DOI: 10.1111/cgf.13320

COMPUTER GRAPHICS *forum*
Volume 37 (2018), number 6 pp. 174–187

Quantitative and Qualitative Analysis of the Perception of Semi-Transparent Structures in Direct Volume Rendering

R. Englund[1] and T. Ropinski[2]

[1]Scientific Visualization Group, Department of Science and Technology, Linköping University, Sweden
rickard.englund@liu.se
[2]Visual Computing Group, Ulm University, Germany
timo.ropinski@uni-ulm.de

Abstract

Direct Volume Rendering (DVR) provides the possibility to visualize volumetric data sets as they occur in many scientific disciplines. With DVR semi-transparency is facilitated to convey the complexity of the data. Unfortunately, semi-transparency introduces challenges in spatial comprehension of the data, as the ambiguities inherent to semi-transparent representations affect spatial comprehension. Accordingly, many techniques have been introduced to enhance the spatial comprehension of DVR images. In this paper, we present our findings obtained from two evaluations investigating the perception of semi-transparent structures from volume rendered images. We have conducted a user evaluation in which we have compared standard DVR with five techniques previously proposed to enhance the spatial comprehension of DVR images. In this study, we investigated the perceptual performance of these techniques and have compared them against each other in a large-scale quantitative user study with 300 participants. Each participant completed micro-tasks designed such that the aggregated feedback gives insight on how well these techniques aid the user to perceive depth and shape of objects. To further clarify the findings, we conducted a qualitative evaluation in which we interviewed three experienced visualization researchers, in order to find out if we can identify the benefits and shortcomings of the individual techniques.

Keywords: scientific visualization, volume visualization

ACM CCS: Computing methodologies → Perception; Human-centred computing → Scientific visualization

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 [BBD*09]. 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 [Hib00]. Thus, while semi-transparency is able to reduce the occlusion of internal structures, it does not necessarily support perceptually resolving inter-object relations. Accordingly, transfer functions as used in the context of Direct Volume Rendering (DVR) usually provide an opacity mapping functionality. Such opacity mapping can on the one hand be used to mask out structures by assigning zero opacity to them. But on the other hand, opacity mapping can also be used to visualize selected structures in a semi-transparent manner, and thus bring out internal structures.

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 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, in previous work, 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 [ER16]. This paper is an extended version of [ER16] in which we present the conducted evaluation from the original paper and relate the obtained results to newly acquired expert feedback, which we have gathered from three international visualization researchers, which each have work for more than 10 years with volumetric data.

The compared 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 in which these techniques have been first described. Thus, we compare standard DVR to Depth of Field rendering [SPGM*11], Depth Darkening [SEA09], Volumetric Halos [BG07], Volume Illustration [RE01] and Volumetric Line Drawings [BKR*05]. An example for the application of each of these techniques is shown in Figure 1, where we apply them to a data set obtained in the area of material science. 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 hypotheses, which are based on Ware's discussion on Depth and Shape perception [War12] and will be discussed in Section 4.3:

- 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.

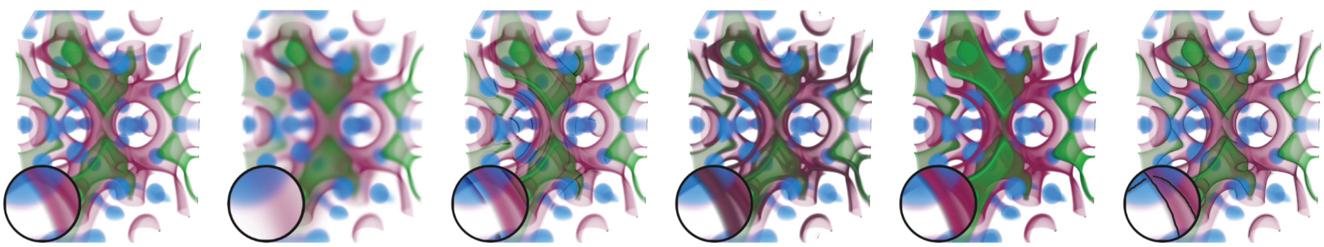
As we were additionally interested in the subjective appeal of the tested techniques, we have also conducted a subjective image comparison task. By exploiting crowdsourcing through the Amazon Mechanical Turk (MTurk) platform, we could include the feedback of 281 participants in our study, which has been used to acquire a total of 22 716 perceptual judgements. Based on the findings of an initial quantitative perceptual evaluation (PEA), we have performed an additional quantitative perceptual evaluation (PEB) for clarification. Finally, we have gathered expert feedback in a qualitative evaluation quantitative evaluation (QE), and linked the obtained feedback to the quantitative results.

In the remainder of this paper, we first discuss work related to our evaluation in Section 2. Section 3 first explains the selection criteria applied to identify the evaluated techniques, before discussing the individual techniques in more detail. Section 4 describes the setup and execution of the perceptual quantitative evaluations and discusses and analyses the results. The qualitative expert evaluation is presented and discussed in Section 5, whereby we interlink all made findings and draw conclusions. In Section 6, we propose guidelines for the effective visualization of semi-transparent structures in volume rendered images based on the quantitative and qualitative evaluations. Finally, we summarize the most important aspects of this paper and present some open questions for future work in Section 7.

As mentioned, this is an extended version of [ER16]. The sections describing the evaluated techniques (Section 3) and the quantitative evaluation (Section 4) is reused from the original submission. Based on the results in the original paper, we have conducted a qualitative user evaluation (Section 5) which strengthens the conclusions and user guidelines defined in that paper.

2. Related Work

We have structured the related work into three categories, which are discussed separately. Firstly, 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.



