

Evaluating the Perception of Semi-Transparent Structures in Direct Volume Rendering Techniques

Rickard Englund

Scientific Visualization Group, Linköping University, Sweden

Timo Ropinski

Visual Computing Group, Ulm University, Germany

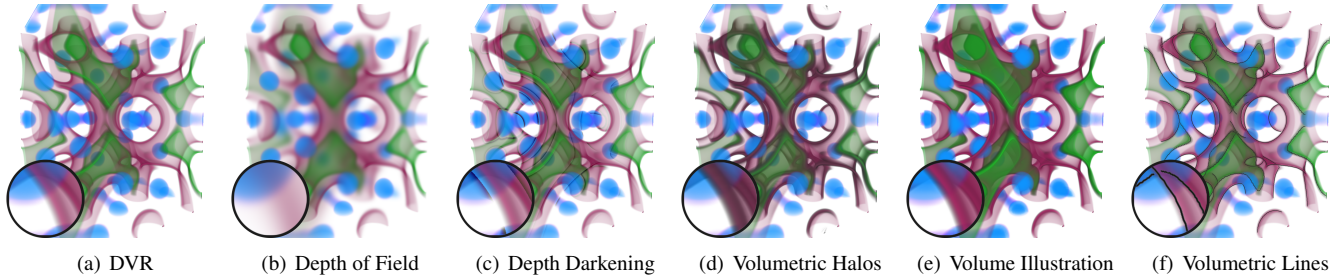


Figure 1: We have investigated the perceptual impact when applying six widely used volume rendering techniques to semi-transparent structures. The tested techniques are standard (a) Direct Volume Rendering (DVR), (b) Depth of Field, (c) Depth Darkening, (d) Volumetric Halos, (e) Volume Illustration, and (f) Volumetric Line Drawings.

Abstract

Direct volume rendering (DVR) provides the possibility to visualize volumetric data sets as they occur in many scientific disciplines. A key benefit of DVR is that semi-transparency can be facilitated in order to convey the complexity of the visualized data. Unfortunately, semi-transparency introduces new challenges in spatial comprehension of the visualized data, as the ambiguities inherent to semi-transparent representations affect spatial comprehension. Accordingly, many visualization techniques have been introduced to enhance the spatial comprehension of DVR images. In this paper, we conduct a user evaluation in which we compare standard DVR with five visualization techniques which have been proposed to enhance the spatial comprehension of DVR images. In our study, we investigate the perceptual performance of these techniques and compare them against each other to find out which technique is most suitable for different types of data and purposes. In order to do this, a large-scale user study was conducted with 300 participants who completed a number of micro-tasks designed such that the aggregated feedback gives us insight on how well these techniques aid the end user to perceive depth and shape of objects. Within this paper we discuss the tested techniques, present the conducted study and analyze the retrieved results.

Keywords: Volume rendering, transparency, depth perception

Concepts: •Human-centered computing → Empirical studies in visualization; Visualization design and evaluation methods;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.

SA '16 Symposium on Visualization, December 05 - 08, 2016, Macao

ISBN: 978-1-4503-4547-7/16/12

DOI: <http://dx.doi.org/10.1145/3002151.3002164>

1 Introduction

Today volume rendering techniques have become mature and it is possible to generate high quality volume rendered images at interactive frame rates. As a consequence, volume rendering is used in a wide spectrum of scientific disciplines, ranging from material science to medicine, to support the interactive exploration of volumetric data sets. In contrast to polygonal models which are frequently used in computer graphics, volumetric data sets also capture the interior of an object of interest. Thus, visualizing them can be considered as more challenging, since internal structures tend to be occluded and inter-object relations are difficult to perceive due to the nested structure of the represented objects. Boucheny et al. even state that perceptually enhanced visualizations are one of the major challenges of tomorrow's engineering systems [2009]. Semi-transparency is often exploited to deal with these visualization challenges, by reducing the occlusion of internal structures. Unfortunately, as our human visual system is better skilled in estimating opaque structures, the perception of semi-transparent structures is far less reliable and perception of these representations tends to be ambiguous [Hibbard 2000]. Thus, while semi-transparency is able to reduce the occlusion of internal structures, it does not necessarily support perceptually resolving inter-object relations.

Due to the problems that occur when perceiving semi-transparent structures, the perceptual improvement of semi-transparent volume renderings has been addressed by many researchers in the past, which led to several extensions of direct volume rendering (DVR). The presented techniques either mimic physical phenomena to which improved perception is accounted to, or they facilitate illustrative mechanisms often inspired by well established illustration techniques. As these techniques vary with respect to the made enhancements, it can also be expected that they have a varying impact on semi-transparency perception. Unfortunately, it is unknown which of these techniques is best to be used in which scenarios. Therefore, we have selected five of the most important perceptual enhancement techniques for DVR, and have measured and compared their perceptual impact when visualizing scenes containing semi-transparent structures. These techniques, all consider depth values in the compositing equation, and have been selected based on initial perceptual findings, and the widespread usage in volume

rendering, as determined based on exchange with other researchers and our own experience, as well as the number of citations of the publications proposing these techniques. Thus, we compare standard DVR to Depth of Field rendering [Schott et al. 2011], Depth Darkening [Svakhine et al. 2009], Volumetric Halos [Bruckner and Gröller 2007], Volume Illustration [Rheingans and Ebert 2001], and Volumetric Line Drawings [Burns et al. 2005]. An example for the application of each of these techniques is shown in Figure 1. While there are certainly many more techniques available which could have been evaluated, these techniques span a wide range of enhancement effects in the continuum between more realistic to more illustrative, and they can thus be considered as a representative subset of the available techniques.

Within the presented evaluation we have focused on two essential parts of the scene perception process: depth perception and shape perception. Depth perception enables the depth discrimination of different scene objects, based on their distance to the viewer. Shape perception on the other hand supports the understanding of the shape of scene elements. As both are essential for an unambiguous scene perception, we have designed our perceptual evaluation such that it enables us to test both qualities independently. Our experiments are driven by the following two hypothesis, which are based on Ware’s discussion on Depth and Shape perception [2012] and will be discussed in Section 6:

- H1: depth perception is improved when back-to-front relations are encoded unambiguously through image contrast.
- H2: shape perception is improved when the curvature of the shape can be inferred at nearby silhouettes.

