



Evaluation of Depth of Field for Depth Perception in DVR

Pascal Grosset, Mathias Schott, Georges-Pierre Bonneau, Hansen Charles

► To cite this version:

Pascal Grosset, Mathias Schott, Georges-Pierre Bonneau, Hansen Charles. Evaluation of Depth of Field for Depth Perception in DVR. IEEE Pacific Visualization 2013, Feb 2013, Sidney, Australia. pp.81-88, 10.1109/PacificVis.2013.6596131 . hal-00762548

HAL Id: hal-00762548

<https://inria.hal.science/hal-00762548>

Submitted on 7 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation of Depth of Field for Depth Perception in DVR

A.V.Pascal Grosset*

School of Computing and SCI Institute,
University of Utah

Mathias Schott†

NVIDIA - School of Computing and SCI Institute,
University of Utah

Charles D. Hansen§

School of Computing and SCI Institute,
University of Utah

Georges-Pierre Bonneau‡

LJK - Grenoble University - INRIA

ABSTRACT

In this paper, we present a user study on the use of Depth of Field for depth perception in Direct Volume Rendering. Direct Volume Rendering with Phong shading and perspective projection is used as the baseline. Depth of Field is then added to see its impact on the correct perception of ordinal depth. Accuracy and response time are used as the metrics to evaluate the usefulness of Depth of Field. The on site user study has two parts: static and dynamic. Eye tracking is used to monitor the gaze of the subjects. From our results we see that though Depth of Field does not act as a proper depth cue in all conditions, it can be used to reinforce the perception of which feature is in front of the other. The best results (high accuracy & fast response time) for correct perception of ordinal depth is when the front feature (out of the users were to choose from) is in focus and perspective projection is used.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture; I.3.m [Computer Graphics]: Miscellaneous; I.4.10 [Image Processing and Computer Vision]: Image Representation—Volumetric

1 INTRODUCTION

As Direct Volume Rendering (DVR) gains acceptance and becomes more pervasive, it is becoming increasingly important to test the effectiveness of the different rendering techniques that claim to improve perception in DVR. While it is easy to judge the horizontal and vertical positions of features in an image, estimating the depth of these features is more difficult. As pointed out by Boucheny et al. [4], estimating depth even in a trivial highly transparent scene can be very hard and DVR images often have highly transparent features. To aid depth perception, humans rely on a number of depth cues [13], many of which are present in DVR images. Table 1 describes the different depth cues available in static monocular scenes.

For solid surfaces, shadows are very useful in helping us estimate relative ordering of features. Lindemann et al. [17] conducted a user study on the use of different illumination methods and found that Directional Occlusion Shading [27] (DOS) brings about 20% improvement in ordinal depth perception. However, for highly translucent surfaces, as shown in Figure 1, DOS is less effective. Depth of Field (DoF) is another visual cue which may aid depth perception.

There are different kinds of depth descriptors. Absolute descriptions are usually quantitative and are defined in units like meters or relative to the viewer’s body. Relative descriptions relate one property to another. This can be further broken down into relative

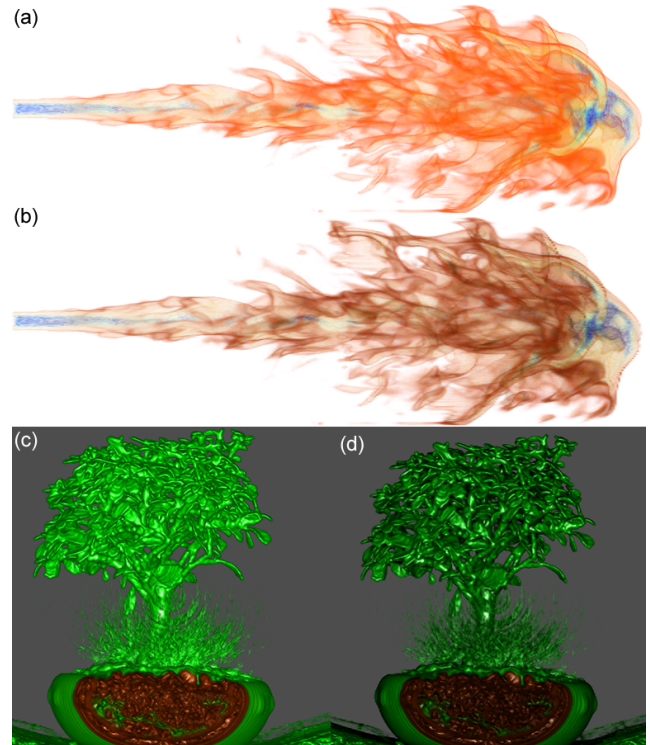


Figure 1: Flame Dataset (a) Original Rendering (b) with occlusion shading. We see that having ambient occlusion does not improve ordinal depth perception in this type of images. (c) shows the bonsai dataset with well defined surfaces. We can see that applying directional occlusion shading (d) helps with depth perception in this case.

and ordinal. “Relative descriptions relate one perceived geometric property to another (e.g., point a is twice as far away as point b). Ordinal descriptions are a special case of relative measure in which the sign, but not the magnitude, of the relations is all that is represented.” [29].

In this paper, we present a study on the use of DoF and its impact on the perception of ordinal depth in DVR and we try to establish under which conditions it is most beneficial. DoF cause regions further from the focal plane to be increasingly blurred while the focal plane region remains sharp. Photographers often use shallow depth of field, that is having an object which is relatively close to the camera in focus while the background is not, to put emphasis on a subject or feature they want viewers to focus.

Many researchers have carried out studies to understand the role of blur [6] and DoF in perceiving depth [11, 12, 19, 21] but their finding is by no means conclusive. The conclusions range from blur being at best a weak depth cue [20] to “an important depth cue which has been previously underestimated” [11]. For DVR, the only work we found that studies the application of DoF to improve depth perception was by Ropinski et al. [25] which used a modified DoF and focussed only on angiography images described in

*e-mail: zetwal@cs.utah.edu

†e-mail: mschott@cs.utah.edu

‡e-mail: georges-pierre.bonneau@inria.fr

§e-mail: hansen@cs.utah.edu

the related work section. To the best of our knowledge, this is the first comprehensive user study on the use of DoF as a depth cue in DVR. The main contributions of this paper are:

- Establish whether DoF is a useful depth cue in DVR images.
- Determine for which kind of datasets is DoF most helpful.