(a) DVR (b) Depth of Field (c) Depth Darkening (d) Volumetric Halos (e) Volume Illustration (f) Volumetric Lines

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 [SPGM*11], (c) Depth Darkening [SEA09], (d) Volumetric Halos [BG07], (e) Volume Illustration [RE01] and (f) Volumetric Line Drawings [BKR*05].*

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 included 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 [NK04]. As our study focusses 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 [DV10] or order-independent enhancements [ME04].

Other approaches of interest for semi-transparency perception belong to the class of focus-and-context techniques [VKG05]. Together with approaches borrowed from shading computation to modulate transparency [BGKG05, dMPDSF11], 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.* [BGKG05]. 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 [CWQ07, CWM*09, ZWM13].

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 [BBD*09], more recent studies also collect continuous feedback with point-based judgement trials [LR11, GSBH13]. Our study is inspired by these evaluations, and we use similar depth trials: ordinal and absolute depth judgements. While some of the previous evaluations of the depth perception of volume rendered images have been conducted with a special domain focus [RSH06, KOCC14], 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. [BGCP11, vTPV12]. 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 [IFP97]. More recently, Bair and House have conducted similar experiments investigating the shape perception of layered grid-textured surfaces [BH07]. 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.

While the aforementioned evaluations were mainly focusing on monoscopic images, recently several evaluations have investigated the impact of stereoscopy [KSTE06, CDW*12, MS13, WPF12, DRN*17], which we did not consider 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 [QB11]. Since its emergence several perceptual studies have been conducted in the graphics community, whereby among others the shape perception of line renderings [CSD*09], and the perception of materials in illustrative graphics has been investigated [BOD*13]. Recently, it could also be shown that crowdsourcing can be an effective way of investigating the perceptual capabilities of visualizations [HB10], which led to the conduction of several interesting crowdsourcing studies in visualization. Prominent examples include the perception of averages in multi-class scatter plots [GCNF13], but also the memorability of visualizations [BVB*13]. 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.* [AZM12]. 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 [EKR16]. They also showed that crowdsourcing can be used as an alternative to lab-based evaluation by replicating previously published, lab-based evaluations [LKJ*05, LR11].

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. [DV10], 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 parametrizations, such as for instance the texture-enhancement proposed by Interrante *et al.* [IFP97]. 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 [JSYR14], 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 [dMPDSF11] or a focus region [VKG05] have also been excluded.

While we have constrained the impact of parameters to be able to generalize our findings, we have furthermore eliminated those techniques which took over parameter control. This decision has been made in order to focus on techniques that support interactive exploration. Only when the user can directly set the DVR parameters, such as transfer function and camera, a direct exploration becomes possible. Therefore, we did not include techniques in our study which automatically alter these parameters, for instance by optimizing the transfer function [ZWM13] or adapting the current camera view [VFSG06]. 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 [TFCRS11], p. 261]. This left us six volume rendering techniques to be evaluated. Each technique described in brief bellow and a visual comparison of them is shown in Figure 1. For a more detail description of each technique, see the appendix in the supplemental material.

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 [SPGM*11]. 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 [LCD06], 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 [SEA09], 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 [BG07], 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 [RE01]. 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 [BKR*05]. We incorporate this technique and extend it by introducing the haloed line effect presented by Appel *et al.* [ARS79], as the halos emitted by the lines are expected to help solving the ambiguity of T and X-junctions [TFCRS11], p. 258].

Based on their individual capabilities, we have sorted the six selected techniques into four groups by keeping our hypotheses 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 aim at improving 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. Quantitative Evaluation

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 Amazon's Mechanical Turk, 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. A total of 281 participants have completed the study and submitted their data.

4.1. Evaluation setup

The tasks for each participant were split into four groups, whereby each group required a participant to fulfil a different type of task. As we are interested in investigating the tested techniques' impact on scene perception, two of the perceptual task types were focused on depth perception, and one perceptual task type was focused on perceiving orientations, an essential part of shape perception [CSD*09]. Furthermore, we were interested in the subjective visual appeal of the tested techniques, and have therefore added the subjective image comparison. In evaluation PEA the participants had to perform all four task types, which were designed to investigate the perception of relative depth, the perception of absolute depth, the perception of orientation, as well as the visual appeal of an image. In evaluation PEB, where we were focusing on the top four ranked techniques identified through PEA, we have exposed the participants only to three perceptual tasks, as we have dropped the visual appeal comparison. During both evaluations, the order of the task groups as well as the individual visual stimuli were randomized.

As stated above, we have decided to not put any time constraints during the conducted evaluations. Time constraints are often applied in comparable studies in order to rule out the influence of cognitive processes [BBD*09]. As we were interested in investigating the six techniques in the context of a data exploration scenario, we felt that time constraints would be rather artificial and thus might distort the results. Consequently, we did not bring up any time-related discussion in the study instructions, and let the users take as much time as they wanted.

