

Reevaluating Reconstruction Filters for Path-Searching Tasks in 3D

D. A. T. Roberts and I. Ivriissimtzis

School of Engineering and Computing Sciences, Durham University, UK

Abstract

In this paper we present an experiment on stereoscopic Direct Volume Rendering (DVR), aiming at understanding the relationship between the choice of reconstruction filter and participant performance on tasks requiring spatial understanding such as 3D path-searching. The focus of our study is on the impact on task performance of the post-aliasing and smoothing produced by the reconstruction filters.

We evaluated five reconstruction filters, each under two different transfer functions and two different displays with a wide range of behaviours in terms of post-aliasing and smoothing. We found that path-searching tasks commonly found in the literature, and as the one we employed here, elicit bias in the responses which should be taken into account when analysing the results. Our analysis, which employed both standard statistical tests and techniques from signal detection theory, indicates that the choice of reconstruction filter affects some aspects of the spatial understanding of the scene.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms, Viewing Algorithms I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Stereo, Depth Cues

1. Introduction

Compared to standard 2D displays, stereoscopic displays can aid the understanding of complex 3D scenes and thus, improve performance in tasks such as path-searching. While the effectiveness of stereoscopic displays in aiding path-searching tasks on 3D surface renderings is well-established [WF96, WM05, HHL10], it was only recently that researchers have started assessing their effectiveness with Direct Volume Rendering (DVR). As a result, the relevant literature is sparse and the results inconclusive.

Assessing the effectiveness of stereoscopic displays with images produced via DVR is a challenging task. Typically, the rendered volumes have several characteristics such as high transparencies, or overlapping spatial features, whose effect on task performance is not immediately obvious. The situation is further complicated by the multitude of parameters that have to be chosen prior to producing DVR images, including the choice of reconstruction filter and transfer function.

Here, we focus on aliasing and blurring whose impact on stereoscopic vision has already been established in non-

DVR contexts. For example, using stereoscopically displayed 3D surfaces, [Pfa00] showed aliasing to produce inconsistent depth perceptions, while [CMHV10], using the Frisby Stereo Test, showed that blurring degrades fine depth perception. Thus, our initial expectation, partially only verified by the experiment, was that aliasing and blurring, which are controlled by the DVR parameter settings and referred to in this context as post-aliasing and smoothing, have a negative impact on DVRs too.

Our experiment is based on a path-searching task similar to those found in the literature for evaluating 3D surface rendered stereoscopic scenes. We compare five DVR reconstruction filters and two transfer functions in terms of accuracy and response times when displayed stereoscopically and in 2D. The scene consists of a graph with spherical nodes and cylindrical edges, with a certain amount of added Perlin noise. The task is to specify whether two nodes, highlighted with the colour red, are connected with a path of length two or not. We seek to answer the following questions: 1) How does the opacity, controlled by the choice of transfer function, relate to task performance? 2) How does post-aliasing and smoothing, controlled by the choice of re-

construction filter, relate to task performance? 3) How do the results between stereoscopic and monoscopic displays compare?

The analysis of the results showed that our experiment was eliciting biased responses from the participants. More specifically, a No answer was more likely than a Yes answer with 63.14% of the responses in 3D and 56.58% of the responses in 2D being No answers. Such bias, which is common in Yes-No experiments, is not necessarily a negative aspect. For example, a surgeon wanting to reduce unnecessary operations may have a preference towards saying no, whereas when detecting a tumor may say yes more often. However, as a result of the bias in our experiment, we found that the standard statistical analysis of the overall accuracy rates is inadequate and techniques from *signal detection theory*, give insights that other techniques miss.

Our analysis shows that in 3D displays the choice of transfer function has the greatest effect on task performance, with the accuracy increasing with opacity. The overall effect of the choice of reconstruction filter on accuracy rates was not strong, however, the choice of filters with large amounts of post-aliasing had a strong effect in maximising the number of correct rejections. Using signal detection theory techniques, we also found that a choice of in-between filters, which avoid large amounts of both post-aliasing and smoothing, had a strong effect in maximising the ability of the participants to discriminate between connected and not-connected nodes. When the whole experiment was duplicated in 2D we found that task performance was significantly lower compared to 3D.