The DoF technique proposed by Schott et al. [26] was implemented in the SLIVR renderer of the VisIt Visualization software [18]. The experiment is run on site to be able to control most experimental parameters including screen quality, luminosity and attention of subjects. The 3 main hypotheses we want to test are:

1. HYP1: DoF will help improve the accuracy of ordinal depth perception in a volume rendered image where there are multiple features.
2. HYP2: DoF will help improve the speed of ordinal depth perception in a volume rendered image where there are multiple features.
3. HYP3: If users view a moving focal plane, correct perception of ordinal depth will improve.

An experiment is set up that has two parts: a static and dynamic part to test these hypotheses that are being put forward. In the static part, to test the first two hypothesis, we used 150 static images of 5 different datasets. In each image two features are selected and the subject is asked to identify which one is closer. For the dynamic part of the experiment, to test the third hypothesis, we change the focal length of the lens and create a video where the focal plane sweeps through the volume from the front to the back and to the front again. The reason why we use a video instead of providing direct control to the test subject to move the focal plane is to avoid possible biases introduced by user interaction [30]. We use 20 videos of about 17 seconds for this part.

The rest of the paper is structured as follows: Section 2 discusses the findings of previous research on depth perception and DoF, and the different approaches for DoF in computer graphics and DVR. Section 3 describes how DoF has been implemented for this user study. In section 4, we describe the setup of the experiment, state our results and discuss their implications. The last section, section 5, states the conclusions and proposes some future work.

Table 1: Monocular Static Depth cues

Depth Cue	Description
Depth of Focus	Objects in focus appear sharp while those not in focus appear blurred.
Familiar Size	When we use our prior knowledge of the world to judge distances. E.g. The smaller a plane looks in the sky, the further away from us we know it is.
Occlusion	Objects in front of others overlap those at the back and so hide part of the back objects.
Perspective	Parallel lines appearing to converge at infinity.
Relative Size	Objects which are far away from us take a smaller area in our field of view.
Shading Brightness	Objects which are far from us tend to appear more dimly lit than objects that are close to us.
Shadow	When we know where the light source is, we make use of where the shadow will fall to decide where the object is.
Texture Gradient	Fine details are clearly seen for objects which are close to us compared to objects that are far away.

2 RELATED WORKS

2.1 Depth perception

To be able to perceive the depth of objects in the real world and in a synthetic image, the human visual system makes use of a number of depth cues. Depth cues have been studied thoroughly [13, 29] and

are generally grouped into two categories: monocular and binocular. Binocular refers to the inter ocular offset to perceive depth in a scene while monocular depth cues refers to having a single image of the scene. For this user study, only one view is being shown on the screen. Furthermore, there is no motion. So we only have static monocular depth cues. These are explained in Table 1.

Depth of focus is one of the depth cues the human visual system uses to perceive depth. In our eyes, the shape of the lens is distorted by ciliary muscles to focus light on the retina. The process of modifying the shape of the lens of our eye is called accommodation. The human visual system gets information of the amount of distortion of the lens from the ciliary muscles and can use this as a depth cue; especially for very close objects [29]. Also, the part of the image not in focus, in front and behind the focal plane will appear blurred. Using the amount of blurriness of an image as a depth cue is not a viable indication of how close or far an object is. As indicated by Mather et al. [21], blur discrimination is poor in human vision.

2.2 Depth of Field in Graphics

Several attempts have been made to reproduce DoF in computer graphics. Barsky et al. [2] have a comprehensive study of different DoF techniques in computer graphics. Some of the earliest applications of DoF are from Cook et al. [7] where they simulated DoF in distributed raytracing applications by sending multiple rays. In interactive computer graphics, a common technique is to sort the scene along depth and apply different amounts of blur to different levels of depth which Barsky et al. [1] and Kraus et al. [16] did. In DVR, DoF has been proposed by Crassin et al. [8] for large volumes in the Gigavoxels Framework. Ropinski et al. [25] is of greater interest since in addition to having DoF, they also conduct a user study on the use of modified DoF to enhance spatial understanding; their application only looked at angiography images. In their work, they try to determine where the user is focussing in an image. The part deemed to be behind the region in focus has DoF effects applied to it. Consequently, there are no DoF effects in front of the focal plane. From their user study they find that DoF helps to improve the correctness of depth perception but they also saw an increase in response time. They attribute this to the user having to get used to some part of the image appearing out of focus since they study other techniques for improving the correct perception of depth and DoF is one of the ones they test. To implement DoF in DVR, we referred to the work of Schott et al. [26] that describes a DoF technique for slice based Volume rendering. A view-aligned slice based volume renderer is used where the view-aligned slices are blurred according to the distance from the viewer. Though this technique can suffer from over-blurring, its results are quasi-indistinguishable from a Monte Carlo ray tracer and so we chose this approach to implement DoF.

2.3 Perceptive studies for DoF

On the subject of whether DoF is a valid depth cue, literature seems to be quite undecided. Table 2 shows a summary of the previous results. The general agreement among researchers is that DoF helps improve depth perception when combined with other depth cues. Mather et al. [21] established that blur discrimination is poor in humans. However, increasing the number of depth cues improves the perception of depth Mather et al. [22]. In their study, the subjects were presented with a set of tiles and used a combination of blur, contrast and interposition as depth cues. They were asked to identify which tile appeared closer to them by clicking on them starting by the closest. Blur on its own gives about 65% correct depth results but when combined with other depth clues like interposition and blur, this rises to about 90%. This was confirmed by Held et al. [12] who conclude that blur on its own does not help to estimate absolute or relative distance but coupled with perspective, it is much more effective. In their latest study Held et al. [11] claim

Table 2: Previous findings on the impact of DoF and blur as a depth cue.

Research	Findings
Held et al. 2012 [11]	Blur is used to make depth estimation significantly more precise throughout visual space.
Mather et al. 2004 [22] Held et al. 2010 [12]	Blur when combined with a number of other depth cues improves depth perception.
Mather 1996 [20] Mather et al. 2002 [21]	Blur can only be used to establish ordinal depth.

that combined with perspective, blur can even be used to guess distances. Since DVR often uses perspective, it could prove to be a very valuable addition. The only study on the impact of blur in DVR that we have found is from Ropinski et al. [25].

