychophysics studies the exposure time is usually in the range of a few hundred of milliseconds but 2s is common in stereoscopic VR studies [23]). Experiment 1 requires the subject to first locate an intersection of a pair of cylinders based on a provided instructional cue, and to then report on the depth relation of the cylinder pair. The 2 second exposure time allows for vergence eye movements [24]. Experiment 2 explores the effect of stereoscopy and head-tracking on the task of depth ordering which requires distinguishing the depth order of multiple cylinders, not just a single pair. Experiment 2 examines the displays' effect within the context of an unlimited exposure time.

We analyzed the results by a repeated measures ANOVA (analysis of variance).The reported F tests use $\alpha=.05$ for significance and indicate the Geisser-Greenhouse correction to protect against possible violation of the sphericity assumption. The post-hoc tests that were conducted were Fisher's least significant differences (LSD) pairwise comparisons with $\alpha=.05$ level for significance.

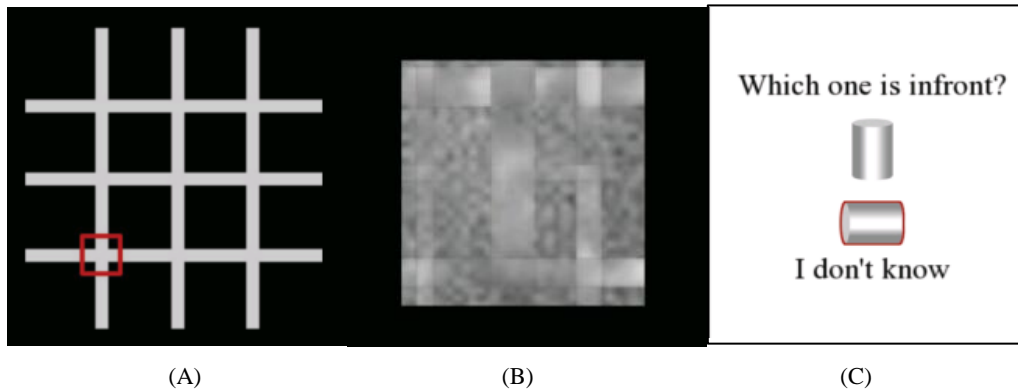## 5. EXPERIMENT 1: DEPTH DISCRIMINATION

(A) (B) (C)

Figure 4: Three screens displayed in Experiment 1's trial. (A) The instructional cue indicating the target cylinder pair to examine. (B) The volumetric cylinders seen through an aperture. (C) The question the participant answers for the trial.

Experiment 1 evaluates how stereoscopy and structure-from-motion affect performance on a depth discrimination task. The participant determines which of two cylinders, one horizontal and the other vertical, is in front of the other. The volumetric dataset contains six cylinders, but in each trial a pair of cylinders is designated as the target pair for the trial. On each trial, the first screen displays a 2D picture (Figure 4444A) where a red box designates which of the nine intersections of the six cylinders is the target pair. The next screen displays the volumetric dataset for 2 seconds (Figure 4444B). The final screen displays a menu with three choices (Figure 4444C): "*the horizontal cylinder is in front*", "*the vertical cylinder is in front*", or "*I don't know*". Note, we choose not to use a force-choice protocol in this experiment because we want to gather data on how often a user feels they cannot determine the depth ordering. A force-choice protocol would have conflated results for trials where participants were guessing at the depth order with those trials in which they felt they could determine a specific ordering.

Experiment 1 has six display conditions. The conditions are: non-stereo without no motion (NS-NM), stereo with no-motion (S-NM), non-stereo with head-tracking (NS-H), stereo with head-tracking (S-H), non-stereo with kinetic-depth

effect (NS-KD) and stereo with kinetic-depth effect (S-KD). The last two conditions were added because in pilot tests, not all users utilized the head-tracking when limited to the 2 second exposure time. In particular, some users did not attempt to use a quick head motion to gain motion parallax cues even when we were careful to specifically remind them this was possible. Hence, the kinetic-depth effect condition automatically rotates the cylinders left and right by 10 degrees. For a small range of motion the visual effect is similar to having the participant quickly move her head left and right. In the non-head-tracking conditions, a participant uses a chin rest. In this condition, the view frustums are calibrated for this fixed head position.