A first implication of the results is that if the maximisation of correct rejections is the primary concern, e.g. during a process of selection by elimination, then filters with higher post-aliasing should be preferred. On the other hand, if the ability to discriminate between connected and not-connected nodes is the primary concern, then filters with a better balance between post-aliasing and smoothing should be preferred. We note that this second observation stems from the signal detection theory analysis and possibly due to the bias in the data, does not translate into a strong effect of the reconstruction filter on the overall accuracy rates of the experiment.

A second implication of the results of the paper is that, unless experimental designs eliciting unbiased responses are employed, such as *m*-Alternative Forced Choice (mAFC) experiments, signal detection theory should be applied to measure the bias and then explicitly take it into account in the analysis. Finally, if for any reason signal detection theory cannot be the analytical tool of choice, the experimental results should nevertheless be presented in a way that bias can be computed. In the case of a binary decision experiments in particular, hits and correct rejections should be separately reported, rather than just accuracy rates.

In summary, the main contributions of the paper are:

- A comprehensive experiment on task performance with stereoscopic DVR, taking into account opacity on the one hand and post-aliasing and smoothing on the other.
- A demonstration that bias should be taken into account in the analysis of results of Yes-No path-searching experiments.
- Evidence supporting the use of stereoscopic display of DVR images over 2D for tasks requiring a high level of spatial understanding.

The rest of the paper is organised as follows. In Section 2, we discuss the relevant literature. In Section 3, we present the experimental setup. In Section 4, we analyse the results which are then discussed in Section 5 and we briefly conclude in Section 6.

2. Background

2.1. Background to DVR

DVR uses volumetric data represented as the nodes of a 3D grid lattice. The nodes store scalar values which can represent measurements of density, heat or some other quantity computed via simulation. Due to the discrete nature of the data, whilst rendering the volume there is often the need to compute values between grid nodes. This can be described by the equation [TBU00]:

$$f(x) = \sum_{k \in \mathbb{Z}^3} f_k \phi(x - k) \quad x \in \mathbb{R}^3 \quad (1)$$

That is, the value at position x is a linear combination of the values f_k of some nodes with the weights given by the reconstruction filter ϕ .

Provided that the sampled volume data is band-limited, the *sinc* function can be used for perfect reconstruction. There are however considerations that usually make this infeasible in practice. Firstly, the band-limited requirement might not be satisfied due to high frequencies in the original stimulus, or limited resolutions of the scanning devices. Secondly, the *sinc* function has unbounded support in the spatial domain. It is therefore customary to use finite support reconstruction filters approximating the ideal *sinc* filter.

Due to the differences between the ideal filter and bounded support filters, artefacts are introduced into the reconstructed data. These reconstruction artefacts can be categorised as post-aliasing, smoothing and ringing [MN88, ML94]. In an image, post-aliasing visually manifests as jaggies or staircase artefacts, smoothing as the blurring of fine details, while ringing is displayed as oscillations between high and low intensities centred around sharp edges in the data. For illustration, Figure 1 shows a DVR image of a blood vessel data set, Figure 4(k) shows a zoomed in section with high amount of smoothing typical of the B-spline filter, while Figure 4(o) shows a zoomed in section with spurious edges typical of the high post-aliasing of the Welch windowed sinc filter.



Figure 1: DVR image of a blood vessel data set.

2.2. Evaluations of Stereoscopic Volume Rendering

A number of studies have concluded that stereoscopy can assist in certain tasks with DVR that require spatial understanding, however the results are not unanimous or consistent.