4.1.1. *Visual stimuli*

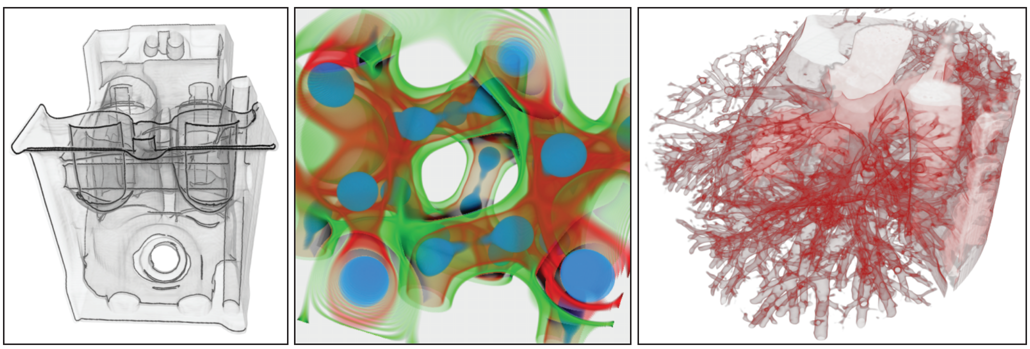
During our study, we have confined ourselves to the evaluation of static images. The main reasoning behind this decision is the goal to avoid that motion-induced parallax effects overrule other depth cues, introduced by the investigated techniques. Such an observation has been made by Boucheny *et al.*, who report that DVR has been ambiguous in static cases, but that this effect could be counterbalanced with motion parallax [BBD*09]. 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 at a unique cue to depth is also in line 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 [KOCC14]. 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 [SSK*15] and all tested images 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. Figure 2 shows a few sample images, which have been used in the study. 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.

4.1.2. *Perceptual tasks*

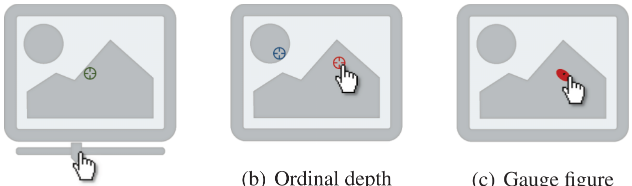
The perceptual tasks have been designed to investigate depth perception, shape perception and visual appeal of the tested techniques. To validate our study setup, we have initially conducted a pilot study with 20 participants. In the following, we will discuss the four task types in detail and refer to how the pilot study has influenced their final design. As most tasks are location-specific, we have used image markers to point out the image area to be judged. For these markers, we have manually selected locations with the goal to create a uniform distribution, both spatially within the images and with respect to actual depth values. Thus, we could maintain 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 colour has been selected to maximize the contrast with the background. To obtain knowledge about the duration of the decision processes, we have also measured the time in milliseconds for each task.

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 3(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. Once the slider had been moved, the participant could proceed to the next



(a) CT scan / Volumetric Line Drawings (b) Simulation / Depth Darkening (c) CT scan / DVR

Figure 2: *Exemplar images as they have been used in the conducted user studies. The data sets have been selected to show familiar (a) and more unfamiliar structures (b, c). Also the degree of density with respect to the represented features varies from low (c) to high (a, b). Transfer functions have been designed that they capture more homogeneous structures (b, c), but also semi-transparent surface-like boundaries (a).*



(a) Absolute depth (b) Ordinal depth (c) Gauge figure

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

task by pressing a button below the visual stimulus. Our pilot study indicated, that this task was hard to understand. Therefore, we gave detailed instructions which were shown under each task setup. A screen grab of the task as seen by the participants can be found in the appendix in the supplementary materials.

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 3(b) each visual stimulus was shown together with two coloured markers. The task was to select the marker closest to the viewer. 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. A screen grab of the task as seen by the participants can be found in the appendix in the supplementary materials.

Shape Perception. To investigate the impact on shape perception of the six tested techniques, we have asked the participants to complete a gauge figure task [KVDK92]. 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 [CSD*09]. Furthermore, some of the observations made by Šoltészová *et al.* when using gauge figure tasks in the area of volume rendering [vPV11] were helpful. Thus, we exposed the participants to visual stimuli, which had a randomly oriented gauge figure located at a pre-defined location (see Figure 3c) 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 judgement, the participants could access these examples which would show a correct and an incorrect gauge orientation for an easy to comprehend shape. A screen grab of the task as seen by the participants can be found in the appendix in the supplementary materials. To not let colour choices influence the orientation of the gauge figure, we have decided to always depict it with the same colours. Furthermore, while Šoltészová *et al.* have used a transparent base for the gauge figure to improve the visibility of the underlying surface [vPV11], we have decided to stick to the gauge figure design exploited by Cole *et al.* [CSD*09], as it makes the perception of the gauge orientation more obvious.

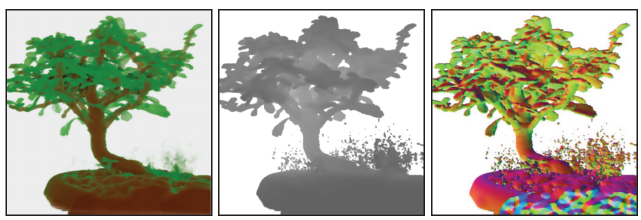
Visual Appeal. To investigate the visual appeal of the tested techniques, we have asked the participants to select the most appealing picture in a two alternative forced choice test. The two alternatives were shown side by side, and the user could select the preferred image by directly selecting it with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

4.1.3. *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 while taking part in the evaluation. A screen grab of these two pages can be found in the appendix in the supplementary materials.

4.1.4. *Result analysis*

Since we were dealing with a vast number of perceptual tasks, an automatic result analysis process was crucial. Therefore, we have also created depth and normal images when generating the visual stimuli used in our evaluations, such that we could determine the *correctness* of a perceptual task automatically. Examples for these images are shown in Figure 4. 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



(a) DVR (b) Depth (c) Gradients

Figure 4: *To enable an automated results analysis, we have created and stored additional information for each stimuli. Here, we show the depth map (b) as well as a map colour coding the gradients as computed with central differences (c) for the DVR rendering of the bonsai data set (a). The depth maps enable us to automatically judge the correctness for the absolute and ordinal depth tasks, while the normal map is used to judge the correctness of the gauge figure task.*

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.