As we were additionally interested in the subjective appeal of the tested techniques, we have also conducted a subjective image comparison task. By exploiting crowdsourcing through the Amazon Mechanical Turk (MTurk) platform, we could include the feedback of 281 participants in our study, which has been used to acquired a total of 22,716 perceptual judgments.

2 Related Work

We have structured the related work into three categories, which are discussed separately. First, we review volume rendering enhancements, before we explain how the presented study is influenced by comparable perceptual evaluations. Finally, we review work relevant to crowdsourcing experiments in the area of visualization.

Volume rendering enhancements. While we address the evaluated techniques in the next section, we like to briefly review other important approaches that we have not include in our study. As we also employ silhouette-based approaches, we would like to point out that several other high quality silhouette extraction techniques exist. Nagy and Klein have proposed the extraction of silhouettes based on a robust two step process [2004]. As our study focuses on techniques exploiting DVR, several other approaches have not been included in our study, such as techniques related to maximum intensity projection (MIP) and X-ray representations. Thus, we did not incorporate MIP-enhancement techniques [Díaz and Vázquez 2010], or order-independent enhancements [Mora and Ebert 2004].

Other approaches of interest for semi-transparency perception belong to the class of focus-and-context techniques [Viola et al. 2005]. Together with approaches borrowed from shading computation to modulate transparency [Bruckner et al. 2005; de Moura Pinto and Dal Sasso Freitas 2011], these techniques belong to the group of illustrative volume rendering enhancements. While we have included some techniques of this class in our study, due to time and resource limitations we have decided to focus on those, which require no

additional parameter settings. One approach we would like to include in future studies is the illustrative context-preserving volume rendering technique introduced by Bruckner et al. [2005]. For now, we have not evaluated this approach, as we felt that the dependency on the shading equation combined with the slice-distance parameter κ_t violates our goal to minimize parameter variations. Also, as we are focusing on interactive volume exploration, we decided to not incorporate techniques which take over control of the user parameters usually set in volume rendering. Accordingly, we did not include any camera steering approaches, or approaches which automatically adapt the transfer function or other relevant rendering parameters [Chan et al. 2007; Chan et al. 2009; Zheng et al. 2013].

Transparency evaluations. Several other evaluations have been conducted in order to investigate the perception of semi-transparent structures in computer generated images. In this paragraph we outline the most relevant ones, and describe how they have influenced our study setup. When considering previous studies, two types of perception are in focus: depth perception and shape perception. While Boucheny et al. have evaluated depth perception of DVR with a three-alternative forced-choice test [2009], more recent studies also collect continuous feedback with point-based judgment trials [Lindemann and Ropinski 2011; Grosset et al. 2013]. Our study is inspired by these evaluations, and we use similar depth trials: ordinal and absolute depth judgments. While some of the previous evaluations of the depth perception of volume rendered images have been conducted with a special domain focus [Ropinski et al. 2006; Kersten-Oertel et al. 2014], our study keeps the domain open, and thus does not make implications about the nature of the used data sets. Apart from these depth perception evaluations, several authors have presented studies in which they have investigated shape perception, e. g., [Baer et al. 2011; Solteszova et al. 2012]. Interrante et al. have for instance focused on the perception of surface shape, by determining how texturing affects the perception of nested semi-transparent surfaces [1997]. More recently Bair and House have conducted similar experiments investigating the shape perception of layered grid-textured surfaces [2007]. As these and other evaluations use the gauge figure task to investigate shape perception, we have also decided to employ this task in our study.

Crowdsourcing perception. In recent years the usage of crowdsourcing techniques has emerged as an effective way to conduct large scale user studies with a heterogeneous group of subjects at low costs [Quinn and Bederson 2011]. Since its emergence several perceptual studies have been conducted in the graphics community, whereby among others the shape perception of line renderings [Cole et al. 2009], and the perception of materials in illustrative graphics has been investigated [Bousseau et al. 2013]. Recently, it could also be shown that crowdsourcing can be an effective way of investigating the perceptual capabilities of visualizations [Heer and Bostock 2010], which led to the conduction of several interesting crowdsourcing studies in visualization. Prominent examples include the perception of averages in multi-class scatter plots [Gleicher et al. 2013], but also the memorability of visualizations [Borkin et al. 2013]. However, in the area of scientific visualization not many such studies have been published. An exception directly related to our study is the human computation based transparency investigation conducted by Ahmed et al. [Ahmed et al. 2012]. By exploiting a simple game, they investigate different blending functions with respect to their impact on the perception of semi-transparent layers. Furthermore, Englund et al. presented a complete system for conducting evaluation of scientific visualization algorithm using crowdsourcing [2016].

3 Evaluated Techniques

Due to the challenges associated with perceiving semi-transparent structures, a multitude of techniques exist which are aimed at improving the perceptual qualities of volume rendered images. Nevertheless, due to time and resource constraints we needed to select a subset of these techniques for our evaluation. To identify this subset we have taken into account several criteria. The main criterion for selecting the tested techniques was their applicability to DVR. As one aspect of our investigation focuses on depth perception, we have decided to discard those techniques, which do not exploit depth in the compositing equation. Accordingly, MIP and X-ray based approaches, e.g. [Díaz and Vázquez 2010], have not been considered. Since we investigate the perception of semi-transparent structures, we do not explicitly cover iso-surface representations, but take into account peak-based transfer functions which result in a similar visual appearance. As a consequence we did not include any technique which exploits surface parameterizations, such as for instance the texture-enhancement proposed by Interrante et al. [1997]. Another important aspect which went into the selection of the tested techniques was the required user parameter tuning. The goal was to keep the influence of user parameters low, as we wanted to obtain generalizable results. As there are virtually no techniques without parameters, we have constrained ourselves to those techniques where parameters are limited to controlling the strength of the effect as opposed to the effect's area of influence. Thus, all tested techniques are applied in a holistic manner with respect to the visualized data set, instead of being constrained to certain regions in the volume. Consequently, we did not take into account any illumination-based approaches [Jönsson et al. 2014], as lighting and material parameters would violate this parameter constraint. Therefore, also more complex focus+context techniques which require the specification of a degree of importance [de Moura Pinto and Dal Sasso Freitas 2011] or a focus region [Viola et al. 2005] have also been excluded. To focus on techniques which support interactive exploration through parameter variations, we have further not included any techniques taking control over the rendering parameters. Finally, while refraction is often cited as a cue for indicating semi-transparent structures, we have excluded such techniques from our study as refraction is considered as perceptually ambiguous [Thompson et al. 2011, p. 261]. This left us six volume rendering techniques to be evaluated. In the following paragraphs we briefly address these techniques, while we provide a more detailed explanation of the technical realization of each technique in the appendix of this paper. See Figure 1 for a visual example of each technique.