Each participant performed 324 trials in blocks of 54 where each block used one of the six display conditions. Display condition block order was counter-balanced using Latin squares.

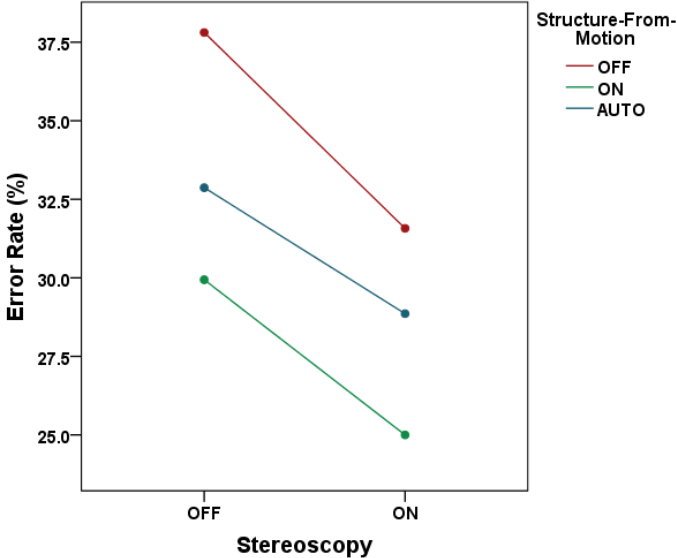## 5.1 Results – Anticipated Interactions

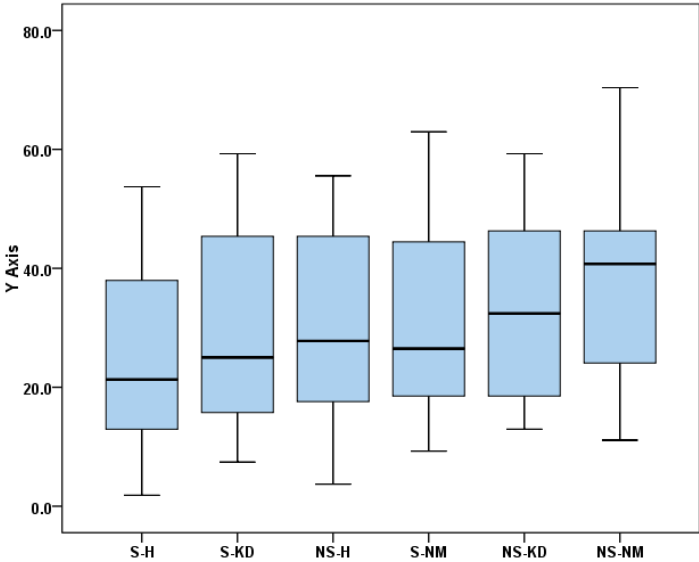Figure 5: Effect of stere and motion on error rate

Figure 6: Effect of display condition on error rate

Table 1: Mean and standard deviation of error rate for each condition

Display Condition	Error rate (%)	
	Mean	**SD**
S-H	25.00	15.82
S-KD	28.87	15.68
NS-H	29.94	16.12
S-NM	32.41	16.61
NS-KD	32.87	15.68
NS-NM	37.81	17.00

Error rate measures the percentage of incorrect depth judgments counting "*I don't know*" answers as incorrect. The two-way repeated measures ANOVA (($2 \times 3 \times S$)) shows no significant interaction between stereo and motion on error rate (p=.861, Figure 5555). Stereoscopy has a significant main effect on error rate (F(1,11)=8.691, p=.013, η_p^2=.441) decreasing the error rate from 33.54% (SD=4.12) to 28.76% (SD=4.52). However, there is no main effect on error rate of the motion condition (p=0.71).