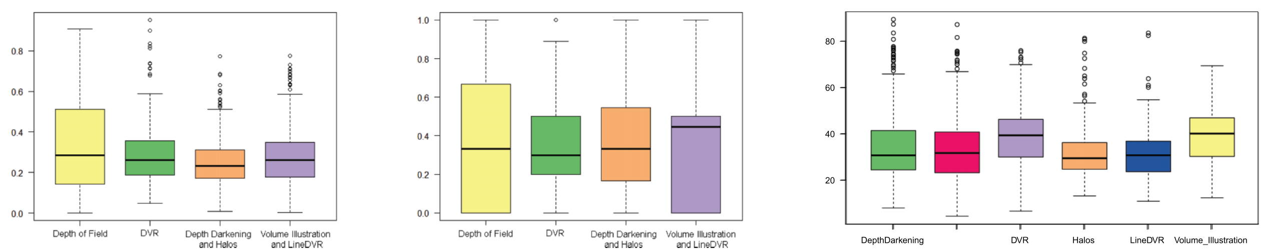
When using crowdsourcing to conduct user evaluations, we heavily decrease the workload of participant recruitment compared to lab-based evaluations. Though, this comes with a slight cost, since participants are recruited over the Internet we do not have any control over their environment, such has screen 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 [KYY12]. Another potential issue of 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 such random clickers in our studies, 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. Secondly, to discard participants who were making random decisions 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%). Based on the functionality provided by Amazon, 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 response data available after discarding the identified participants, we have analysed 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.

4.2. Quantitative evaluation results

To evaluate the six techniques, we have recruited 300 participants. The data have 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 judgements, which took in average 16.5 min for a payment of \$1.05. In the second wave, we have omitted the visual appeal judgement, and 200 participants had to perform 120 judgements each, which took in average 26 min and brought them \$2.00. Of the 300 recruited participants, 19 did not complete the study resulting in 281 participants. Thus, in total we have collected 22 716 judgements.

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 5(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



(a) Average error of the absolute depth perception results.

(b) Average correctness of the ordinal depth perception task.

(c) Average angular error in degrees of the shape perception task.

Figure 5: *Boxplots showing the results for the two depth perception tasks and the shape orientation task.*

gave a power value of 89%. Accordingly, we did not perform the Tukey *post hoc* tests, but refer to Figure 5(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 [$F(3, 173) = 0.466$, $p = 0.707$].

Shape Perception Task. The rANOVA showed a significant difference for 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 Halos where 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 analysed groups are shown in Figure 5(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 analysing 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).

4.3. Quantitative result discussion

In the presented evaluations, 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.* [BBD*09], which focuses on achromatic transparency, we have studied the impact on coloured images, a procedure backed up by the findings of Fulvio *et al.* [FSM06]. Looking at the results by keeping our two hypotheses 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 judgements. 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 Bülthoff found that the effects influencing the depth separation of thin semi-transparent filters and translucent objects vary [FB05]. 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 enable the viewer 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 [SA02]. 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 Bülthoff [FB05]. 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 in line 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 judgements, 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 Bülthoff [LB99].

Investigating the bad performance of the Volumetric Line Drawing makes us think that the mental fusion of the lines and the volumetric structures might be cognitively challenging. Therefore, in the future we would like to experiment with coloured lines, whereby the colour tones adapt to the colours used in the enhanced volume. We would thus expect an effect similar to Volume Illustration, which has performed significantly better in our evaluation. Thus, we might see an influence of the frequency of the line effects, as the

Volumetric Lines would be rather high frequent effects as compared to the smoother nature of the Volume Illustration technique.

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 [CSD*09] 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 judgement needs to be improved in the future. While Šoltészová *et al.* have successfully applied gauge figure tasks in the area of volume rendering [vTPV12], 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.

5. Qualitative Evaluation

So far we have performed and discussed quantitative studies to analyse our hypotheses H1 and H2. As stated in the beginning of this paper, these hypotheses have been derived from the literature. Since these sources were not dealing with volume rendering in particular, but rather the representation of 3D structures in general, we wanted to clarify what specifics might be different in the area of volume rendering. To do so, we have conducted a qualitative evaluation with three experienced visualization researchers.

5.1. Evaluation setup

The three experts (E1, E2, E3), have published in different subareas of volume rendering since 12 (E3) and 13 (E1+E2) years. The respective subareas are focus+context techniques (E1), volumetric illustrations (E1+E3) as well as volumetric illumination models (E2+E3). To obtain their feedback, we have conducted formalized interviews, in which the experts were exposed to volume rendered images which have been generated with the six described techniques. In the beginning of the interview, we have briefly introduced the experts to our goal, i.e. the evaluation of volume rendered images showing semi-transparent structures, and have shown them sample images, where all different rendering techniques where shown on the same screen in two rows with three images each. Thus, a direct side-by-side comparison of the rendering techniques was possible. To not introduce a bias, we have neither explained the individual techniques, nor have we indicated which images were rendered with which techniques. However, as the three experts are knowledgeable in the field, we cannot guarantee that they were not able to identify the techniques themselves. After the introduction, the three experts have been asked the following four questions in the given order during the actual interview process:

- Q1: Which technique do you believe does communicate semi-transparent structures best?
- Q2: Can you point out most easy and most difficult areas with respect to perception?
- Q3: What properties do you think are essential to communicate semi-transparent structures?
- Q4: What techniques in your field would you rate as most promising to communicate semi-transparent structures?

5.2. Expert feedback

In this section, we briefly describe the feedback obtained from each of the three experts, before we discuss the feedback and link it to the findings made in the quantitative studies.