When comparing the depth of two vessels in angiography data sets like the one in Figure 1, Ropinski reported that participants who were inexperienced with stereoscopic displays produced increased error rates and response times when compared to monoscopic depth cues [RSH06]. Similarly, for a similar task performed by novices and experts Kersten reports that enhanced monoscopic depth cues were capable of producing better results than stereo [KOCC14]. In [CWD*14] in contrast, for depth discrimination tasks using simulated angiography-like data sets stereoscopic displays proved more effective in most cases against 2D displays. Recently, [ABKP15] investigated methods to enhance MR angiography images by combining contour enhancement with stereopsis for DVR. Using a relative depth task it was found that whilst stereopsis improved accuracy, it was most effective when combined with contour enhancement.

Purely absorptive volume rendering has also been evaluated when combined with the kinetic depth effect in stereoscopic and monoscopic conditions [KSTE06]. Stereoscopic displays were found to improve the ability of users to determine which direction a volume rendered cylinder was rotating. Similar results for the same task were found in a later study with multi-view auto-stereoscopic displays [ABG*09].

Further conflicting results have been reported when participants have been required to determine the relative depth

of volume rendered translucent cylinders. An initial study reported that stereo display of the cylinders is beneficial for this type of task [CDW*12]. However, in a similar study with cylinders combined in a computed tomography data set, stereoscopy was found to have no significant effect for correct depth discrimination [EJH*13].

Even within the same study the results of tasks for stereoscopic volume rendering are not consistent. In an early investigation of volume rendering with stereoscopy, Hubbard found that when users were required to determine which sphere from a set of three was nearest, stereo only improved the results when the spheres were contained within a semi-transparent shell [HHM97]. A recent study [LSSB12] evaluating a range of quantitative tasks for volume rendering with different immersive display setups found that complex search tasks benefited from higher levels of immersion. Yet when slicing the data sets a decrease in performance was found when stereoscopic displays were used. The follow-up study found that different display modalities benefit different volume rendering styles [LBS14].

When producing stereoscopic images there is a large number of parameters affecting the image properties. These include camera separation, camera model, viewing distance, as well as hardware properties of individual stereoscopic displays. In DVR, there is a further set of considerations including transfer function, sampling rate, reconstruction filter and data set resolution. The variation of the results found in the literature may reflect the large number of parameters used in the studies. A study on reconstruction filters in DVR with stereoscopic displays has found that the choice of filter alone can have an impact on a spatial search task [RIH14], but it is not clear which properties of the filters correlate with task performance, or what impact more realistic data sets would have on the task.

As an alternative to stereoscopic DVR, other studies attempt to improve depth perception with DVR via the use of monocular depth cues. These include using the depth of field by blurring specific features in the scene [RMSD*08, GS13] and the use of local and global illumination models [RMSD*08, RDRS10, LR11]. Other additions include the use of feature halos surrounding regions of interest in the volume [SE03, BG07] and using the kinetic depth effect to assist in the understanding of structure [KSTE06, BBD*09]. However, unlike using stereoscopic displays, each of these methods change the visual appearance of the volume and although they can increase the perception of depth, details in the scene can be obscured.

2.3. Tasks used to Evaluate DVR

There are several considerations related to the choice of the task involved in the experimental testing of volume rendering. First, the type of volumetric data to use, either real or simulated can have a large impact on the experiment. A large

source of volumetric data is from the medical domain and as such the majority of evaluations of depth perception with DVR use either real medical data sets, or simulated data sets that have properties similar to data produced by medical scanners. Real data sets often have characteristics that can be difficult to reproduce in simulated data sets, therefore certain studies have primarily used these. In prior evaluations that have used real angiography data sets, a common task is to differentiate the depth of two blood vessels [RSH06, KOCC14] or for other types of data to determine which highlighted feature of the volume is nearest [GS13, LR11]. Other tasks using real data sets involve asking users to determine the direction a volume is facing [DV10], or using a Likert scale to subjectively evaluate DVR images [MSRH08].

Alternatively, simulated data sets can allow for a greater amount of control on the difficulty level of the task, as well as being able to control the artefacts in the images used. A relatively common task found in the literature is to use a large Perlin noise filled cylinder and then having participants determine which direction the cylinder is rotating [BBD*09, KSTE06, ABG*09]. Whilst the use of noise simulates some properties of real data sets, the use of a single large cylinder is abstract. By using multiple overlapping cylinders a more realistic data set can be produced reminiscent of blood vessels, in which case participants are required to determine the correct depth ordering of the cylinders [CWD*14, BBD*09].