3 DEPTH OF FIELD FOR DVR

The mechanics of DoF can be derived from the Thin-Lens equation:

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{z_f} \quad (1)$$

where f is the focal length, s is the distance from the lens to the focal plane and z_f is the distance from the lens to the object as shown in Figure 2 (a). When light from a point in the scene passes through a camera lens, it should ideally be focussed on a single point on the image plane. However, if the image plane is not at the correct position, the point is mapped to a circular region instead, c on the diagram in Figure 2 (a). The diameter of the region can be determined according to the equation 2:

$$c(z) = A \frac{|z - z_f|}{z} \quad (2)$$

where $c(z)$ is the diameter of the circle of confusion, A is the aperture of the lens as shown in Figure 2 (b). This figure is quite revealing from a perception point of view. It shows that on both sides of the focal distance z_f , we can have regions having similar diameter for the circle of confusion which would translate to the same amount of blur.

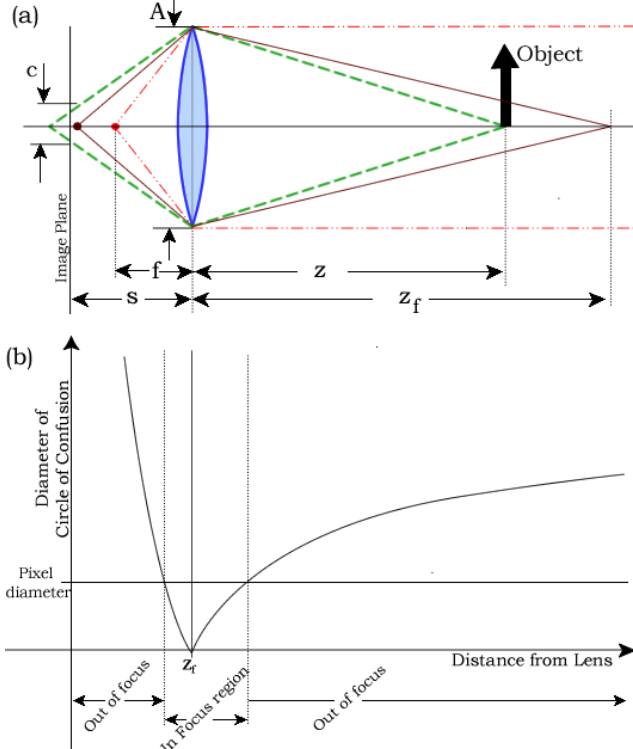


Figure 2: (a) Lens Setup, (b) Circle of Confusion

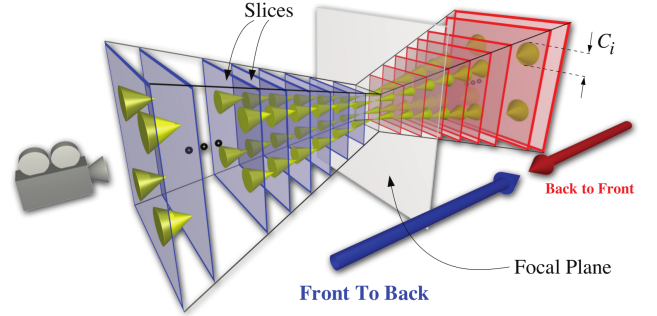


Figure 3: Geometric setup of the DVR implementation of DoF (image courtesy of Mathias Schott [26])

DoF is implemented as shown in Figure 3 and explained in [26]. A GPU slice based volume renderer is used and the scene is broken down into 2 sections namely before and after the focal plane. For the part between the focal plane and the front of the volume (where the camera is), the volume is processed from the front of the volume to the focal plane (in a front to back manner) and each slice is blurred according to its position. For the part behind the focal plane, the volume is processed from the end of the volume to the focal plane (in a back to front manner). Each rendered slice is blended with a blurred region of the previous slice. The blur kernel's size decreases as the slice approaches the focal plane. Two directions are needed to ensure that the in focus region does not contribute to the blurred region. A more detailed description of the algorithm with pseudocode can be found in [26].

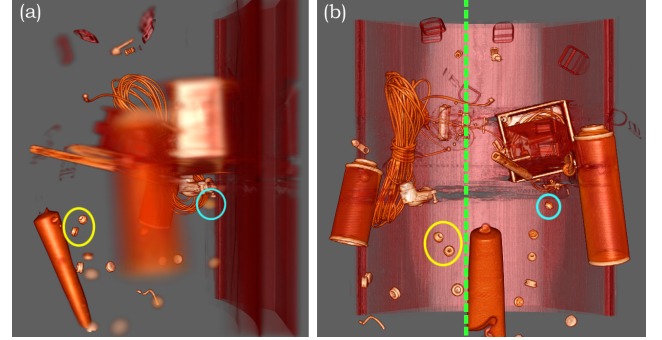


Figure 4: (a) Backpack displayed within Psychopy. The features to choose from have been circled. (b) Side view of the same dataset (focus plane shown as dashed line) not shown in the experiment. We can see that the features are quite far apart.

4 USER STUDY

The aim of the experiment is to determine whether DoF provides a better understanding of the ordinal depth of different features in a volume rendered image. More specifically, we want to be able to check these three hypotheses: 1) DoF helps improve the accuracy of the determination of ordinal depth; 2) DoF helps improve the speed of the determination of ordinal depth; 3) if users can change the position of the focal plane, correct perception of ordinal depth will improve.

Our focus is only on ordinal depth as absolute depth, for different datasets, can mean very different things but yet they can appear to be of similar size in a volume rendered image. For example, the size of the backpack dataset could be about 40 cm while the size of the aneurism dataset would be about 10 cm.