Expert Feedback E1. Regarding question Q1, E1 has requested additional information regarding the question. He has asked if we are interested in ordinal distance, but has stated that he has also taken into account the *possibility to see the most* in his reasoning. Based on these assumptions, he has chosen Volume Illustration as the technique, he feels communicates the overall 3D structure the best. He also comments that the resulting density of the emphasized silhouettes actually results in some occlusion, and thus reveals less of the overall structure, but that the first layers are clearly arranged. Thus, according to E1, while Volume Illustration depicts less layers, the visible layers are *more clearly arranged*. Regarding Q2, E1 has expressed that for the images shown in Figure 1, the top left region showing the green structure reaching out from the middle would be the best discriminative feature, which would be much more difficult to interpret in the other representations. When he is confronted with Q2, he feels that it is important to also point out that there are some structures which are hard to see, or even omitted, when applying Volume Illustration. In contrast, they are sometimes visible in the other representations, due to the less accumulated opacity. Regarding Q3, E1 has stated that the overall density, and less transparency, was the most beneficial property of the selected technique. He explains his opinion by referring to the Volumetric Line Drawing technique, were he could derive the ordering from the silhouettes, and then needs to *infer* that ordering to the structures, which he found more cumbersome. Asking for suggestions for Q4, E3 points out that an animation showing the build-up of the semi-transparent structures from back-to-front could be helpful. Furthermore, he refers to parallax effects which could be exploited to improve the perception of semi-transparent structures, as he stated that *movement gives me a lot of depth cues in addition*.

Expert Feedback E2. With respect to Q1, E2 judges that the Depth Darkening best communicates the 3D structure. He justifies this opinion with the *enhanced edges and the shadows* which support him in the perception of the overall structure. Despite not being asked, E2 also relates DVR as second best and Volumetric Line Drawings as third best. For the Volumetric Line Drawings, he also points out that the enhanced edges are particularly beneficial. Regarding Q2, E2 refers to particular objects he cannot unambiguously depth sort. In Figure 1, these are the small blue spheres in the centre of the images, which are difficult to perceive with the other techniques. While E2 has already justified his selection of the most effective techniques by referring to enhanced edges and shadows, he confirms this also in the context of Q3. When asked for other potentially helpful approaches in Q4, he suggests to encode the opacity with colour or modulate the brightness with respect to depth, but again states that he likes the shadows as perceived in the Depth Darkening approach.

Expert Feedback E3. E3 has replied to Q1 by also selecting Volume Illustration as the technique, which communicates the 3D structures the best. When asked for problematic structures, he pointed out the same structure in the images shown in Figure 1 as did expert E1. With respect to Q3, he replies that the *enhancement near the edges helps to clarify occlusion relationships which makes it easier to distinguish which structures are in front or in the back*. He also points out that the Volumetric Line Drawings push this even further, but that he finds the shown lines less revealing, as it is *partially distracting since they introduce too much additional clutter in the image*. With respect to question Q4, E3 suggests the use of halos as published by Bruckner and Gröller [BG07]. After pointing out that one of the techniques actually uses this approach, E3 closely inspects the used parameters. Being asked about the optimal parameter setup for this technique, he answers that one aspect for halos in particular, halo colours can be used to complement the contrast in the image. So if edges go to the darker colour ranges, white halos might be a better choice. He further suggests to investigate if all depth relations are equally important, as there might be some salient structures that require better depth communication, while this could be less important for contextual structures. He suggest to distinguish these structures in order to use the visual channels efficiently, as any variance adds additional information which *potentially guides the viewer to less relevant structures*. Asking how to select the relevant structures, he suggest the usage of predictors, such as salience computation techniques.