For this study we use a path-searching task with simulated volume rendered data sets. Such tasks are becoming a standard for evaluating high fidelity displays [WF96, WM05, CCAL12, HHL10], as well as seeing use within the medical field [ABG*09, vSvDZS*10] and studies of underground cave systems [RKS13]. Despite this, spatial understanding tasks are not yet in common use for DVR evaluations.

2.4. Signal Detection Theory

The path-searching experiment reported in this study can be classified as a Yes-No experiment. Historically the analysis of such experiments have been limited to the statistical analysis of raw accuracy results and in some cases response times of each participant. By applying techniques from Signal Detection Theory [MD04], a wider range of analysis can be performed, providing a different insight into the results gathered. In particular, different types of errors can be evaluated, as well as the sensitivity and bias of the participants, providing a more finer analysis of the results than statistical analysis of raw accuracy alone.

Signal Detection Theory originated as a method of analysing experiments where auditory or visual stimuli have to be distinguished from some background noise. Since this initial use, it has been applied to the analysis of memory tasks, cognition as well as tasks within the medical domain [ABKP15].

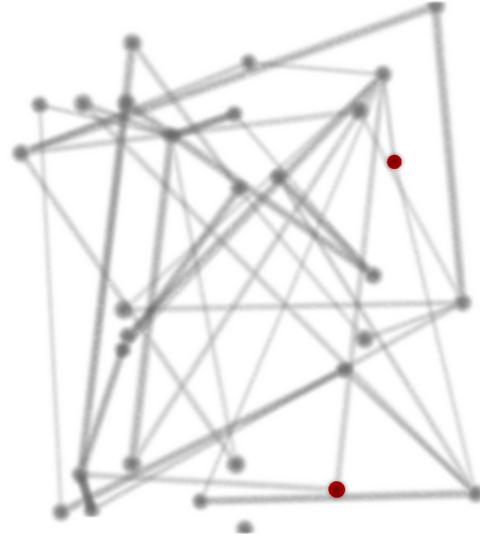


Figure 2: Example of a test graph image with 30 nodes. The two highlighted nodes are connected by a path of two.

3. Experimental Methodology

3.1. Path-Searching Task Description

In each image presented to the participants a direct volume rendered graph is displayed. The task of the participant is to determine whether two highlighted nodes in the graph are connected by a path of two. Each trial is set so that the chance of the nodes being connected by a path of two is 50%. The 'y' key on the keyboard is used for the participant to indicate there is a path of two where as the 'n' key is used to indicate they cannot see a path.

The structural layout of the graph follows that of [WF96, HHL10]. The number of nodes in all trials was set to 90 and they were evenly divided into three sets, two leaf sets and a single set of intermediate nodes. When viewed stereoscopically, the intermediate nodes were displayed on the zero-disparity plane without depth, the first leaf group was displayed at 50mm in front of the display and the second leaf group at 50mm behind the display. In the 2D experiment, all nodes were displayed without depth on the screen plane. The x and y coordinates were randomly distributed.

The chosen path-searching task is similar to the search for aneurisms in real-life angiography data sets. Both of the tasks require a high-degree of spatial understanding and require the observers to understand the layouts of the scene [GS13, RSH06]. For illustration, an example test image with only 30 nodes is shown in Figure 2.

3.2. Stimulus Generation

The graph data sets were generated using the voxelisation package *vxtl* [NDS10] with resolution $256 \times 256 \times 256$.

In order to produce images that are representative of blood vessel data sets, two features have been used. The first is that the lines connecting the nodes in the graph have been randomly assigned a diameter from the choice of [0.0015, 0.003, 0.006], where the size of the scene in normalized coordinates is [0, 1]. These diameters represent fine, medium and thick blood vessels.