To test these three hypotheses, we will carry out two experiments. In the static part, we show the test subjects a number of images in which we ask them to select which of two circled features (located at different depths in the image) is in front. Figure 4 shows an example of this. In the dynamic part, we show a video of a DVR dataset where the focal plane sweeps from the front to the back and back to the front. Here again two features are circled at different depths and the subject is asked to decide which one is

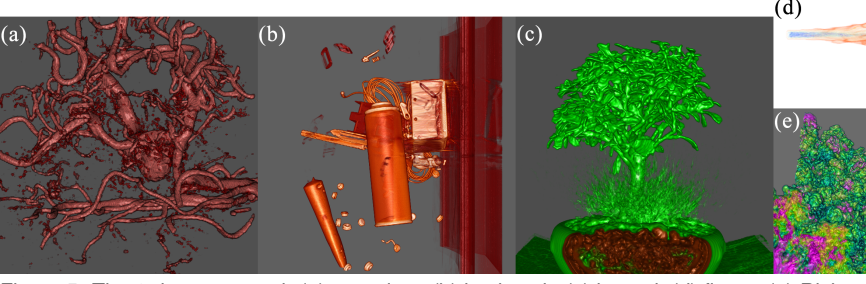


Figure 5: The 6 datasets used: (a) aneurism, (b) backpack, (c) bonsai, (d) flame, (e) Richtmyer-Meshkov instability & (f) thorax. The reason for having a different background color is explained in Section 4.1

in front. As discussed by Knill [14], the influence of depth cues varies depending on the task and what might weigh more for one task might be less important for another. Consequently, to make our experiment as general as possible, we decided to have minimal interaction. Sections 4.1 to 4.3 describe the parameters and setup of the experiment.

4.1 Stimuli Description

To generate the stimuli, DoF was implemented as described in Section 3 in the SLIVR renderer of VisIt 2.4.2. PsychoPy 1.7.4 [24] was used to present the images and movies to the subjects, and to collect their answers.

As can be seen in Figure 5 the background for each dataset is different. The background (except the flame dataset where the background interfered too much due to the highly transparent nature of the image) was carefully chosen by computing the Michelson contrast. This was done as follows: the image is generated in VisIt with a background which can be easily removed from the image such as pure green (RGB: 0 1 0) or pure blue (RGB:0 0 1) depending on the image. The Michelson contrast [23] is computed for all the colors which do not match the pure green or blue as follows:

$$M_c = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \quad (3)$$

where M_c is the Michelson Contrast, L_{max} is the maximum luminance and L_{min} is the minimum luminance where luminance [5] is calculated as follows:

$$L = 0.2126 * Red + 0.7152 * Green + 0.0722 * Blue \quad (4)$$

where *Red*, *Green* and *Blue* are the RGB components of the color. The background color is assigned to be a grey RGB value (same red, green and blue) for which calculation of the Michelson contrast with the background is the same as calculating the Michelson contrast with only the dataset ignoring the background.

Five of the six datasets were used for the first static images part of the experiment. This is due firstly to time constraints: we did not want the experiment to last too long but still manage to have enough data per dataset. Secondly some features of the torso dataset would disappear when blurring was applied for DoF; some of the circled features would disappear and the subject would see an empty ellipse or sometimes the feature behind the selected feature.

The datasets and their associated transfer function were selected so that we have a mixture of shapes which are familiar (like the bonsai which is tree shaped) and unfamiliar (like the flame dataset) to non-volume rendering users. Also, each resulting image has a different number and type of depth cues. The depth cues that we can expect are: Occlusion, Perspective, Relative Size, Familiar Size and Texture Gradient. For this test we did not use shadows or changes in shading. The perspective projection settings will be controlled so that all datasets have the same settings. Table 4 shows a taxonomy of the depth cues for the images (Figure 5) that we are using and Table 3 describes each of the images. The minimum separation between the features to be selected in an image was 5% of the whole

Table 3: Images description

Image	Description
aneurism	Has lots of occlusion which should be quite helpful to perceive depth but the complex network of veins is quite confusing.
backpack	The earbuds are totally disconnected and appear to float in mid air which is quite unfamiliar. Moreover, even for people who are familiar with depth of field, they normally see it as a progression over the image. This is not the case here. The background is not blurred, only the volume is. However, there is still a certain amount of occlusion that could be helpful.
bonsai / Richtmyer-Meshkov instability	People are familiar with these kind of shapes. The Richtmyer-Meshkov instability looks like a landscape (though it is not one) and everyone is familiar with seeing trees. Moreover, they tend to have similar depth cues as shown in Table 4.
flame	This is a combustion dataset that is extremely hard to understand. The chosen transfer function is what chemical engineers have used to visualize this data. One of the complaints was that it is very hard to understand this static image in a publication and so they would like to know if DoF could help in this case.
thorax	This image is complex due to the presence of many structures some of which are quite transparent and faded completely when DoF was applied. So, we used it only in the videos.

Table 4: Images and their associated depth cues. We have 3 levels for each: High, Med and Low indicating how useful each of the depth cue is expected to be in each image.

Dataset	Occlusion	Relative Size	Familiar Size	Texture Gradient
aneurism	High	Med	Med	Med
backpack	Low	Med	Low	Low
bonsai	High	High	High	High
flame	Low	Med	Low	Low
thorax	Med	Med	Low	Low
Richtmyer-Meshkov instability	Med	Med	Med	High

depth of the volume and the maximum separation was 60%. The number of test images for each separation range is as follows: [0 - 10%]: 12, [10% - 20%]: 23, [20% - 30%]: 22, [30% - 40%]: 37, [40% - 50%]: 49, [50% - 60%]: 7.

With perspective projection, objects which are closer to the viewer appear bigger than objects which are far. This will be the baseline to which DoF is added to see whether ordinal depth is perceived better. We have also included orthographic projection to allow us to verify the net impact on only having DoF. Also, we expect to see a performance improvement from orthographic to perspective projection which would match what other researchers have found; increasing the number of depth cues will increase correct perception. Also, one of the things a good user study [15] should take care of is verifying that the participants are committed to answering truthfully. That can be done in this case by verifying that perspective projection gives an improvement compared to using orthographic projection.

4.2 Experimental procedure

Environment setup The experiment was conducted on site so as to ensure that we have similar environment settings for all participants of the experiment. The study was carried out in a room where the curtains were closed to shut off light from the outside (which will vary during the day) and illuminated by white fluorescent light. This is to make the screen more visible as well as to increase the performance of the eye-tracker which requires specific light condition for better accuracy of the tracker. Eye tracking is not the most important part of the study but was included to see if it would help us understand the results of the experiment. The T2T (Talk to Tobii) [10] package was used to interface with the Tobii T60 eye-tracker [28].

Apparatus A Macbook pro was used to run PsychoPy which was connected to the eye-tracker. A gamepad was used as input device as it is more ergonomic than a keyboard and mouse. On the gamepad, the subject presses any of the left buttons to select the left feature and any of the right buttons to select the right feature. This is important since the subject will spend on average 30 minutes for the whole experiment and so we want to make it as easy and less tiring for them as possible to try to minimize the impact of boredom and fatigue. Also, interaction with a gamepad is less likely to introduce perceptual bias as it offers a simple and straight-forward method of entering selection.