DVR. We have used standard DVR as the most basic technique to compare the evaluated extensions against.

Depth of Field. Schott et al. have presented a technique which integrates a Depth of Field effect directly into volume rendering [2011]. As perceptual models indicate a perception benefit of such techniques, we have included Depth of Field rendering in our study.

Depth Darkening. By unsharp masking the depth buffer [Luft et al. 2006], Depth Darkening achieves an effect visually comparable to ambient occlusion, as it emphasizes cavities through the darkening effect. As it has also been applied in the area of volume rendering [Svakhine et al. 2009], we have chosen this approach for our evaluation.

Volumetric Halos. By integrating halo effects directly into GPU-based volume rendering, it becomes possible to extend these effects to the entire volume [Bruckner and Gröller 2007], and not only the silhouettes. To investigate the impact of this extension, we have included Volumetric Halos.

Volume Illustration. Volume Illustration is a group of techniques first presented by Rheingans and Ebert [2001]. We have integrated their approach for silhouette-enhancement in our study.

Volumetric Line Drawings. Burns et al. presented a technique to extract contour lines from volumetric data [2005]. We incorporate this technique and extend it by introducing the haloed line effect presented by Appel et al. [1979], as the halos emitted by the lines are expected to help solving the ambiguity of T and X-junctions [Thompson et al. 2011, p. 258].

Based on their individual capabilities we have sorted the six selected techniques into four groups by keeping our hypothesis H1 and H2 in mind. We have grouped Volumetric Halos and Depth Darkening together, since both introduce illumination-like darkening to encode depth clues. Furthermore, the Volume Illustration and Volumetric Line Drawings techniques both tried to enhance depth perception by enhancing silhouettes. The remaining two techniques, DVR and Depth of Field, are contained within their own groups. According to H1 and the nature of the techniques, we expect the group containing Volumetric Halos and Depth Darkening to perform best in depth related tests. On the other hand, the group containing Volume Illustration and Volumetric Line Drawings is expected to perform best in shape related tasks.

4 Materials and Methods

To compare the selected techniques' influence on the perception of volume rendered images, we have conducted an evaluation as a randomized within subject design, whereby for each of the six compared techniques we have generated 15 volume rendered images, resulting in a total of 90 different images. A total of 300 participants has been recruited through MTurk, from where they have been redirected to our web server to be able to conduct the four different perceptual tasks: ordinal depth estimation, absolute depth estimation, surface orientation, and beauty ranking. During each task we did not enforce any time constraints on the participants. The tasks were grouped based on task type, but the order of the groups and the shown images was randomized. After completing all tasks the participants were exposed to a brief survey through which they could give us their feedback. 281 participants have completed the study and submitted their data.

4.1 Visual Stimuli

During our study we have confined ourselves to the evaluation of static images. The main reasoning behind this decision is the goal to avoid that motion-induced parallax effects overrule other depth cues, introduced by the investigated techniques. Such an observation has been made by Boucheny et al., who report that DVR has been ambiguous in static cases, but that this effect could be counterbalanced with motion parallax [2009]. As DVR serves as the baseline in our study, we have to ensure that we do not introduce other cues overpowering the depth cues provided by DVR. Taking into account that, due to the nature of the standard volume rendering integral, occlusion is the only depth cue existing in DVR, we would like to conserve the capabilities of this cue to make it comparable to the other occlusion-exploiting techniques. This proceeding to aim a unique cue to depth is also inline with previous studies, which limit available depth information in order to reduce the parameters potentially influencing the results. For instance, Kersten-Oertel et al. have conducted a series of two studies, whereby they found kinetic depth cues confusing in the first study, and have therefore eliminated these cues in the second study in order to consider visualization solutions that require no interaction to provide effective depth and structural understanding [2014]. As a practical side-effect this also enables us to make assumption about the used high-quality imagery in papers and magazines.

The evaluated techniques have been implemented in the open source framework Inviwo [Sundén et al. 2015] and all tested images

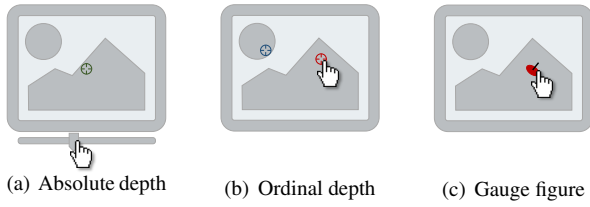


Figure 2: Illustration of the 3 tasks participants had to conduct to rate depth and shape perception. Interaction affordances are depicted by the mouse cursor.

have been generated within the software. The visual stimuli have been carefully generated by using the tested techniques, whereby we have applied a perspective projection. To support generalization of our results, we have aimed at a variety of data sets and transfer function setups during the generation process. With respect to the selected data sets, we have taken into account familiarity and data density as the most important selection criteria. The goal was to span these scales and expose the users to data sets which they were familiar with, but also more abstract 3D structures. From a more perceptual point of view, we were also interested in investigating the effects based on feature distribution, and we have therefore chosen data sets with dense and sparse feature distributions. Thus, the tested data sets range from medical data, through engineering data to material sciences data sets, with a different degree of familiarization and feature density. Also with respect to the used transfer functions, the goal was to span a range of possible effects. Though, as the perception of semi-transparent structures was the main focus of our study, all used transfer functions employ semi-transparency, whereby this can span from more homogeneous cloud-like structures to more surface-like semi-transparent layers.