The one-way repeated measures ANOVA (($6 \times S$)) shows a main effect of the general display condition (NS-NM, S-NM, NS-H, S-H, NS-KD, S-KD) has a significant effect on error rate (F(2.512,27.635)=3.549, p=.034, η_p^2=.244). Table 1 shows means and standard deviations of error rate for each condition and Figure 6666 illustrates mean error rate across all six conditions. LSD post-hoc tests show S-KD is significantly better than NS-NM (p=.012) and S-NM (p=.034). In addition, S-H is significantly better than NS-H (p=.038), S-NM (p=.038) and NS-NM (p=.009). S-NM is significantly better than NS-NM (p=.038).

5.2 Results – Unanticipated Interactions

We observed that some participants appeared more confident with their task performance and more comfortable with using our semi-immersive VR environment. Also, our participants came from two pools, a computer science (CS) pool all of whom were computer science majors, and a psychology (PSYC) pool, which were psychology and liberal arts majors. A one-way between-subjects ANOVA shows a main effect of pool: CS pool average error rate is 27% while PSYC pool average is 36% (p=.001).

Participants from the two pools were randomly assigned a display condition order and they participated over the same time period, therefore, we analyzed the pre-questionnaire. An example question is: "How often do you play 3D computer games?" Answers are on a 7 point scale with 1 being "Never" and 7 being "A Great Deal". On a number of these questions CS pool scored significantly higher on this scale for game playing experience. In particular some results were: 2D game playing (mean 4.8 vs. 2.3 at p<.001), gaming on a PC (mean 4.8 vs. 3.2 at p=.017), gaming on a console (mean 4.3 vs. 2.7 at p=.001), gaming with motion capture devices (mean 2.5 vs. 1.5 at p=.048), and stereoscopic 3D TV usage (2,2 vs. 1.0 at p=.05).We apply a mixed three-factor within subjects ANOVA to evaluate the effect of pool × motion × stereo ($2 \times (3 \times 2 \times S)$)) on error rate. There is no significant three-way interaction (F(2,20)=1.014). Regarding the 2-way interactions, stereo × pool is not significant. (The significant main effect of stereo is reported in sub-section *5-1*). However, motion × pool is significant (F(2, 20)=3.690, p =.043, η_p^2=.270).

The simple main effects for motion are as follows. For the PSYC pool motion is not significant but for the CS pool the main effect of motion is significant (p=.048). LSD pairwise comparisons for the CS pool shows head-tracking conditions are better than no-motion conditions, with average error rates 18.82% vs. 32.25% (p=.027).

Plausibly CS pool subject's greater experience with gaming trains a person to better attend to various depth cues when viewing computer generated 3D images and increases their sense of confidence in using VR type technologies. Alternatively, the CS majors might have simply been more interested in the technology employed and hence were somewhat more motivated. However, given that participants perform 324 trials, we suspect CS pools greater experience played a larger role than interest level.

In summary, for all participants stereo had a generally positive, significant main effect while only for the CS pool does the motion condition have a generally positive significant effect. In general, the CS pool participants performed better overall than PSYC pool participants.
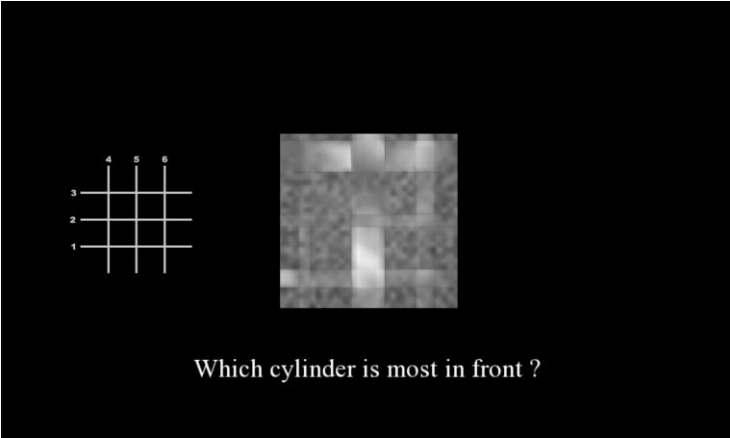
6. EXPERIMENT 2: DEPTH ORDERING

Figure 7: Single screen used in depth ordering experiment.