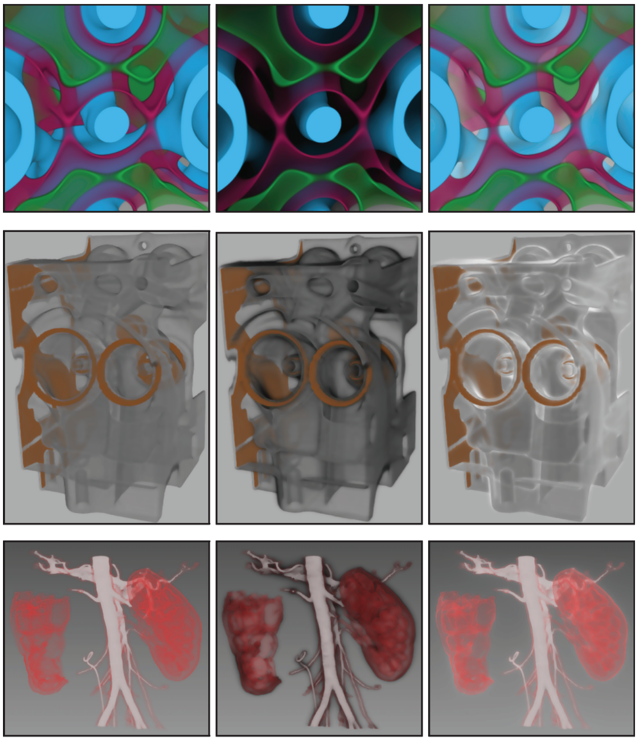
5.3. Discussion

In this section, we discuss the feedback received by the three experts and relate it to our hypothesis as well as the findings made in our quantitative studies. While E1 and E3 select Volume Illustration as the most promising technique, E2 has selected Depth Darkening. Although these techniques are considerably different, they both share the property that enhancements are done along edges. Consequently, all three experts have pointed out the edges when asked about the most important features during Q3. We believe that this is another indicator that H2 might hold, i.e. that shape perception is improved when the curvature of the shape can be inferred at nearby silhouettes. With respect to hypothesis H1, i.e. depth perception is improved when back-to-front relations are encoded unambiguously through image contrast, we see the feedback of E1 and E2 as supporting evidence. While E2 especially emphasizes, besides the edges, the shadow-like nature of depth darkening, E1 in particular refers to the higher density at edges in the Volume Illustration technique. We see both these properties as aspects which increase the contrast, and thus are in line with hypothesis H1. However, the comments made by E1 and E3 go even beyond what we have hypothesized in H1 and H2. Both explicitly stated that the association between edges and structures (E1) as well as the integration of the edges into the structure (E3) are of high relevance. We also feel that these are important aspects, especially for volume rendered images. We further suspect that the Volumetric Line Drawing is problematic with respect to this property, as lines are extracted and merged with the surfaces without working on the integration. In a post-interview discussion with E3, we have further elaborated on this aspect. As he felt that the lines in the Volumetric Line Drawing even better emphasize the edges, we have discussed the option to use these lines as an initial edge indicator, which can be low pass filtered to better merge it with the structures. This would also lead to less clutter in this technique.

As clutter has not only been brought up by E3, but also by E1, we feel that its avoidance also plays an essential role. E1 in particular preferred the Volume Illustration approach, as it is more dense, and communicates fewer layers, which can be perceived more clearly. Thus, it should be thought about how this inherent conflict, communicating semi-transparent structures with less transparency, can be solved. While Volume Illustration seems already to be a step in the right direction, we feel that the comments raised by E3 could be important in the future. Thus, a modulation with respect to some degree of interest function could be relevant to reduce clutter, while still optimizing transparency perception for the relevant structures. Although we wanted to avoid such a technique, which takes over the control over the rendering parameters, within our studies, such techniques should probably be investigated in the future.

Based on the comments of E3 with respect to the parameters of the Volumetric Halos, we have investigated setting the halos to white for a few selected data sets. The obtained results are shown in Figure 6, where we show the original DVR (a), black halos (b) and white halos (c). As can be seen in the images and also suggested by expert E3, the contrast-enhancing effects of the white halos are in particular obvious when rendering dark objects. In the contrary, they are less prominent towards the brighter background. Furthermore, naturally the with halos reveal more features of the internal structures, which are dimmed out by the black halos. Interestingly, this dimming out can be related to the focus on the front layers, positively pointed out by E1 for Volume Illustration. Due to these varying effects, a combination of different halo colours in the same image might be interesting to be investigated in the future.

Finally, we would like to point out, that the comments of E1 with respect to parallax effects has ensured us, that the focus on static images within this study was the right decision. Since all techniques



(a) DVR (b) Black halos (c) White halos

Figure 6: *As suggested by expert E3, we have experimented with white halos. The images show three example scenes rendered with DVR (a), black halos as used in our user studies (b) and white halos as suggested by E3. As stated by E3 the contrast-enhancing effects of the white halos are in particular obvious when rendering dark objects. In the contrary, they are less prominent towards the brighter background.*

have rather similar properties, movement as additional depth cue might have overpowered all other effects.

6. Usage Guidelines

To make our findings applicable, we have derived usage guidelines for the visualization of semi-transparent structures in volume rendered images. 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 in line 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 [RSH06]. We see this as an indicator that techniques simulating natural illumination effects might be beneficial. This would also be in line with a study investigating the effect of advanced illumination on volume rendered images, where the authors conclude that directional occlusion shading is most beneficial [LR11]. 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 analysed 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 semi-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 judgements, we conclude that none 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 studies [RSH06, KOCC14].

While Volume Illustration was for most of the conducted tasks in the top ranked group, we have experienced that this approach works far better for imaged data as compared to simulation data. In the latter, the nearly uniform distribution of gradient magnitudes makes it difficult to emphasize relevant structures and thus leads to less prominent effects. However, we feel that the Volume Illustration technique resembles the effect of X-junctions rather closely [And97], which might lead to an improved perception for surface-like structures.

7. Conclusions and Future Work

In this paper, we have presented results from two large-scale user evaluation, in which we have quantitatively investigated the perceptual qualities of well-known volume rendering techniques in the context of semi-transparent rendering. Therefore, 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 based on our quantitative findings. 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. Interestingly, these recommendations could not be confirmed with the expert feedback acquired in our qualitative evaluation. There, two of three have judged that Volume Illustration most effectively communicates the 3D structure of the shown objects. As we did not ask the experts to focus on specific regions, but judge the images as a whole, this indicates that there is more to understanding 3D structure than simply integrating depth and shape judgements at individual positions. We feel that this nicely underlines that often observed effects of Gestalt laws used in visual perception.

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 [LR11]. Since it has been shown in previous studies related to volume rendering that dynamic images has a significant impact on depth perception [BBD*09], due to the parallax effect, we limited ourself to static images. However, it would be interesting to investigate how the depth and shape perception is affected by introducing motion as an additional cue to these techniques. While we were only able to cover a limited subset of the available techniques, extending our study to other techniques would also be interesting.

But also other measures should be taken into account in the future. It would be for instance interesting to investigate the influence on size perception, but also the effects for colour-deficient observers. Finally, we would be interested in how our results correlate with automatic image quality measures [WQC*10]. This could then also be extended to the automatic finding of these measures, which would be a first step towards synthesizing new visualization algorithms based on large-scale user feedback.

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 (http://www.inviwo.org).

References

[And97] Anderson B. L.: A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception 26* (1997), 419–454.

[ARS79] Appel A., Rohlf F. J., Stein A. J.: The haloed line effect for hidden line elimination. *ACM SIGGRAPH Computer Graphics 13* (1979), 151–157.

[AZM12] Ahmed N., Zheng Z., Mueller K.: Human computation in visualization: Using purpose driven games for robust evaluation of visualization algorithms. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2104–2113.