4.2 Perceptual Tasks

The perceptual tasks have been designed to investigate depth perception, shape perception and visual appeal of the tested techniques. To validate our study setup, we have initially conducted a pilot study with 20 participants. In the following we will discuss the four task types in detail and refer to how the pilot study has influenced their final design. For the task where markers is used we have manually selected location with the purpose of creating a uniform distribution, both spatially in the stimuli and in actual depth values of while keeping an adequate level of difficulty, such that the task could be completed but the correct result was not too obvious. As we have generated the same number of images for each technique, and used the same marker position for the corresponding images, a possible bias introduced through marker positioning would not affect the techniques independently. To increase the visibility of the markers, their color has been selected to maximize the contrast with the background. In addition to the given feedback, we have also measured the time in milliseconds for each task.

4.2.1 Absolute Depth Perception

To investigate the absolute depth perception in a scene we have conducted an absolute depth perception task in a forced choice setup. As shown in the task layout in Figure 2 (a) the visual stimuli were shown together with a slider below them, which had to be used to estimate the depth of the point highlighted by the marker shown on top of the image. Our pilot study indicated, that this task was hard to understand. Therefore, we gave detailed instructions which were shown under each task setup: *Please adjust the slider to indicate the highlighted position's distance to the viewer. The slider's val-*

ues range from the position closest to the viewer to the position most distant from the viewer. In case semi-transparent structures overlap in a position, consider the structure closest to you. Please look at this example before you proceed. Hint: Adjust the slider to the left if the position is close to you and to the right if it is far from you. (Close or Far? If this image was taken with a camera, your position is behind/in the camera). The linked examples showed generic example images together with correct and incorrect slider positions. Once the slider had been moved, the participant could proceed to the next task by pressing a button below the visual stimulus.

4.2.2 Ordinal Depth Perception

To investigate the relative depth perception in a scene we conducted an ordinal depth perception task as a two alternative forced choice experiment. As shown in the task layout in Figure 2 (b) each visual stimulus was shown together with two colored markers. The task was to select the marker closest to the viewer, which we have expressed through the following instructions: *Please click on the marker that highlights the position closest to you. In case semi-transparent structures overlap in a position, consider the structure closest to you.* The participants had to select the chosen marker through a mouse click and could proceed to the next task by pressing the next button at the bottom of the page. Once a marker has been selected, the selection status was displayed under the image, and the selection could be corrected with another click.

4.2.3 Orientation

To investigate the impact on shape perception of the six tested techniques, we have asked the participants to complete a gauge figure task [Koenderink et al. 1992]. As surface orientation is one of the factors influencing shape perception, we see this task as an indicator of the qualities of the tested techniques in this area. The design of our gauge figure task is influenced by the design used by Cole et al. to investigate the perception of line drawings [2009]. Furthermore, some of the observations made by Šoltészová et al. when using gauge figure tasks in the area of volume rendering [2011] were helpful. Thus, we exposed the participants to visual stimuli, which had a randomly oriented gauge figure located at a predefined location (see Figure 2 (c)) while asking them to orient the figure according to the foremost semi-transparent surface orientation. As our pilot study has indicated that this task is difficult to understand, we have decided to include expressive examples. From within each judgment the participants could access these examples which would show a correct and a incorrect gauge orientation for an easy to comprehend shape. The instructions for this task read as follows: *Please rotate the gauge figure to indicate the surface orientation at its corresponding position. A correct orientation would align the red base with the surface and make the black indicator point away from the surface. In case semi-transparent structures overlap in a position, consider the structure closest to you. Please look at this example before you proceed.* To not let color choices influence the orientation of the gauge figure, we have decided to always depict it with the same colors. Furthermore, while Šoltészová et al. have used a transparent base for the gauge figure to improve the visibility of the underlying surface [2011], we have decided to stick to the gauge figure design exploited by Cole et al. [2009], as it makes the perception of the gauge orientation more obvious.

4.2.4 Visual Appeal

To investigate the visual appeal of the tested techniques, we have asked the participants to select the most appealing picture in a two alternative forced choice test. The two alternatives were shown side by side, and the user could select the preferred image by directly

selecting it with the mouse. To proceed to the next task, the user had to press a button on the bottom of the page. As this task is inherently subjective, we have kept the instructions simple as follows: *Please click on the image that appeals most to you.*

4.3 Instructions and Survey

Besides the instructions and examples provided for each individual perceptual task, the participants have been given initial instructions and some background about the evaluations on a welcome screen. We further conducted a survey at the end of each study, where we have asked the participants to rate the given instructions, the difficulty of executing the evaluation, and the satisfaction regarding the payment. Furthermore, we gave them the possibility to provide optional textual feedback on the evaluation. We have used this information during our analysis to discard participants, who have expressed that they had problems taking part in the evaluation.

4.4 Result Analysis

As we have also created depth and normal images when generating the visual stimuli used in our evaluations, we could determine the *correctness* of a perceptual task automatically. For the absolute depth test the error was determined through the difference between the correct and the specified depth value. We have transformed this error into a percentage which expresses the error with respect to the depth extend of the scene. For the ordinal depth test, we have taken into account the average number of correct selections per participant, and for the gauge figure task we have used the angular deviation from the central difference gradient to encode the error. During the visual appeal task the participants saw each technique six times, which means that the possible value range for this measure is between 0 and 6, whereby 6 means good.