In Experiment 2, participants perform a depth ordering task on the six volumetric cylinders. Because the trial duration is unlimited and participants have ample time to use head-coupled motion parallax, the kinetic-depth-effect (e.g. auto-rotation) conditions are not included. The four display conditions are: non-stereo without head-tracking (NS-NH), stereo without head-tracking (S-NH), non-stereo with head-tracking (NS-H), and stereo with head-tracking (S-H). Each cylinder is labeled with a number (1-6). The participant must designate which of the six cylinders is at a particular position either: the front, the middle, or the back. The particular position queried is randomly determined per trial. (Two answers are counted as correct for 'middle'). For each trial, the cylinders are rendered with random depth ordering. Figure 777 shows the displayed screen.

The participant designates which cylinder is at the queried position by pressing the corresponding number key on the keyboard. In non-head-tracking conditions participants use a chin rest as in Experiment 1. Each participant undergoes 36 trials per display condition which means 144 trials total. Trials are in blocks by display condition and the block ordering uses Latin squares.
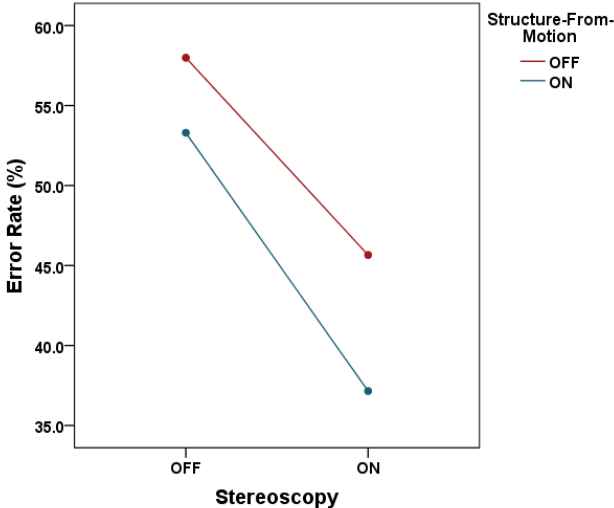
6.1 Result

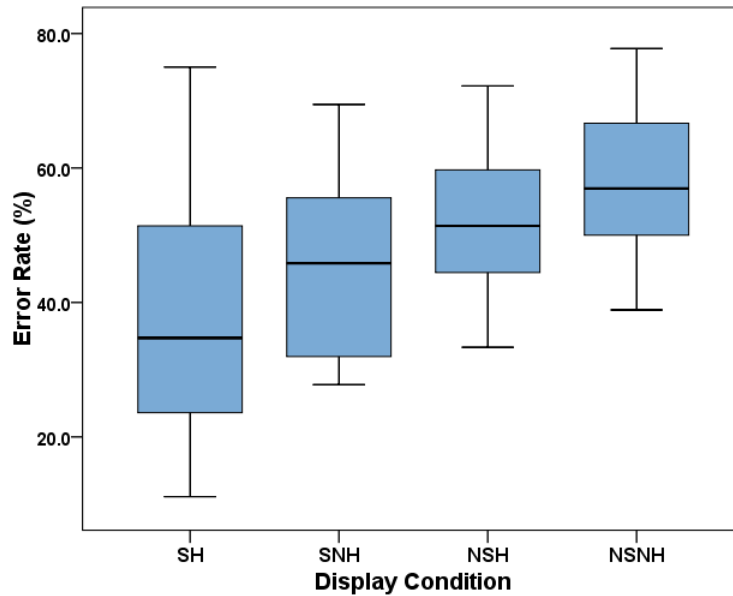Figure 8: Effect of stereo and motion on error rate