Secondly, volumes generated via medical scanners and other devices often exhibit some form of noise, for example in volumes created using ultrasound scanners this is a primary cause of concern. In order to simulate this property the generated graph volumes are modulated by a texture with resolution ($256 \times 256 \times 256$). Each point in the texture is given a value according to the Perlin noise function, with the parameters controlling the frequency and harmonics of the noise chosen as $a = b = 2$. By modulating the generated graphs with the noise the produced volumes have an irregularity that approximates those of natural data sets.

3.3. Reconstruction Filters to be Evaluated

The five reconstruction filters to be evaluated were chosen for their wide range in the amount of smoothing and post-aliasing they introduce, as well as being in common usage in DVR. In order to quantitatively assess the amount of smoothing and post-aliasing each filter exhibits, the frequency domain metrics of Marschner and Lobb were used [ML94].

There, smoothing is assessed by measuring the energy the filter extrudes in the pass-band:

$$S(h) = 1 - \frac{1}{|R_N|} \int_{R_N} |\hat{h}|^2 dV \quad (2)$$

where h is the filter to be measured, \hat{h} is the frequency domain representation of h , R_N is the Nyquist region and $|R_N|$ is the volume of R_N . Post-aliasing is assessed by measuring the energy the filter extrudes in the stop-band:

$$P(h) = \frac{1}{|\bar{R}_N|} \int_{\bar{R}_N} |\hat{h}|^2 dV \quad (3)$$

where \bar{R}_N is the complement of R_N .

The *trilinear* interpolation scheme is the simplest and most common filter capable of recreating a continuous function. It produces images of acceptable quality and as it is implemented in hardware in GPUs it has become the defacto standard reconstruction filter for DVR. The *Catmull-Rom spline* satisfies the interpolation constraint whilst balancing smoothing and aliasing. The *cubic B-spline* is an approximating filter exhibiting minimal post-aliasing at the cost of significant smoothing. Due to this it is used primarily in DVR when there is a large amount of noise present. Derived from the approximating B-spline filter via the use of generalized interpolation [BTU99], the *interpolating B-spline* has a higher degree of smoothing and greater accuracy than the trilinear and Catmull-Rom filters whilst satisfying the interpolation constraint.

The choice of the last filter is motivated by the desire to test a reconstruction filter that exhibits high post-aliasing with minimal smoothing. A number of choices are available that include the pass-band optimal filters of [Car93] as well as the premultiplied cubic filters of [Csé08]. Such filters are however not in common usage in DVR. Instead, much research has gone into the study of windowed-sinc filters whereby a finite window with some radius is applied to the ideal *sinc* reconstruction filter to produce a more practical filter. By choosing an appropriate window function and radius a considerable degree of control over the frequency domain artefacts can be exercised. The Welch window function is defined as:

$$Welch_{\tau}(x) = \begin{cases} 1 - \left(\frac{|x|}{\tau}\right)^2 & |x| \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where τ is the radius of the window and for the purposes of this experiment is set to 4. This produces a reconstruction filter that has low smoothing, less than that of the interpolating B-spline, whilst exhibiting significant post-aliasing. The smoothing and post-aliasing properties of each reconstruction filter according to the Marschner and Lobb metrics are shown in Figure 3. Zoomed in images of the graph stimulus used for the experiment generated with the different reconstruction filters are shown in Figure 4.

For each reconstruction filter in the experiment, two transfer functions have been used to produce two test images that are typical of DVR. The first produces semi-transparent renderings with a white colour value of 0.35 and an increasing linear opacity of 0 – 0.6. The second transfer function has the same colour settings and a step function with 0.95 at non-zero values for opacity, producing more opaque images. The choice of transfer functions is motivated by two common styles of volume rendering. Semi-transparent rendering is typically used in DVR to allow multiple layers to be visible, while in iso-surface rendering the surfaces of interest are made opaque. To be able to compare the effects of the reconstruction filter and transfer function on different display methods the experiment has been performed twice—once in 3D and once when presented in 2D.