When using crowdsourcing to conduct user evaluations we heavily decrease the workload of participant recruitment compared to lab-based evaluations. Though, this comes with a slight cost, since participants are recruited over the Internet we do not have any control over their environment, such as screen size, resolution, brightness etc. The effect of this is commonly neglected when conducting crowdsourcing studies since the larger pool of participants will average out this effect [Kim et al. 2012]. A common problem in crowdsourcing are people who randomly click through the evaluation with the goal of finishing each task in minimum time possible and collect the reward. Therefore, it is expected that we need to discard more users when using crowdsourcing compared to lab-based evaluations. To detect random clicker in our study, we have filtered out participants from the different tasks based on three criteria. First, have used the questionnaire feedback to discard participants who indicated they had problems with the tasks due to technical issues, misunderstandings, visual deficiency etc. Participants who stated that they had trouble understanding a certain task have only been discarded from the analysis of that task. Second, to discard participants who were clicking random options without paying attention to the actual tasks, we implemented an automatic discarding technique. For each task group we have used the trial that had the lowest mean error as a baseline, and discarded those participants who got high error on that trial. For the absolute depth and surface orientation trials we have discarded those participants whose error was higher than 90% of the error of the worst participant. For the ordinal depth trial we discarded the participants who selected the wrong marker in the baseline trial. As a third filter criterion, we have discarded all participants from a task if their answer deviates by more than two standard deviations from the overall mean for that task. We have applied a similar criterion to filter out participants who replied more than two standard deviations faster than the

mean response time. However, this did not result in any discards. Thus based on the three filter criteria, the amount of remaining participants is as follows: absolute depth: 215 (76.5%), ordinal depth: 193 (68.7%), gauge figure: 202 (71.9%). The discarded participants have been flagged on MTurk and did not received any payment. Due to its subjectivity, we have not discarded any participants from the visual appeal task. For the remaining data we have analyzed the correctness and the time taken through a repeated measure analysis of variance (rANOVA) combined with Tukey post-hoc tests with 95% confidence intervals and a power analysis.

5 Evaluation Results

To evaluate the six techniques we have recruited 300 participants. The data has been collected in two subsequent waves, whereby in the first wave 100 participants had to perform all four perceptual tasks, whereby each participant had to make 60 judgments, which took in average 16.5 minutes for a payment of \$1.05. In the second wave we have omitted the visual appeal judgment, and 200 participants had to perform 120 judgments each, which took in average 26 minutes and brought them \$2.00. Out of the 300 recruited participants 19 did not complete the study resulting in 281 participants. In total we have collected 22,716 judgments.

Absolute Depth Perception. The rANOVA showed a significant difference between the average error of the tested techniques ($F(3, 878) = 7.959, p < 0.001$), and a power analysis revealed a power of 100%. Performing the Tukey post-hoc tests revealed that there were three significant clusters of the four tested groups. The Volumetric Halos and Depth Darkening group was with a mean error of 24.87% significantly better performing than all other groups. We could further detect a significant difference between the Volume Illustration and Volumetric Line Drawings group which performed significantly better than Depth of Field with a mean error of 28.20% compared to 33.08%. We could not find any significant difference between these two groups and DVR (28.90%) (see Figure 3(a)). With respect to time we found no significant difference for the tested techniques ($F(5, 244) = 0.615, p = 0.689$) nor our chosen groupings ($F(3, 246) = 0.536, p = 0.658$). Furthermore, comparing the mean error between Volumetric Halos (25.27%) and Depth Darkening (24.45%) shows no significant difference ($p = 0.999$) while between Volume Illustration (25.97%) and Volumetric Line Drawings (32.25%) there is a significant difference ($p = 0.027$).

Ordinal Depth Perception. The rANOVA could not show a significant difference between the average correctness of the tested techniques ($F(5, 690) = 1.91, p = 1.123$) nor our chosen groupings ($F(3, 692) = 1.158, p = 0.325$), while the power analysis gave a power value of 89%. Accordingly, we did not perform the Tukey post-hoc tests, but refer to Figure 3(b) for the calculated mean values. With respect to time we found no significant difference for the tested techniques ($F(5, 171) = 1.336, p = 0.251$) nor our chosen groupings ($F(3, 173) = 0.466, p = 0.707$).

Orientation. The rANOVA showed a significant difference techniques the average error of the respected groupings ($F(3, 752) = 3.474, p = 0.016$), but also with respect to the tested techniques ($F(5, 750) = 6.771, p < 0.001$), whereby the power analysis gave a power of 99%. While looking at the results of the techniques within our groups we found a significant difference in the mean angular error between Volumetric Halos (32.55°) and Depth Darkening (35.25°) and between Volume Illustration (39.03°) and Volumetric Line Drawings (31.69°) there is a significant difference ($p = 0.013$). Therefore we run the analysis per technique rather than the groups. The Tukey post-hoc tests revealed two overlapping clusters where Volumetric Line Drawings and Volumetric

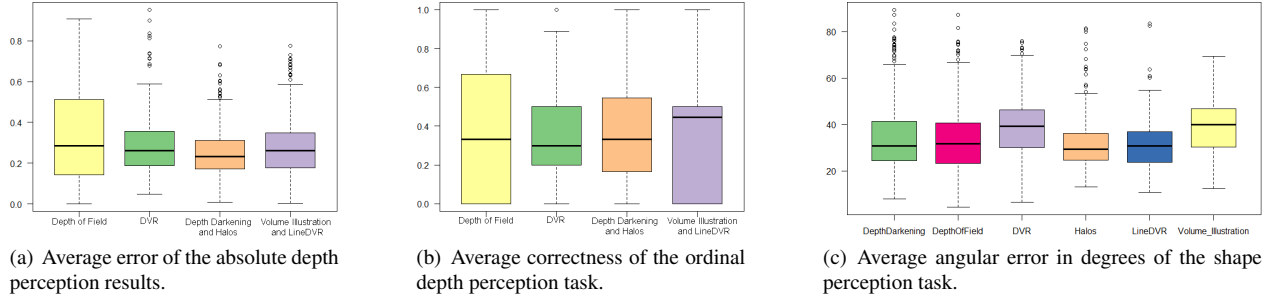


Figure 3: Boxplots showing the results for the two depth perception tasks and the shape orientation task.