Participants & Design 25 subjects (6 females, 19 males) participated in the user study. All of them had good eye sight or corrected vision; 7 wore glasses and none were color blind. All but one test subjects reported to be right handed. The age range was as follows: 15 - 20: 1, 21 - 30: 21, 31 - 40: 3. All the participants had some experience with computer graphics through games or movies but most of them were not familiar with volume rendering: none of them were students, researchers or users developing or working with volume rendering.

A within-subject design was used where all the participants did all the tests. The experiment, carried out over 2 days with 25 participants, used a two-alternative forced-choice methodology [3]. Each participant spent approximately 30 mins doing the experiment (along with calibrating the eye-tracker and training for the experiment) and each test was followed by a debriefing session.

4.3 Tasks

Static Experiment To test the first two hypotheses, of whether DoF helps improve the speed and accuracy of perception of ordinal depth in a scene. We conduct the following experiment. We have 150 images of 5 datasets (aneurism, backpack, bonsai, flame and Richtmyer-Meshkov instability) under different 6 different conditions namely:

1. orthographic projection
2. orthographic projection & front feature in focus (DoF Front)
3. orthographic projection & back feature in focus (DoF Back)
4. perspective projection
5. perspective projection & front feature in focus (DoF Front)
6. perspective projection & back feature in focus (DoF Back)

In each image, we present the subject with two features. Each feature is surrounded by an ellipse and located at different horizontal positions such that it is clearly distinguishable which one is on the left and which one is on the right. The features are located at different depths and participants are asked to choose the one which they perceive as being in front. To select the left feature, the subject just has to click any of the left buttons on the gamepad. Same for the right feature. Figure 4 (a) shows an example image from the user study. On choosing a feature, the color of the circle changes. The order in which the different images were shown was randomized so as not to have all images under one specific condition or for

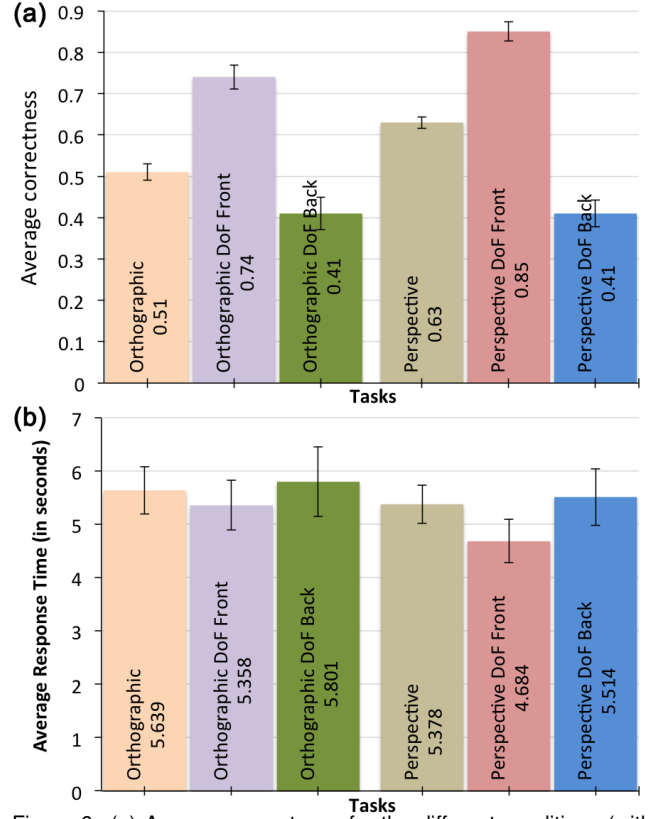


Figure 6: (a) Average correctness for the different conditions (with static images) of the experiment with standard error. (b) Average response time for the different datasets and conditions (with static images) of the experiment with standard error.

a particular dataset following each other. However, all participants were shown the test images in the same pre-randomized order. The submitted answer and the time taken is recorded for each test image.

Dynamic Experiment To test the third hypothesis, whether being able to change the position of the focal plane will improve the correct perception of relative depth, we carry out a second experiment. To minimize user interaction that could introduce bias in the experiment [30], we made a video of the focal plane sweeping from the front to the back and back to the front. This is referred to as the dynamic part of the experiment. We have 20 videos of the 6 datasets (aneurism, backpack, bonsai, flame and Richtmyer-Meshkov instability & thorax) each lasting approximately 17 seconds. We record only the answers, not the time, as the subjects are asked to watch each video completely before answering.

Both experiments were conducted as follows:

1. Calibration of the eye-tracker. The subject is asked to stare at a red circle on the screen which appears for 4 seconds at one position and then jumps to another spot. The calibration stage lasts for about one minute.
2. Training for the static part of the experiment. Firstly, the overall task is explained to the user. The input device to be used (gamepad) is introduced and the subject is briefed on how to use it. Next 7 training images are presented, 4 with DoF and 3 without. In each image, two features are circled with an ellipse and the subject is asked to select which one is in front. We request the correct answer for each before proceeding to the next image to ensure that the subject understands the task at hand. On completing this phase, we inform the subject that we will start with the experiment but now they will not be required to give the correct answer to proceed to the next image.

3. The static experiment. Each of the 150 images is shown to the participant who is asked to select which feature appears to be in front. The circle around the chosen feature changes color on being selected and the next test image appears. Response time and answers are recorded along with the eye tracking data.
4. Training for the dynamic experiment. A short animation describing the motion of the focal plane is presented to the user along with an explanation on what they have to do.
5. The dynamic experiment. Each of the 20 videos is shown and the participant is asked to choose which of the two circled features appears to be in front. The answer is recorded along with the eye tracking data.
6. The experiment ends with a debriefing session where the subject is asked for verbal feedback on the experiment which was noted down by the experimenters.

4.4 Experiment 1: Static

4.4.1 Results

The average correctness and completion times recorded during the experiment are used to analyze the results.