3.4. Equipment and Viewing Conditions

A True3DI 24" HD-SDI monitor with a resolution of 1920×1200 pixels for each eye and a refresh rate of $60Hz$ was used to display the images for the experiment. Participants were required to wear linear polarized glasses during both the 3D and 2D experiments and were placed at a distance of $60cm$ from the display. Light levels were kept consistently low throughout the experiments. Figure 5 shows a participant taking part in the experiment.

3.5. Participant Selection

In total 60 participants were recruited to take part with 30 performing the experiment in 3D and 30 in 2D with a small

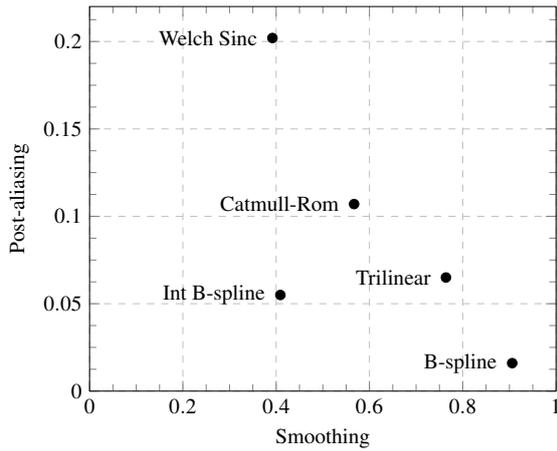


Figure 3: Marschner and Lobb metrics for the reconstruction filters of the experiment. Higher values indicate an increased presence of smoothing/post-aliasing.

overlap between the two sets. There were 31 women and 29 men with ages ranging from 18 – 39 and a mean age of 20.75. Each participant was paid £7.50 to take part in the experiment. All participants were screened for vision using the Bailey and Snellen chart and for stereo-vision using the Titmus fly stereo test. All participants were informed that accuracy and response time would be recorded during the task. Whilst some of the participants had experience with stereoscopic displays, they were unfamiliar with the DVR image generation process and had minimal knowledge of reconstruction filters.

The use of novices follows that of similar evaluations of volume data sets [LSSB12,LBS14]. Apart from the difficulty in obtaining a significant number of domain experts in volume data, their prior knowledge can obfuscate the results. On the other hand, search tasks have been identified as being relevant to domain experts whilst not requiring specialised knowledge [LBS14].

3.6. Procedure

Prior to taking part in the full experiment, participants were required to perform a trial test using graphs with a reduced difficulty level containing only 30 nodes for training purposes. The trial test contained six images with three images containing a path of two and three not. In both the 3D and 2D experiments there were five trials for each reconstruction filter per the two transfer functions leading to a total of ten trials. Each trial contained twelve images, after which there was a short break to reduce the risk of fatigue. A Latin Square design was used to determine the exact ordering of the trials.

Table 1: Summary of results. TF and Rec denote transfer function and reconstruction filter, respectively. Asterisks denote significant p -values, $p < 0.05$.

Condition	3D		2D	
	Acc	RT	Acc	RT
Overall				
TF	0.010*	0.369	0.038*	0.064
Rec	0.576	0.451	0.893	0.915
TF x Rec	0.576	0.451	0.017*	0.758
Hits (True-Positive)				
TF	0.183	0.623	0.681	0.160
Rec	0.906	0.180	0.928	0.543
TF x Rec	0.451	0.215	0.052	0.476
Misses (False-Negative)				
TF	-	0.450	-	0.975
Rec	-	0.046*	-	0.018*
TF x Rec	-	0.841	-	0.260
Correct Rejections (True-Negative)				
TF	0.002*	0.646	0.020*	0.132
Rec	0.621	0.485	0.234	0.894
TF x Rec	0.022*	0.337	0.079	0.904
False-Alarms (False-positive)				
TF	-	0.112	-	0.352
Rec	-	0.457	-	0.100
TF x Rec	-	0.766	-	0.229

3.7. Hypotheses

Based on the findings of the prior work, as discussed in Section 2.2, our initial hypotheses were:

- Reconstruction filters that exhibit extremes of either blurring or aliasing will display lower overall accuracy rates.
- The accuracy rates will be higher in 3D than in 2D.
- The transfer function with high opacity will display higher task performance

4. Results

Section 4.1 presents the results of a standard set of statistical tests run over the overall accuracy rates and response times recorded in the experiments. As the bias in the responses is not taken into account, the findings are relatively weak. In Section 4.2 we categorise the results into hits, misses, correct rejections and false alarms and run the same statistical tests on each category. Complementing this approach, in Section 4.3 we analyse the results with techniques from Signal Detection Theory, determining participant sensitivity and quantifying the participant bias.