Figure 9: Effect of display condition on error rate

Table 2: Mean and standard devication (SD) of error rate and answering time

Display Condition	Error Rate (%)		Answering Time (sec)	
	Mean	SD	Mean	SD
S-H	37.33	19.51	13.89	2.46
S-NH	44.97	14.88	12.61	2.30
NS-H	53.30	11. 89	15.56	4.43
NS-NH	57.47	11.95	14.11	3.56

We analyze the effect of the display condition on answering time and error rate. Error rate is computed as the number of incorrect answers divided by total number of questions in each trial (36 questions per condition).

A $2 \times 3 \times 3$ repeated measures ANOVA shows no interaction effect on error rate among stereoscopic, head-tracking and question. Head-tracking has a significant main effect on error rate ($F(1,15)=5.372$, $p=.035$, $\eta_p^2=.264$). Stereoscopy has a significant main effect on error rate ($F(1,15)=19.471$, $p=.001$, $\eta_p^2=.565$).

A 2×2 repeated measures ANOVA (($2 \times 2 \times S$)) shows no interaction effect on error rate between stereoscopic and head-tracking display. Figure 8887 indicates the lack of interaction. Head-tracking has a significant main effect on error rate ($F(1,15)=5.372$, $p=.035$, $\eta_p^2=.264$) decreasing the error rate from 51.22% (SE=3.07) to 45.31% (SE=3.15). In addition, stereoscopy has a significant main effect on error rate ($F(1,15)=19.471$, $p=.001$, $\eta_p^2=.565$) decreasing the error rate from 55.38% (SE=2.07) to 41.15% (SE=4.13). There is no effect of display order condition and unlike in Experiment 1, no interactions with participant pool (CS vs. PSYC) are significant.

Table 222Table 2 shows mean and standard deviation of error rate and answering time for all four display conditions. The one-way repeated measures ANOVA (($4 \times S$)) shows a significant main effect of the combined display condition on error rate ($F(3,45)=11.047$, $p<.001$ $\eta_p^2=.424$). LSD pairwise comparisons reveal the mean error rate for condition S-H (stereo with head-tracking) is significantly lower than all other three display conditions (NS-NH ($p<.001$), S-NH ($p=.012$) and NS-H ($p=.004$)). The error rate of S-NH is significantly lower than NS-NH ($p<.001$). Unexpectedly, head-tracking alone (NS-H) does not lead to significant improvement in accuracy over the no stereo no head-tracking (NS-NH) condition ($p=.308$). Figure 8887Figure 8 illustrates mean error rate across all four conditions.

Average task completion time across all conditions is 14.4s. We expected the addition of stereo to reduce response time. While the stereo means were faster the differences were not significant. Further, tests for three-way (including participant pool), two-way and one-way ANOVA are not significant.

7. DISCUSSION

In a prior shorter presentation off this study [25], we erroneously used a non-repeated measures ANOVA to test for display × pool interaction (4 × 2) and found a significant interaction. However, here we use a corrected and more nuanced procedure, a three-way mixed design, namely pool × stereo × motion (2 × (2 × 2 × S)), which yields no participant pool interactions. Instead, the more statistically powerful repeated measures ANOVA finds a significant effect on error rate over all participants. The general trends of Section 6.1's repeated measures analysis on error rate for entire participant pool are in-line with the prior non-repeated measures analysis of the separated CS pool (e.g. the more sensitive repeated measures shows stereo and head-tracking help over the all pool, not just the CS pool).

The task in Experiment 2 is more difficult than in Experiment 1. Over all display conditions error rate range is 37.1% to 58% compared to 25% to 38% in Experiment 1. In Experiment 2 chance guessing would be expected to yield an error rate of 77% while Experiment 1 would be 50%. This indicates even in the worse condition—no stereo, no motion--participants perform better than chance.