The overall comparison of average correctness under the 6 conditions for all the subjects reveals that there is a significant difference in the results for average correctness [one-way, repeated-measures ANOVA, $F(2.2,52.8) = 49.754$, $p < 0.001$]. Figure 6 (a) shows the results for average correctness per task. For the individual datasets, when the average correctness under the 6 different conditions per dataset is compared, we see a statistically significant difference for average correctness of the results [two-way, repeated-measures ANOVA, $F(20,480) = 35.153$, $p < 0.001$]. Figure 7 (a) shows the average correctness and Figure 7 (b) the mean response time for each dataset. Running ANOVA for on the response time shows that there is a low statistically significant difference between the means for perspective and perspective DoF front [one-way, repeated-measures ANOVA, $F(1,24) = 6.6$, $p < 0.017$]. Standard error is computed as follows: σ/\sqrt{n} where n is the number of observations.

We hoped the eye-tracker would show users trying to find separation between the in focus and out of focus region to make a decision. Unfortunately such behaviors did not show up. For most test-subjects, we noticed that their gaze jumped from one of the circled features to the other. In some cases, their gaze would linger on some of the central figures like the canister in the backpack dataset or they would try to follow the veins in the aneurism dataset.

4.4.2 Discussion

The results show that, as expected, perspective projection is better than orthographic projection but hypotheses 1 and 2 are not fully validated.

Perspective projection is expected to be better than orthographic projection as it causes objects in front to appear bigger and humans are used to having bigger things in front of smaller ones. This is in line with what Mather et al. [22] reported: combining depth cues improves response speed and accuracy.

The first hypothesis, DoF will help improve the accuracy of ordinal depth perception in a volume rendered image where there are multiple features, is only partly validated. Therefore HYP1 is supported by the experiment if DoF is in front but is not supported by the experiment if DoF is on the back. The reason for that, we believe, is due to depth cue conflict [9]. Drascic et al., reported that when depth cues provide conflicting information, there is an increase in uncertainty and decrease in accuracy. Humans are used to seeing objects far away as blurred while those in front are usually well defined. When DoF is on the back feature, the front feature appears blurry contradicting what we are used to. Boucheny et al. [4] report a similar incident in their user study for volume rendering. In

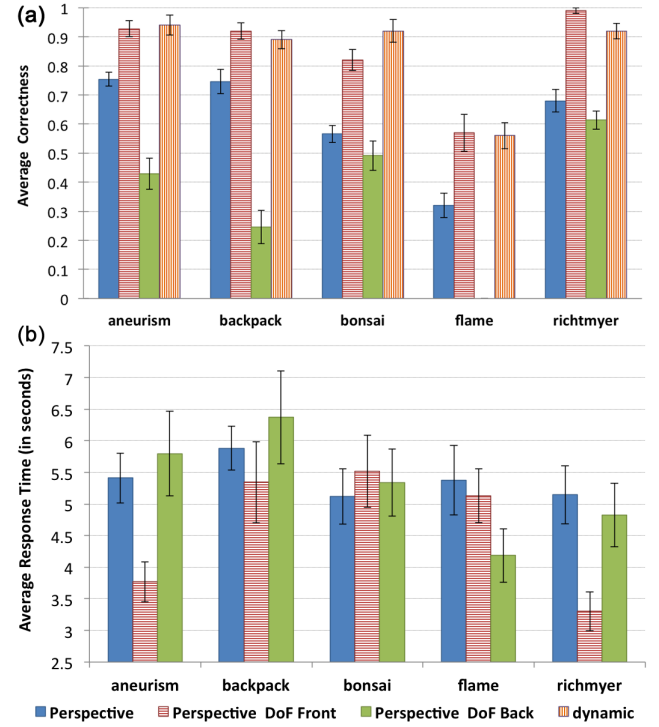


Figure 7: (a) Average correctness for the different datasets, including static and dynamic (b) Mean response time taken for the different datasets and conditions. Note: the 0 value for the flame dataset for Perspective DoF Back indicates that all answers were wrong.

their first experiment, they had two cylinders arranged in different z depth and the subject was asked to state which one was in front of the other. Whenever the small cylinder was in front, the percentage of correct answers would drop drastically, from 70% to 30%. This is likely due to depth cue conflict since we are used to seeing objects close to us as big and far away objects as small.

For the second hypothesis, DoF will help improve the speed of ordinal depth perception in a volume rendered image where there are multiple features, is not fully validated. We see that both accuracy and speed improved for the DoF focused on the front feature in all cases except the bonsai. While the average correctness for the bonsai dataset increased from 57% for perspective projection with DoF to 82% for perspective projection with DoF front, the participants took more time to reply. This might have been due to the participants spending more time analyzing the image since that had more information at their disposal. Decrease in speed when DoF is on the back feature can be explained by depth cue conflict. The flame dataset is an exception to that. However, this can be explained by what the participants told us during the debriefing phase. Many of them found the flame dataset with or without DoF still extremely hard to understand and very often gave up on trying to find a correct answer.

We also see that we can identify some pattern in our results that seem to be linked to the datasets. Namely, we see that the Richtmyer-Meshkov instability and the bonsai seem to behave quite similarly: both have a slight difference between perspective and DoF with perspective on the back feature, and a marked improvement for perspective with DoF front. We believe that this is because these two have shapes with which we are quite familiar. The Richtmyer-Meshkov instability looks like a landscape (though it is a CFD dataset) and from the eye tracking data, we can see that the subjects sweep it as a landscape. Moreover, during the debriefing, it was often referred to as a landscape by the participants. The bonsai looks like an ordinary tree. So familiarity with the shape helped the subjects in identifying the correct location of the features; familiarity is also one of the main depth cues that people use

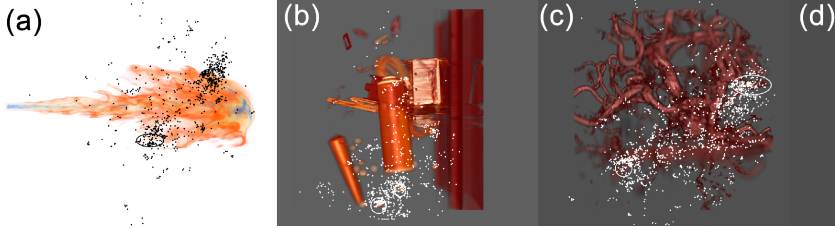


Figure 8: Tracking examples: (a) flame, (b) backpack, (c) aneurism, (d) bonsai & (e) Richtmyer-Meshkov instability where the white and black dots represent gaze for all the users.

to correctly determine depth. On looking at the response time, we see that the response speed for perspective DoF front is higher for the Richtmyer-Meshkov than for any other dataset and condition.