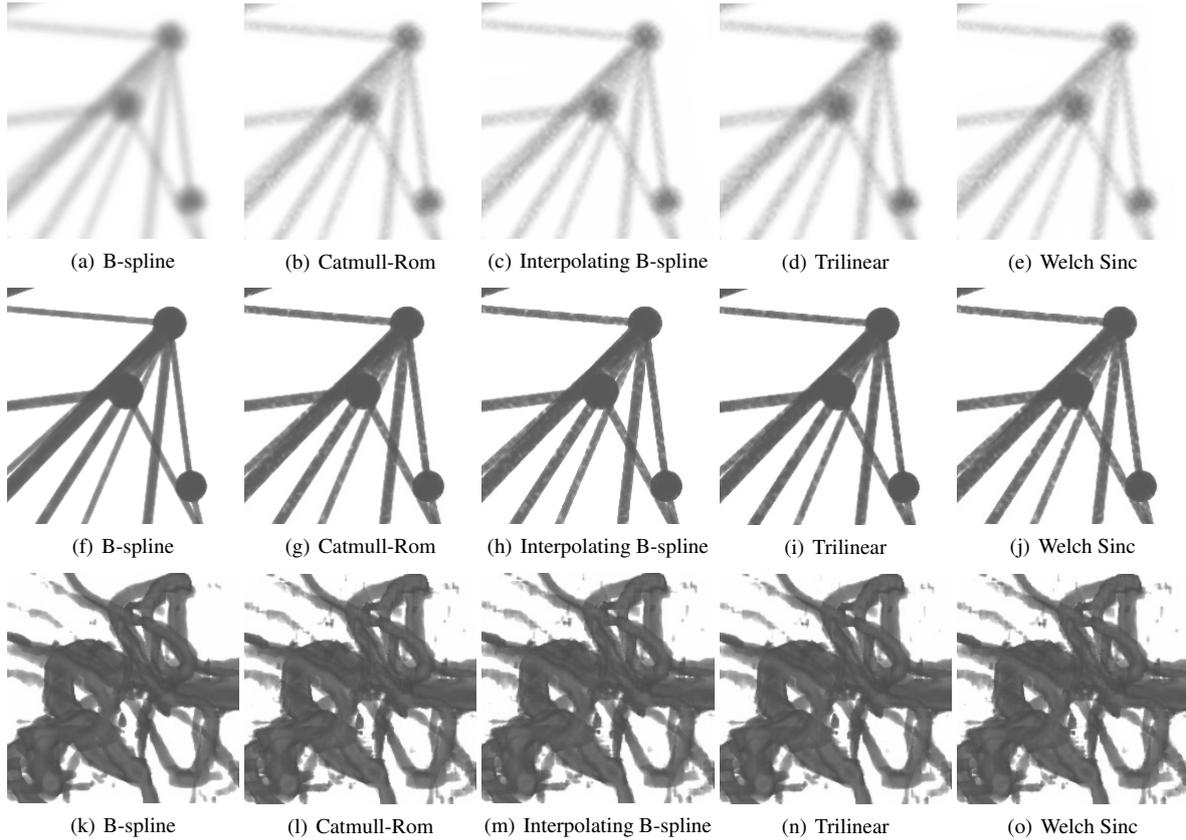


Figure 4: Zoomed in view of the graph stimulus and blood vessel data set with different reconstruction schemes. Figures a-e have been generated with the semi-transparent function, Figures f-j have been generated with the opaque transfer function.