In both the prior report and our current analysis, there is no main effect on response time; however, additionally in Section 6.1, we test and find no three-way nor two-way interactions which could have theoretically masked the main effect. Further, we note here that there is a non-significant trend for shorter mean response time with stereo and head-tracking. Possibly, a more statistically powerful experiment could find a small effect which is being masked by the depth ordering task's relative difficulty. Another possibility is that with the NS-NH condition yielding an average 58% error rate participants are essentially giving up after 14s thus capping the response time. Perhaps a study that gives user feedback on the correctness of their answer and allows for a limited number of additional attempts to answer the depth ordering question correctly might find a larger variation in the response time required to obtain a correct answer.

7.1 General Discussion

Our results indicate that semi-immersive VR display conditions affect how well a user perceives depth within a volumetric dataset. More specifically, the results support the hypothesis that stereo with head-tracking significantly improves accuracy in depth discrimination task and depth ordering task in a volumetric dataset. The SH conditions generally outperform all other combinations and conditions with stereo generally out perform their non-stereo counter-parts.

Prior Work Comparisons

Ware et al. [2] compared display conditions' effect on a participant's ability to determine whether two nodes in a network are connected. The conditions are 2D rendering, 3D rendering (no motion, no stereo), stereo, passive rotation, stereo plus passive rotation, hand controlled rotation, stereo plus hand controlled rotation, head-coupled motion parallax, stereo plus head-coupled motion parallax. The network had 75 nodes and 100 arcs. The results confirmed stereo plus motion is the most effective and show that which method is used for producing motion is not particularly important.

While the 3D network data and our volume data clearly differ, some useful comparisons can be drawn. Our Experiment 1 and 2 are consistent with network study's finding that stereo plus motion is most effective. However, in the network experiment motion alone showed a greater advantage than stereo alone. In contrast, in our experiments, the motion alone conditions did *not* demonstrate significant improvement over the stereo alone conditions. For Experiment 1 this might be explained by our shorter exposure time (2 sec) compared to the network experiment where user response time varied from 5 to 15 seconds depending on the node count. With a shorter exposure time, there is simply less time to gather structure-from-motion cues. However, in our Experiment 2 the average task time is a similar 14.4 seconds. Here motion cues can perhaps become more useful. However, again the motion only condition did not exhibit significant improvement for depth ordering accuracy. Overall in both our volume data experiments, the stereo conditions had more significant pair-wise comparisons which suggest that for volumetric data, unlike for 3D network data, stereo may be more significant than motion while stereo plus motion still yields the best accuracy.

Ware et al. [3] repeat a modified version of their network experiment comparing stereo, the kinect-depth-effect, stereo with kinect-depth-effect, and plain 3D (no stereo or motion) for viewing 3D graphs of varying sizes. A major difference

is the use of a 3840 × 2400 Wheatstone stereoscope rather than the 1024 × 768 time-multiplexed display in the earlier study. This change caused the improvement due to stereo plus motion to be roughly an order magnitude, rather than merely the threefold improvement found earlier.

Their participants had up to 5 seconds to view each trial after which the screen went blank until the participant responded. Average response time ranged between 1.5 and 3 seconds. This a similar range to our depth discrimination task's limit of 2 seconds, but less than our depth ordering tasks average of 14.4 seconds. Note, our Experiment 1's kinect-depth-effect rotates through 20 degrees in 2 seconds. The 3D network experiment rotates 360° per 36 seconds, implying 30° for 3 seconds view. In the high-res network experiment for inexperienced observers stereo was the most useful cue, while for the experienced observers (the experimenters themselves) motion was the most useful (i.e. gave greater incremental improvement in accuracy). The former result of inexperienced subjects is inline with our Experiment 1 and 2 results where post-hoc comparisons show the stereo conditions' better performance to be statistically significant. We did not include ourselves in our experiments, but as we note earlier in Experiment 1, for the CS pool subjects (who had more gaming experience) head-tracking alone did show a significant effect, although it was not stronger than stereo. This suggests repeating our study with highly experience observes to see whether motion proves a stronger cue.