The backpack dataset proved to be very challenging to understand. While we usually have a floor or ground in the real world that we use as frame of reference for the horizon, here, everything floats in mid air. Perspective projection aided viewers as the closer small spherical-like objects appeared bigger than the remote ones. The result is strikingly similar to the aneurism dataset. Both of these datasets have similar average correctness values for the perspective, perspective with DoF front and perspective with DoF back. The mean response time are quite similar except for DoF front. We believe that this is due to similarities between these two datasets: both lack a floor like structure and have shapes which are not very common.

The flame was the hardest for the subjects to understand as the high degree of transparency hindered most of the depth cues that we have. Adding DoF helped increase correct perception from about 33% to 55%. While 55% is similar to a chance guess, it is nevertheless much better than having no DoF at all. When DoF is on the back feature, participants never answered correctly. As mentioned in Table 1, most of the depth cues that we normally would rely on have a very low impact here. Hence performance at correctly determining ordinal depth of features is poor. Also, as noted previously, DoF on the back feature was the fastest. It is the only dataset that does not have any predictability in the results that we saw.

So after breaking down the results per dataset, we observe the same behavior; DoF on the front object helps reinforce the correct perception that an object is closer to the viewer. When DoF is applied on the back object, depth cue conflicts reduces performance.

4.5 Experiment 2: Dynamic

4.5.1 Results

Figure 9 shows the average success rate for the different datasets. While the accuracy rate is around 90% for aneurism, backpack, bonsai and Richtmyer-Meshkov instability, accuracy for the translucent datasets of the flame and thorax are lower. From Figure 7 (a) we see that having DoF in the dynamic part, is not much better than DoF on the front feature except for the bonsai dataset. Doing an ANOVA to compare the means for DoF Front and dynamic [two-way, repeated-measures ANOVA, $F(2.28, 68.2) = 1.58$, $p < 0.204$] reveals that there is no statistically conclusive difference between the means of these two. However, as evinced in Figure 10, the performance with videos seem to improve after repeated exposure. Note that time is not used here as a performance metric as the subjects were advised to watch the whole video which evened out the response time.

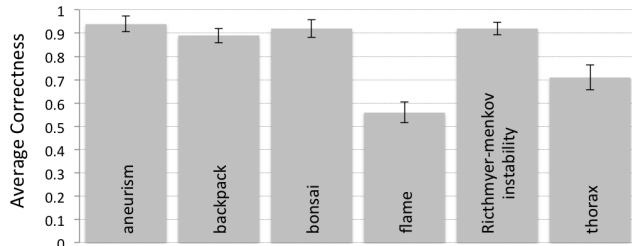


Figure 9: Average correctness for the different datasets.

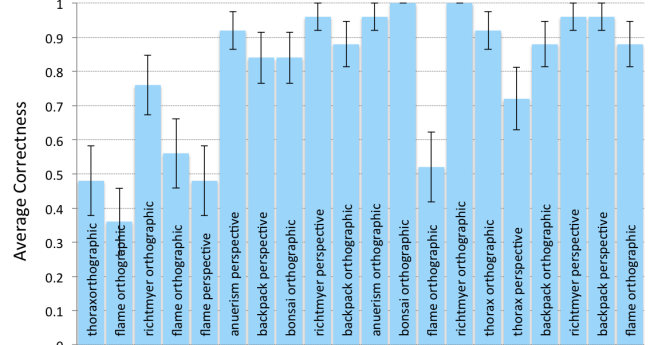


Figure 10: Average correctness for the different datasets. The datasets from left to right shows the ordering in which the videos were shown to the participants.

4.5.2 Discussion

Hypothesis 3: if users view a moving focal plane, correct perception of ordinal depth will improve, is not validated by this experiment. From Figure 7 (a), we see that a static DoF focused on the front feature is often better than having a moving plane, except for the aneurism and the bonsai. Overall though, dynamic is always better than DoF on the back feature which degrades accuracy by a lot.

During the debriefing session we found that the most accurate answers were from subjects who quickly understood that since the focal plane was sweeping from the front to the back and back to the front, the feature which appears non-blurred first is the one in front. If they have missed the ordering on the first pass, they would try to see it on the second pass. The less successful subjects usually had to see more videos to understand the mechanics of the moving plane; that the front feature is the first that will first appear in focus. Also, from the debriefing session we learned that some subjects were focussing on one feature and did not see when the other feature was changing from out of focus to in focus. This could have been better had the subject had direct control on moving the focal plane. They would have been able to check exactly each of the one feature moves in and out of focus.

We also see that transparency is still a major issue even in the videos. For datasets that have hard surfaces, accuracy improves except for the flame and the thorax where we still see quite low accuracy rate. One of the issues with the translucent surfaces is they can disappear after DoF has been applied. For these datasets, we probably need a finer control over the focus to improve the results.

Here again we have mixed results. While with videos, DoF performance is more consistent; the subjects correctly identified which of the two circled features is the front one. However, it did not match our expectations of having better performance when compared to DoF front in static cases for all the datasets.

4.6 Guidelines

Based on what we have observed, we would recommend that to improve perception of ordinal depth, it is very important that DoF be used on the front object and not the back for static images. Using DoF on the front object successfully reinforces the correct perception of depth. If the feature that needs to be focused on is near the back of the volume, rotating the volume by 180 degrees about the y-axis to bring that feature to the front might be something worth

considering. We would also encourage users to have as many depth cues as possible. In line with other research on depth perception, we see that the more depth cues we have, it becomes easier for the users to correctly understand the arrangement of objects.

Using a video would be much beneficial if this option is available, especially if the back feature needs to be in focus. The negative impact of having DoF on the back object that we notice in static images is sensibly reduced here as the users can see features moving in and out of focus. However, the users should be told exactly what to expect so that they are not surprised by an animation popping up. As shown in Figure 10, good results improve over time as the participants got used to the videos.

5 CONCLUSION AND FUTURE WORK

We have conducted a user study on the impact of DoF for ordinal depth perception in DVR. From our results we see that using DoF on the front object reinforces correct perception of depth in DVR. However, putting DoF on the back object leads to depth cue conflicts and the results are worse than not using DoF. Appropriate use of DoF provides a consequent improvement in terms of correct depth perception for general cases in DVR. For the dynamic part of the experiment, we saw a general overall improvement though performance is still worse for highly translucent datasets.