[BBD*09] Boucheny C., Bonneau G.-P., Droulez J., Thibault G., Ploix S.: A perceptive evaluation of volume rendering techniques. *ACM Transactions on Applied Perception 5*, 4 (2009), 23:1–23:24.

[BG07] Bruckner S., Gröller M.: Enhancing depth-perception with flexible volumetric halos. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1344–1351.

[BGCP11] Baer A., Gasteiger R., Cunningham D., Preim B.: Perceptual evaluation of ghosted view techniques for the exploration of vascular structures and embedded flow. *Computer Graphics Forum 30*, 3 (2011), 811–820.

[BGKG05] Bruckner S., Grimm S., Kanitsar A., Gröller M. E.: Illustrative context-preserving volume rendering. In *Proceedings of EG/IEEE VGTC Symposium on Visualization* (Leeds, UK, 2005), pp. 69–76.

[BH07] Bair A., House D.: Grid with a view: Optimal texturing for perception of layered surface shape. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1656–1663.

[BKR*05] Burns M., Klawe J., Rusinkiewicz S., Finkelstein A., DeCarlo D.: Line drawings from volume data. *ACM Transactions on Graphics 24*, 3 (2005), 512–518.

[BOD*13] Bousseau A., O'shea J. P., Durand F., Ramamoorthi R., Agrawala M.: Gloss perception in painterly and cartoon rendering. *ACM Transactions on Graphics 32*, 2 (2013), 18:1–18:13.

[BVB*13] Borkin M. A., Vo A. A., Bylinskii Z., Isola P., Sunkavalli S., Oliva A., Pfister H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2306–2315.

[CDW*12] Cho I., Dou W., Wartell Z., Ribarsky W., Wang X.: Evaluating depth perception of volumetric data in semi-immersive VR. In *AVI '12: Proceedings of Advanced Visual Interfaces* (Capri Island, Italy, 2012), pp. 266–269.

[CSD*09] Cole F., Sanik K., DeCarlo D., Finkelstein A., Funkhouser T., Rusinkiewicz S., Singh M.: How well do line drawings depict shape? *ACM Transactions on Graphics 28*, 3 (2009), 28:1–28:9.

[CWM*09] Chan M.-Y., Wu Y., Mak W.-H., Chen W., Qu H.: Perception-based transparency optimization for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1283–1290.

[CWQ07] Chan M.-Y., Wu Y., Qu H.: Quality enhancement of direct volume rendered images. In *Proceedings of the Sixth Eurographics/IEEE VGTC Conference on Volume Graphics* (Prague, Czech Republic, 2007).

[dMPDSF11] de Moura Pinto F., Dal Sasso Freitas C. M.: Illustrating volume data sets and layered models with importance-aware composition. *Visual Computer 27*, 10 (2011), 875–886.

[DRN*17] Díaz J., Ropinski T., Navazo I., Gobbetti E., Vázquez P.-P.: An experimental study on the effects of shading in 3D perception of volumetric models. *Visual Computer 33* (2017), 47–61.

[DV10] Díaz J., Vázquez P.: Depth-enhanced maximum intensity projection. In *Proceedings of IEEE/EG Volume Graphics* (Norrköping, Sweden, 2010), pp. 93–100.

[EKR16] Englund R., Kottravel S., Ropinski T.: A crowdsourcing system for integrated and reproducible evaluation in scientific visualization. In *Proceedings of IEEE Pacific Visualization Symposium* (Taipei, Taiwan, 2016), IEEE, pp. 40–47.

[ER16] Englund R., Ropinski T.: Evaluating the perception of semi-transparent structures in direct volume rendering techniques. In *Proceedings of SIGGRAPH ASIA 2016 Symposium on Visualization* (Macau, 2016), ACM, Article No. 9.

[FB05] Fleming R. W., Bülthoff H. H.: Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception 2*, 3 (2005), 346–382.

[FSM06] Fulvio J. M., Singh M., Maloney L. T.: Combining achromatic and chromatic cues to transparency. *Journal of Vision 6*, 8 (2006), 760–776.

[GCNF13] Gleicher M., Correll M., Nothelfer C., Franconeri S.: Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2316–2325.

[GSBH13] Grosset A., Schott M., Bonneau G.-P., Hansen C. D.: Evaluation of depth of field for depth perception in DVR. In *Proceedings of IEEE Pacific Visualization Symposium* (Sydney, NSW, Australia, 2013), pp. 81–88.

[HB10] Heer J., Bostock M.: Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *ACM Human Factors in Computing Systems (CHI)* (Atlanta, GA, USA, 2010), pp. 203–212.

[Hib00] Hibbard B.: Confessions of a visualization skeptic. *ACM SIGGRAPH Computer Graphics 34* (2000), 11–13.

[IFP97] Interrante V., Fuchs H., Pizer S. M.: Conveying the 3D shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization and Computer Graphics 3*, 2 (1997), 98–117.

[JSYR14] Jönsson D., Sundén E., Ynnerman A., Ropinski T.: A survey of volumetric illumination techniques for interactive volume rendering. *Computer Graphics Forum 33*, 1 (2014), 27–51.

[KOCC14] Kersten-Oertel M., Chen S. J.-S., Collins D.: An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *IEEE Transactions on Visualization and Computer Graphics 20*, 3 (2014), 391–403.

[KSTE06] Kersten M., Stewart J., Troje N., Ellis R.: Enhancing depth perception in translucent volumes. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 1117–1124.

[KVDK92] Koenderink J. J., Van Doorn A. J., Kappers A. M.: Surface perception in pictures. *Perception & Psychophysics 52*, 5 (1992), 487–496.