The volume depth discrimination's average range of error, 25% to 37%, across display condition is similar to the range of error found in the largest tested 3D networks (1000 nodes), but is beyond the range in the 33 node network, roughly 5% to 15% error. Given the similar error rate range and trial duration, this suggests a similar level of task difficulty between the 1000 node task the volumetric depth discrimination task. The volume depth ordering task appears even harder given its 37.1% to 58% error rates.

Our depth ordering task took significantly longer per trial (average 14.4s) than the high-res 3D network task (1.5s to 3s) but it still improves in accuracy with stereo. In the 3D network task, display condition does significantly affect response time, with the stereo group performing 15% faster, but in our depth ordering task response time did not improve. This further suggests the volume depth ordering task is more difficult and possibly the increased difficulty swamp any improvement due to display condition. Another possible explanation is subjects are essentially giving up in the worst case condition (no stereo, no motion). This hypothesis could be further explored as discussed in Section 6-2.

Finally, the large improvement that the two 3D network studies found when going from a 1024 × 768 time-multiplexed stereo display to a 3840 × 2400 Wheatstone display (sans cross-talk) strongly suggests one may find similar greater enhancements in volumetric depth tasks for the stereo plus motion case with higher resolution systems.

8. CONCLUSION AND FUTURE WORK

In this paper, we examine the effect of stereoscopy and structure-from-motion on depth discrimination and depth ordering tasks for a volumetric dataset. The stereo plus motion condition is the most effective in both experiments. And we found that both stereoscopy and motion by itself improves user's depth judgment in the depth ordering task. In the depth discrimination task, however, head-tracking by itself helps user's depth judgment only for our CS-pool participants who report playing more computer games.

The work of Ware and Mitchell suggests stereoscopy's enhancement of depth perception of volumetric data may be even greater for display resolutions approaching that of the human eye. This is open to further experimentation.

A challenging area is evaluating how display conditions and volumetric software rendering parameters interact to effect perception when both are varied. There are a large number of potential independent variables and interactions to evaluate. For all such studies, our results indicate we should develop a more robust pre-experiment questionnaire that would allow separating participants into groups based on degree of experience with gaming, VR type technologies, and expertise in viewing stereoscopic volume data. Future work should include a range of more elaborately generated volumetric datasets to better mimic real-world data sets while at the same time providing randomly varying volumetric structures to avoid spurious learning effects across repeated trials.

REFERENCES

[1] Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I., [3D User Interfaces: Theory and Practice], Addison Wesley (2005).

[2] Ware, C. and Franck, G., "Evaluating stereo and motion cues for visualizing information nets in three dimensions,"

ACM Trans. Graph., 15(2), 121-140 (1996).

[3] Ware, C. and Mitchell, P., "Visualizing graphs in three dimensions," ACM Trans. Appl. Percept., 5(1), 1-15 (2008).

[4] Ware, C. and Mitchell, P., "Reevaluating stereo and motion cues for visualizing graphs in three dimensions," APGV '05: Proceedings of the 2nd symposium on Applied perception in graphics and visualization, 51-58 (2005).

[5] Wartell, Z., Houtgast, E., Pfeiffer, O., Shaw, C., Ribarsky, W., and Post, F., "Interaction Volume Management in a Multi-scale Virtual Environment," in [Advances in Information and Intelligent Systems], ZbigniewW. Ras and William Ribarsky, Eds., Springer Berlin Heidelberg, 251, 327-349 (2009).

[6] Mora, B. and Ebert, D. S., "Instant Volumetric Understanding with Order-Independent Volume Rendering," Computer Graphics Forum, 23(3), 489-497 (2004).