In the future there are a number of experiments that we would like to try with regards to different interactions models, and comparisons against other depth cues. We would also like to study these using more datasets such as the engine and different models of blur.

Lindemann et al. [17] investigated the role of shadows while this paper investigated the role of DoF and blur. It would be interesting to compare the performance of these two and to combine these two as well. Additionally, with a completely interactive system, we could also study the role of parallax induced by rotation. As pointed out previously, translucent volumes still pose a major challenge when depth is concerned and more experimentation is needed to see which depth cues would be helpful in these cases.

ACKNOWLEDGEMENTS

We would like to thank Timo Ropinski for providing us with some guidelines and sample images from their user study [17].

This research was sponsored by the DOE NNSA Award DE-NA0000740, KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST), DOE SciDAC Institute of Scalable Data Management Analysis and Visualization DOE DE-SC0007446, NSF OCI-0906379, NSF IIS-1162013, NIH-1R01GM098151-01.

REFERENCES

- [1] B. A. Barsky, D. R. Horn, S. A. Klein, J. A. Pang, and M. Yu. Camera models and optical systems used in computer graphics: Part ii, image-based techniques. In *Computational Science and Its Applications - ICCSA 2003*, Lecture Notes in Computer Science, pages 256–265. Springer Berlin / Heidelberg, 2003.
- [2] B. A. Barsky and T. J. Kosloff. Algorithms for rendering depth of field effects in computer graphics. In *Proceedings of the 12th WSEAS international conference on Computers*, ICCOMP'08, pages 999–1010, Stevens Point, Wisconsin, USA, 2008. World Scientific and Engineering Academy and Society (WSEAS).
- [3] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, 113(4):700–765, October 2006.
- [4] C. Boucheny, G.-P. Bonneau, J. Droulez, G. Thibault, and S. Ploix. A perceptive evaluation of volume rendering techniques. *ACM Trans. Appl. Percept.*, 5(4):23:1–23:24, Feb. 2009.
- [5] B. Caldwell, M. Cooper, L. Guarino Reid, and G. Vanderheiden. Web content accessibility guidelines (wcag) 2.0. <http://www.w3.org/TR/WCAG20/>, 12 2008.
- [6] K. J. Ciuffreda, B. Wang, and B. Vasudevan. Conceptual model of human blur perception. *Vision Research*, 47(9):1245 – 1252, 2007.
- [7] R. L. Cook, T. Porter, and L. Carpenter. Distributed ray tracing. *SIG-GRAPH Comput. Graph.*, 18(3):137–145, Jan. 1984.
- [8] C. Crassin, F. Neyret, S. Lefebvre, and E. Eisemann. Gigavoxels: ray-guided streaming for efficient and detailed voxel rendering. In *Proceedings of the 2009 symposium on Interactive 3D graphics and games*, I3D '09, pages 15–22. ACM, 2009.
- [9] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In *SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III*, pages 123–134, 1996.
- [10] L. Filippin. T2tpkg. <https://sites.google.com/site/t2tpkg/>.
- [11] R. T. Held, E. A. Cooper, and M. S. Banks. Blur and disparity are complementary cues to depth. *Current Biology*, 22(5):426 – 431, 2012.
- [12] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks. Using blur to affect perceived distance and size. *ACM Trans. Graph.*, 29(2):19:1–19:16, Apr. 2010.
- [13] R. B. J. Howard and I. P. P. *Seeing in Depth*, volume 2 Depth Perception. I Porteous, 2002.
- [14] D. Knill. Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. *Journal of Vision*, 5(2), 2005.
- [15] R. Kosara, C. Healey, V. Interrante, D. Laidlaw, and C. Ware. Visualization viewpoints. *Computer Graphics and Applications, IEEE*, 23(4):20 – 25, july-aug. 2003.
- [16] M. Kraus and M. Strengert. Depth-of-field rendering by pyramidal image processing. *Computer Graphics Forum*, 26:645–654, September 2007.
- [17] F. Lindemann and T. Ropinski. About the Influence of Illumination Models on Image Comprehension in Direct Volume Rendering. *IEEE TVCG(Vis Proceedings)*, 17(12):1922–1931, 2011.
- [18] LLNL. Visit visualization tool, 2012.
- [19] J. A. Marshall, C. A. Burbeck, D. Ariely, J. P. Rolland, and K. E. Martin. Occlusion edge blur: a cue to relative visual depth. *J. Opt. Soc. Am. A*, 13(4):681–688, Apr 1996.
- [20] G. Mather. Image blur as a pictorial depth cue. *Proceedings: Biological Sciences*, 263(1367):pp. 169–172, 1996.
- [21] G. Mather and D. R. Smith. Blur discrimination and its relation to blur-mediated depth perception. *Perception*, 31(10):1211 – 1219, 2002.
- [22] G. Mather and D. R. Smith. Combining depth cues: effects upon accuracy and speed of performance in a depth-ordering task. *Vision Research*, 44(6):557 – 562, 2004.
- [23] A. A. Michelson. *Studies in Optics*. Univ. Chicago Press, Chicago, IL, 1927.
- [24] J. W. Peirce. Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, 162(1–2):8 – 13, 2007.
- [25] T. Ropinski, F. Steinicke, and K. Hinrichs. Visually supporting depth perception in angiography imaging. In *Smart Graphics*, pages 93–104. Springer, 2006.
- [26] M. Schott, A. V. P. Grosset, T. Martin, V. Pegoraro, S. T. Smith, and C. D. Hansen. Depth of field effects for interactive direct volume rendering. *Comput. Graph. Forum*, pages 941–950, 2011.
- [27] M. Schott, V. Pegoraro, C. D. Hansen, K. Boulanger, and K. Bouatouch. A directional occlusion shading model for interactive direct volume rendering. *Comput. Graph. Forum*, 28(3):855–862, 2009.
- [28] T. Technology. Tobii t60 and t120 eye tracker. http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_T60_T120_EyeTracker_UserManual.pdf.
- [29] W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual Perception from a Computer Graphics Perspective*. A. K. Peters, Ltd., Natick, MA, USA, 1st edition, 2011.
- [30] A. Wykowska, A. Schubö, and B. Hommel. How You Move Is What You See: Action Planning Biases Selection in Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6):1755–1769, 2009.