Halos were significantly better than the other techniques, except against Depth of Field (35.25°). The second group consists of Depth of Field, Depth Darkening, DVR (38.75°) and Volume Illustration. The means for all analyzed groups are shown in Figure 3(c). With respect to time we found no significant difference for the tested techniques ($F(5, 271) = 0.295, p = 0.916$) nor our chosen groupings ($F(3, 273) = 0.250, p = 0.861$).

Visual Appeal. When analyzing the responses we have received for the visual appeal task, rANOVA showed that the difference in mean values was significant among the tested techniques ($F(5, 665) = 13.86, p < 0.001$). The post-hoc analysis revealed three groupings, the first group having Depth Darkening (2.620) together with Volume Illustration (2.330) and Volumetric Line Drawings (2.307). The second grouping, which overlaps with the first grouping, contains Volume Illustration, Volumetric Line Drawings and DVR (2.17). The third group was perceived significantly less appealing than the other contains Volumetric Halos (1.688) and Depth of Field (1.614).

6 Result Discussion

In the presented evaluation we have studied the impact of six state-of-the-art volume rendering techniques, all developed with the goal of improving the perceptual qualities of volume rendered images. In contrast to the previous extensive study reported by Boucheny et al. [2009], which focuses on achromatic transparency, we have studied the impact on colored images, a procedure backed up by the findings of Fulvio et al. [2006]. Looking at the results by keeping our two hypothesis H1 and H2 in mind reveals interesting findings.

Depth perception. With respect to H1, we expected the group containing Depth Darkening and Volumetric Halos to perform best. While this was the case in the absolute depth task, we could not obtain significant results in the ordinal depth task. Our interpretation of this situation is, that the difference occurs from a difference in the visual scanning of the image under investigation. In the absolute depth task a point has to be related to the surroundings by estimating front and back of the shown data set. We assume that deriving these front and back values can be better supported by depth encoding techniques, since the absolute differences are in general larger. In contrast with the ordinal depth test, participants had to relate the depth of two points *in* the volume to each other. Naturally, here the differences in depth are much smaller and the accuracy of the convolution-based techniques might not be appropriate to communicate this difference accurately. As a consequence, we conclude that Volumetric Halos and Depth Darkening can be considered as less adequate for precise depth judgments. Nevertheless, we would like to emphasize that they have performed 10% better, with respect to overall depth value range, when compared to the worst performing technique, i.e., Depth of Field. This is not only a significant

result, but it also indicates a big real world difference, which makes Volumetric Halos and Depth Darkening an ideal candidate for communicating depth in general.

Fleming and Bülhoff found that the effects influencing the depth separation of thin semi-transparent filters and translucent objects vary [2005]. Their study has revealed that X-junctions, background visibility and surrounding contrast are not relevant for translucent objects. They further conclude that translucent objects are too complex to perform inverse optics for the understanding, but instead that image statistics for selected regions become more important. This is also in line with the perceptual model proposed by Singh and Anderson, which emphasizes the difference between perceptual transparency and the simplified Metelli model [2002]. We believe that this effect led to the relatively poor performance of DVR which in essence resembles the Metelli effect. Furthermore, while the Volumetric Line Drawings should in principal emphasize the effect of X-junctions, this effect might be much less relevant for volumetric structures according to Fleming and Bülhoff [2005]. Nevertheless, the group with the line techniques Volume Illustration and Volumetric Line Drawings has performed significantly better than Depth of Field and better than DVR, especially Volume Illustration which performed just slightly worse than Volumetric Halos. Thus, we conclude that the addition of lines can support depth perception in volume rendered images. However, as it is inline with our hypothesis H1, these techniques are outperformed by techniques employing image contrast.

While we were first surprised about the poor performance of Depth of Field with respect to depth judgments, we believe that this is due to the inherent *ambiguous depth representation*. While Volumetric Halos and Depth Darkening essentially resemble a ramp of darker values based on depth differences, Depth of Field effects are fundamentally different. The degree of blurriness changes in both directions from the focal plane, back and front, in the same manner. Thus, just by judging the degree of blurriness, it is impossible to decide back or front. Instead it is necessary to take into account the gradient of blurriness, in order to compare the depth of two reference points. In contrast, when using Volumetric Halos and Depth Darkening, the viewer can assume dark-means-deep, as similarly observed also by Langer and Bülhoff [1999].

Shape perception. Based on our analysis of the results achieved in the orientation task, we found that our grouping was not appropriate for this task and we instead performed the analysis per technique. Based on the findings made by Cole et al., who showed the strength of line drawings in the shape perception process [2009] and our hypothesis H2, it was no surprise that Volumetric Line Drawings did perform well. While Volumetric Line Drawing had the lowest angular error it did not perform significantly better than Volumetric Halos or Depth of Field. Also, based on H2 we expected the Volume Illustration technique to perform well, though this technique

scored with the highest mean angular error. In section 3 we paired Volumetric Line Drawings and Volume Illustration in one group and expected them to perform equally well for hypothesis H2, but they are at the opposite ends of the scale, and a significant difference exists between them. When looking at the image results of both these techniques (see Figure 1), it becomes clear that the silhouette enhancing nature of the Volumetric Line Drawings is much more prominent than with Volume Illustration. We believe that this clear contrast with an emphasis on silhouettes and contours supports better inference of the curvature, which can then be extrapolated to the surfaces under investigation. Volumetric Illustration provides less contrast and thus might make this interference more difficult.

Regarding the general low significance of the gauge figure task, we conclude that this type of judgment needs to be improved in the future. While Šoltészová et al. have successfully applied gauge figure tasks in the area of volume rendering [2012], we believe that this application is in general problematic. Especially, when dealing with semi-transparent scenes, surface orientation is hard to perceive and therefore the gauge figure task might be too challenging to be able to achieve reliable results.

Visual appeal. Depth Darkening is the technique which has been rated as most appealing by most participants. But also Volume Illustration and Volumetric Line Drawings have been assessed as more appealing than the other techniques. This is surprising, these techniques result in rather different appearances and are therefore less alike as for instance Volumetric Halos and Depth Darkening.