[7] Maciejewski, R., Choi, S., Ebert, D. S., and Tan, H. Z., "Multi-Modal Perceptualization of Volumetric Data and Its Application to Molecular Docking," WHC '05: Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 511-514 (2005).

[8] Davis, E. T. and Hodges, L. F., "Human Stereopsis, Fusion, and Stereoscopic Virtual Environments," in [Virtual Environments and Advanced Interface Design], Woodrow Barfield and Thomas A. Furness III, Eds., Oxford University Press, ch. Human Stereopsis, Fusion, and Stereoscopic Virtual Environments, 145-172 (1995).

[9] Grossman, T. and Balakrishnan, R., "An evaluation of depth perception on volumetric displays," AVI '06: Proceedings of the working conference on Advanced visual interfaces, 193-200 (2006).

[10] Stewart, A. J., "Vicinity shading for enhanced perception of volumetric data," Visualization, 2003. VIS 2003. IEEE, 355-362 (2003).

[11] Svakhine, N. A., Ebert, D. S., and Andrews, W. M., "Illustration-Inspired Depth Enhanced Volumetric Medical Visualization," IEEE Transactions on Visualization and Computer Graphics, 15(1), 77-86 (2009).

[12] Bruckner, S. and Gröller, E., "Enhancing Depth-Perception with Flexible Volumetric Halos," IEEE Transactions on Visualization and Computer Graphics, 13(6), 1344-1351 (2007).

[13] Boucheny, C., Bonneau, G.-P., Droulez, J., Thibault, G., and Ploix, S., "A perceptive evaluation of volume rendering techniques," ACM Trans. Appl. Percept., 5(4), 1-24 (2009).

[14] Hancock, D. J., "Distributed volume rendering and stereoscopy display for radiotherapy threatment planning," The University of Manchester, Ph.D. dissertation (2001).

[15] Kersten, M., Stewart, J., Troje, N., and Ellis, R., "Enhancing Depth Perception in Translucent Volumes," IEEE Transactions on Visualization and Computer Graphics, 12(5), 1117-1124 (2006).

[16] Hubbold, R. J. and Hancock, D. J., "Stereo display of nested 3D volume data using automatic tunnelling," (1999).

[17] Kruger, J. and Westermann, R., "Acceleration Techniques for GPU-based Volume Rendering," VIS '03: Proceedings of the 14th IEEE Visualization 2003 (VIS'03), 38 (2003).

[18] Roettger, S., Guthe, S., Weiskopf, D., Ertl, T., and Strasser, W., "Smart hardware-accelerated volume rendering," VISSYM '03: Proceedings of the symposium on Data visualisation 2003, 231-238 (2003).

[19] Ling, F. and Yang, L., "Improved on Maximum Intensity Projection," AICI '09: Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence, 491-495 (2009).

[20] OpenSceneGraph. http://trac.openscenegraph.org/projects/osg//

[21] Kniss, J., Kindlmann, G., and Hansen, C., "Multidimensional Transfer Functions for Interactive Volume Rendering," IEEE Transactions on Visualization and Computer Graphics, 8(3), 270-285 (2002).

[22] Pfister, H., Lorensen, B., Bajaj, C., Kindlmann, G., Schroeder, W., Avila, L. S., Martin, K., Machiraju, R., and Lee, J., "The Transfer Function Bake-Off," IEEE Comput. Graph. Appl., 21(3), 16-22 (2001).

[23] Ware, C., [Information Visualization: Perception for Design; 2nd Edition], Morgan Kaufmann (2004).

[24] Yeh, Y. Y. and Silverstein, L. D., "Limits of fusion and depth judgment in stereoscopic color displays," Hum. Factors, 32, 45-60 (1990).

[25] Cho, I., Dou, W., Wartell, Z., Ribarsky, W., and Wang, X., "Evaluating depth perception of volumetric data in semi-immersive VR," Proceedings of the International Working Conference on Advanced Visual Interfaces, 266-269 (2012).