[KYY12] Kim S.-H., Yun H., Yi J. S.: How to filter out random clickers in a crowdsourcing-based study? In *Proceedings of the 2012 BELIV Workshop* (Seattle, Washington, USA, 2012), ACM, Article No. 15.

[LB99] Langer M. S., Bülthoff H. H.: Depth discrimination from shading under diffuse lighting. *Perception 29*, 6 (1999), 649–660.

[LCD06] Luft T., Colditz C., Deussen O.: Image enhancement by unsharp masking the depth buffer. *ACM Transactions on Graphics 25*, 3 (2006), 1206–1213.

[LKJ*05] Laidlaw D. H., Kirby R. M., Jackson C. D., Davidson J. S., Miller T. S., Da Silva M., Warren W. H., Tarr M. J.: Comparing 2D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics 11*, 1 (2005), 59–70.

[LR11] Lindemann F., Ropinski T.: About the influence of illumination models on image comprehension in direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1922–1931.

[ME04] Mora B., Ebert D. S.: Instant volumetric understanding with order-independent volume rendering. *Computer Graphics Forum 23*, 3 (2004), 489–497.

[MS13] Marreiros F. M. M., Smedby O.: Stereoscopic static depth perception of enclosed 3D objects. In *ACM Symposium on Applied Perception* (Dublin, Ireland, 2013), pp. 15–22.

[NK04] Nagy Z., Klein R.: High-quality silhouette illustration for texture-based volume rendering. *Journal of WSCG 12*, 2 (2004), 301–308.

[QB11] Quinn A. J., Bederson B. B.: Human computation: A survey and taxonomy of a growing field. In *ACM Human Factors in Computing Systems (CHI)* (Vancouver, BC, Canada, 2011), pp. 1403–1412.

[RE01] Rheingans P., Ebert D.: Volume illustration: Nonphotorealistic rendering of volume models. *IEEE Transactions on Visualization and Computer Graphics 7*, 3 (2001), 253–264.

[RSH06] Ropinski T., Steinicke F., Hinrichs K.: Visually Supporting Depth Perception in Angiography Imaging. In *Proceedings of Smart Graphics* (Vancouver, Canada, 2006), pp. 93–104.

[SA02] Singh M., Anderson B. L.: Toward a perceptual theory of transparency. *Psychological Review 109*, 3 (2002), 492–519.

[SEA09] Svakhine N., Ebert D., Andrews W.: Illustration-inspired depth enhanced volumetric medical visualization. *IEEE Transactions on Visualization and Computer Graphics 15*, 1 (2009), 77–86.

[SPGM*11] Schott M., Pascal Grosset A., Martin T., Pegoraro V., Smith S. T., Hansen C. D.: Depth of field effects for interactive direct volume rendering. *Computer Graphics Forum 30*, 3 (2011), 941–950.

[SSK*15] SUNDÉN E., STENETEG P., KOTTRAVEL S., JÖNSSON D., ENGLUND R., FALK M., ROPINSKI T.: Inviwo—An extensible, multi-purpose visualization framework. *Poster at IEEE Visualization* (2015).

[TFCRS11] THOMPSON W., FLEMING R., CREEM-REGEHR S., STEFANUCCI J. K.: *Visual Perception from a Computer Graphics Perspective*. A. K. Peters, Natick, MA, 2011.

[VFSG06] VIOLA I., FEIXAS M., SBERT M., GRÖLLER M. E.: Importance-driven focus of attention. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 933–940.

[VKG05] VIOLA I., KANITSAR A., GRÖLLER M. E.: Importance-driven feature enhancement in volume visualization. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (2005), 408–418.

[vPV11] ŠOLTÉSZOVÁ V., PATEL D., VIOLA I.: Chromatic shadows for improved perception. In *Proceedings of Non-Photorealistic Animation and Rendering* (Vancouver, Canada, 2011), pp. 105–115.

[vTPV12] ŠOLTÉSZOVÁ V., TURKAY C., PRICE M., VIOLA I.: A perceptual-statistics shading model. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2265–2274.

[War12] WARE C.: *Information Visualization: Perception for Design*. Morgan Kaufmann, San Francisco, Calif; Elsevier Science, Oxford, 2012.

[WPF12] WANG N., PALJIC A., FUCHS P.: An evaluation of rendering and interactive methods for volumetric data exploration in virtual reality environments. In *Proceedings of SPIE* (2012), vol. 8294, pp. 82940W–82940W–13.

[WQC*10] WU Y., QU H., CHUNG K.-K., CHAN M.-Y., ZHOU H.: Quantitative effectiveness measures for direct volume rendered images. In *Proceedings of IEEE Pacific Visualization Symposium* (Taipei, Taiwan, 2010), pp. 1–8.

[ZWM13] ZHENG L., WU Y., MA K.-L.: Perceptually-based depth-ordering enhancement for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (2013), 446–459.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure 1. We have investigated the perceptual impact when applying six widely used volume rendering techniques to semi-transparent structures.

Figure 2. An illustration of the Necker Cube illusion.

Figure 3. The tested volumetric line drawings are obtained by fusing a DVR image (a) with extracted silhouette lines (b), to which an additional halo effect is applied to emphasize X and T junctions (c).

Figure 4. To emphasize X and T junctions, we have added halos within the Volumetric Line Drawing technique. While lines without halos (a) do not support depth sorting, this ambiguity is removed when adding halos (b).

Figure 5. The welcome screen and end survey as shown to the participants when they are about to start the evaluation and when they finished it.

Figure 6. Screen grab of the webpage for the three tasks the participant had to complete for the evaluation. The linked examples showed generic example images together with correct and incorrect answers.