6.1 Usage Guidelines

For both, the absolute depth and the orientation task, the group containing Volumetric Halos and Depth Darkening has performed best. We believe that this overall good performance does not only support our hypothesis H1, but is also inline with the findings made earlier by Ropinski et al., which state that supporting occlusion as the strongest depth cue can be beneficial in the context of volume rendering [2006]. We see this as an indicator that techniques simulating natural illumination effects might be beneficial. This would also be inline with a study investigating the effect of advanced illumination on volume rendered images, where the authors conclude that directional occlusion shading is most beneficial [Lindemann and Ropinski 2011]. Interestingly the effects achieved by this technique are visually very close to Depth Darkening and Volumetric Halos. However, we should also point out that the instructions may have been biased towards Depth Darkening. As this technique naturally resolves the depth of the first depth layer, and our instructions asked the participants to refer to the first depth in ambiguous cases, this might have influenced the results. Unfortunately, we do not see an alternative which would resolve such a bias, as in the presence of semi-transparency there will always be ambiguous cases.

Based on our experience when generating the study images, and based on the analyzed results, we further conclude that the perception of semi-transparent structures is challenging for the human visual system, and while perceptual rendering enhancements can improve the situation, none of the tested techniques could outperform all other in all tested tasks. We believe that transparency perception is too complex in relation to the few visual cues given in volume rendered images. Nevertheless, we see Depth Darkening and Volumetric Halos as the clear winners of this study, as they proved the best depth perception and good shape perception. However, when requiring highly accurate depth judgments, we conclude that non of the tested techniques are sufficient, as we could not achieve significant results in the ordinal depth test. In such cases, it would be necessary to take into account more quantitative techniques, such as pseudo-chromadepth which has proven beneficial in other stud-

ies [Ropinski et al. 2006; Kersten-Oertel et al. 2014].

7 Conclusions and Future Work

In this paper we have presented results from a large-scale user evaluation, in which we have investigated the perceptual qualities of well known volume rendering techniques in the context of semi-transparent rendering. We have recruited 281 participants which have worked on 22,716 perceptual micro tasks related to depth and shape perception. Furthermore, they were asked to rate the visual appeal of the evaluated techniques. Our findings show that enhancement techniques simulating natural lighting phenomena result in a clear advantage over the other tested techniques. Our group containing of Volumetric Halos and Depth Darkening has performed best for absolute depth and shape perception, which supports our hypothesis H1. As mentioned above, one could argue that the presented evaluations contain a slight bias towards the more surface-aware Depth Darkening technique. Therefore, and due to the fact that it is a true volumetric approach, we recommend to apply Volumetric Halos in order to improve the perceptual qualities of volume rendered images. In cases where mainly shape perception is desired, we conclude based on the findings that Volumetric Line Drawings might be beneficial, which supports our hypothesis H2.

In the future we would like to proceed with our work, and continue investigating the perceptual influence in the context of volume rendering. Interesting could be a combination with stereoscopic depth cues, as well as volumetric illumination models evaluated in previous studies [Lindemann and Ropinski 2011]. While we were only able to cover a limited subset of the available techniques, extending our study to other techniques would also be interesting.

Acknowledgements

This work was supported through grants from the Excellence Center at Linköping and Lund in Information Technology (ELLIIT), the Swedish e-Science Research Centre (SeRC) and have been developed in the Inviwo framework (www.inviwo.org).

References

- AHMED, N., ZHENG, Z., AND MUELLER, K. 2012. Human computation in visualization: Using purpose driven games for robust evaluation of visualization algorithms. *IEEE TVCG* 18, 12.
- APPEL, A., ROHLF, F. J., AND STEIN, A. J. 1979. *The haloed line effect for hidden line elimination.*, vol. 13. ACM.
- BAER, A., GASTEIGER, R., CUNNINGHAM, D., AND PREIM, B. 2011. Perceptual evaluation of ghosted view techniques for the exploration of vascular structures and embedded flow. *Computer Graphics Forum* 30, 3, 811–820.
- BAIR, A., AND HOUSE, D. 2007. Grid with a view: Optimal texturing for perception of layered surface shape. *IEEE TVCG* 13, 6, 1656–1663.
- BORKIN, M. A., VO, A. A., BYLINSKII, Z., ISOLA, P., SUNKAVALLI, S., OLIVA, A., AND PFISTER, H. 2013. What makes a visualization memorable? *IEEE TVCG* 19, 12.
- BOUCHENY, C., BONNEAU, G.-P., DROULEZ, J., THIBAUT, G., AND PLOIX, S. 2009. A perceptive evaluation of volume rendering techniques. *ACM TAP* 5, 4, 23:1–23:24.
- BOUSSEAU, A., O’SHEA, J. P., DURAND, F., RAMAMOORTHY, R., AND AGRAWALA, M. 2013. Gloss perception in painterly and cartoon rendering. *ACM TOG* 32, 2, 18:1–18:13.

- BRUCKNER, S., AND GRÖLLER, M. 2007. Enhancing depth-perception with flexible volumetric halos. *IEEE TVCG* 13, 6.
- BRUCKNER, S., GRIMM, S., KANITSAR, A., AND GRÖLLER, M. E. 2005. Illustrative context-preserving volume rendering. In *EG/IEEE VGTC Symposium on Visualization*, 69–76.
- BURNS, M., KLAWE, J., RUSINKIEWICZ, S., FINKELSTEIN, A., AND DECARLO, D. 2005. Line drawings from volume data. *ACM TOG* 24, 3, 512–518.
- CHAN, M.-Y., WU, Y., AND QU, H. 2007. Quality enhancement of direct volume rendered images. In *Proceedings of the Sixth Eurographics / IEEE VGTC Conference on Volume Graphics*.
- CHAN, M.-Y., WU, Y., MAK, W.-H., CHEN, W., AND QU, H. 2009. Perception-based transparency optimization for direct volume rendering. *IEEE TVCG* 15, 6, 1283–1290.
- COLE, F., SANIK, K., DECARLO, D., FINKELSTEIN, A., FUNKHOUSER, T., RUSINKIEWICZ, S., AND SINGH, M. 2009. How well do line drawings depict shape? *ACM TOG* 28, 3.
- DE MOURA PINTO, F., AND DAL SASSO FREITAS, C. M. 2011. Illustrating volume data sets and layered models with importance-aware composition. *Vis. Comput.* 27, 10, 875–886.
- DÍAZ, J., AND VÁZQUEZ, P. 2010. Depth-enhanced maximum intensity projection. In *IEEE/EG Volume Graphics*, 93–100.
- ENGLUND, R., KOTTRAVEL, S., AND ROPINSKI, T. 2016. A crowdsourcing system for integrated and reproducible evaluation in scientific visualization. In *IEEE Pacific Visualization Symposium*, IEEE, 40–47.
- FLEMING, R. W., AND BÜLTHOFF, H. H. 2005. Low-level image cues in the perception of translucent materials. *ACM TAP* 2, 3.
- FULVIO, J. M., SINGH, M., AND MALONEY, L. T. 2006. Combining achromatic and chromatic cues to transparency. *Journal of Vision* 6, 8, 1.
- GLEICHER, M., CORRELL, M., NOTHELFER, C., AND FRANCONERI, S. 2013. Perception of average value in multiclass scatterplots. *IEEE TVCG* 19, 12, 2316–2325.
- GROSSET, A., SCHOTT, M., BONNEAU, G.-P., AND HANSEN, C. D. 2013. Evaluation of depth of field for depth perception in dvr. In *IEEE Pacific Visualization Symposium*, 81–88.
- HEER, J., AND BOSTOCK, M. 2010. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *ACM Human Factors in Computing Systems (CHI)*, 203–212.
- HIBBARD, B. 2000. Confessions of a visualization skeptic. In *ACM SIGGRAPH*, 11–13.
- INTERRANTE, V., FUCHS, H., AND PIZER, S. M. 1997. Conveying the 3d shape of smoothly curving transparent surfaces via texture. *IEEE TVCG* 3, 2, 98–117.
- JÖNSSON, D., SUNDÉN, E., YNNERMAN, A., AND ROPINSKI, T. 2014. A survey of volumetric illumination techniques for interactive volume rendering. *Computer Graphics Forum* 33, 1.
- KERSTEN-OERTEL, M., CHEN, S. J.-S., AND COLLINS, D. 2014. An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *IEEE TVCG* 20, 3.
- KIM, S.-H., YUN, H., AND YI, J. S. 2012. How to filter out random clickers in a crowdsourcing-based study? In *Proceedings of the 2012 BELIV Workshop*, ACM, 15.
- KOENDERINK, J. J., VAN DOORN, A. J., AND KAPPERS, A. M. 1992. Surface perception in pictures. *Perception & Psychophysics* 52, 5, 487–496.
- LANGER, M. S., AND BÜLTHOFF, H. H. 1999. Depth discrimination from shading under diffuse lighting. *Perception* 29, 6.
- LINDEMANN, F., AND ROPINSKI, T. 2011. About the influence of illumination models on image comprehension in direct volume rendering. *IEEE TVCG* 17, 12, 1922–1931.
- LUFT, T., COLDITZ, C., AND DEUSSEN, O. 2006. Image enhancement by unsharp masking the depth buffer. *ACM TOG* 25, 3, 1206–1213.
- MORA, B., AND EBERT, D. S. 2004. Instant volumetric understanding with order-independent volume rendering. *Computer Graphics Forum* 23, 3, 489–497.
- NAGY, Z., AND KLEIN, R. 2004. High-quality silhouette illustration for texture-based volume rendering. *Journal of WSCG* 12, 2, 301–308.
- QUINN, A. J., AND BEDERSON, B. B. 2011. Human computation: a survey and taxonomy of a growing field. In *ACM Human Factors in Computing Systems (CHI)*, 1403–1412.
- RHEINGANS, P., AND EBERT, D. 2001. Volume illustration: Non-photorealistic rendering of volume models. *IEEE TVCG* 7, 3.
- ROPINSKI, T., STEINICKE, F., AND HINRICHS, K. 2006. Visually Supporting Depth Perception in Angiography Imaging. In *Smart Graphics*, 93–104.
- SCHOTT, M., PASCAL GROSSET, A., MARTIN, T., PEGORARO, V., SMITH, S. T., AND HANSEN, C. D. 2011. Depth of field effects for interactive direct volume rendering. *Computer Graphics Forum* 30, 3, 941–950.
- SINGH, M., AND ANDERSON, B. L. 2002. Toward a perceptual theory of transparency. *Psychological review* 109, 3, 492.
- SOLTESZOVA, V., TURKAY, C., PRICE, M., AND VIOLA, I. 2012. A perceptual-statistics shading model. *IEEE TVCG* 18, 12.
- SUNDÉN, E., STENETEG, P., KOTTRAVEL, S., JÖNSSON, D., ENGLUND, R., FALK, M., AND ROPINSKI, T., 2015. Inviwo - An Extensible, Multi-Purpose Visualization Framework. Poster at IEEE Vis.
- SVAKHINE, N., EBERT, D., AND ANDREWS, W. 2009. Illustration-inspired depth enhanced volumetric medical visualization. 77–86.
- THOMPSON, W., FLEMING, R., CREEM-REGEHR, S., AND STEFANUCCI, J. K. 2011. *Visual Perception from a Computer Graphics Perspective*. A K Peters.
- VIOLA, I., KANITSAR, A., AND GRÖLLER, M. E. 2005. Importance-driven feature enhancement in volume visualization. *IEEE TVCG* 11, 4, 408–418.
- ŠOLTÉSZOVÁ, V., PATEL, D., AND VIOLA, I. 2011. Chromatic shadows for improved perception. In *Proceedings of Non-Photorealistic Animation and Rendering*, 105–115.
- WARE, C. 2012. *Information visualization: perception for design*. Elsevier.
- ZHENG, L., WU, Y., AND MA, K.-L. 2013. Perceptually-based depth-ordering enhancement for direct volume rendering. *IEEE TVCG* 19, 3, 446–459.