Figure 5: Picture of a participant performing the experiment

4.1. Overall Accuracy and Response Time

First, the accuracy rates and response times of the 2D and 3D experiments are analysed separately, applying a two-way repeated measures of analysis of variance (ANOVAs) for the within subjects variables of transfer function and reconstruction filter. Next, the combined results of the experiments, are analysed with a one-way ANOVA and any significant p -values are discussed at the end of the relevant subsections. Mauchly's Test of Sphericity was used to test for sphericity with the Greenhouse-Geisser correction being applied when required.

In the two-way ANOVA, the reconstruction filter and transfer function, classified as either semi-transparent or opaque, were the independent variables and the accuracy and response times were dependent variables. The p -values are summarised in Table 1. For each condition, means and standard deviations for the correct responses are shown in Figure 6(a) and for the response times in Figure 6(b).

In both 2D and 3D, the transfer function had an effect on accuracy. In 3D, statistical significance was found with

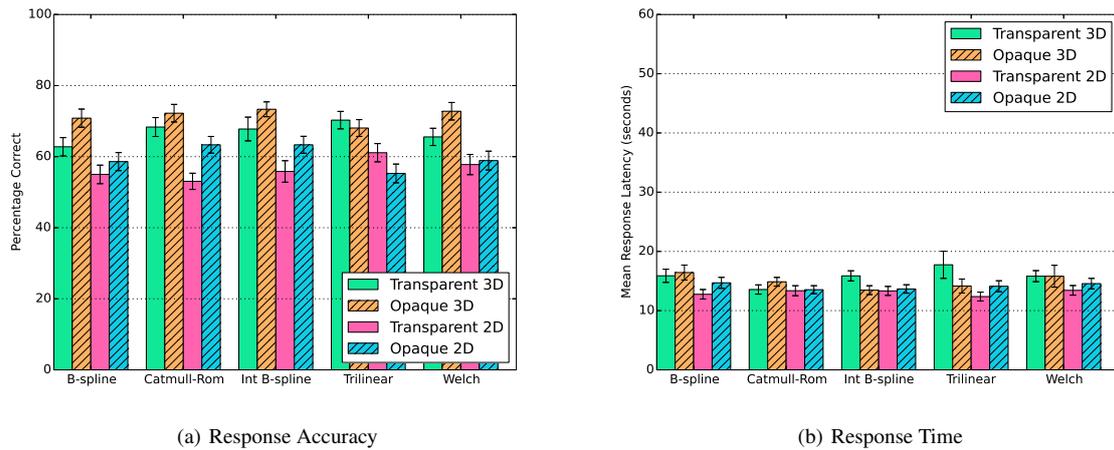


Figure 6: Overall task performance with standard errors.

$F(1,29) = 7.657$, $p = 0.010$, partial $\eta^2 = 0.209$. The accuracy increased from semi-transparent (8.033 ± 0.191) to opaque rendering (8.573 ± 0.177). In 2D statistical significance was found with $F(1,29) = 4.741$, $p = 0.038$ partial $\eta^2 = 0.141$. Again, accuracy increased from semi-transparent (6.787 ± 0.155) to opaque (7.187 ± 0.137). There were no significant effects on response time.

The one-way ANOVA of the display variable showed that participants performed better in 3D than in 2D with $F(1,29) = 43.288$, $p < 0.001$, partial $\eta^2 = 0.599$ increasing from 2D (6.987 ± 0.114) to 3D (8.303 ± 0.156).

4.2. Hits and Correct Rejections

The results of the experiment can be classified into four categories. Using signal detection theory terminology, the four categories are ‘Hits’, ‘Misses’, ‘Correct Rejections’ and ‘False-Alarms’. These correspond to True-Positives, False-Negatives, True-Negatives and False-Positives respectively. The means and standard deviations of the accuracy rates can be seen in Figure 7. Only the Hits and Correct Rejections are given as the Miss and False-Alarm rates can be calculated as (50%-Hits) and (50%-Correct Rejections).

4.2.1. Categorised accuracy

No statistical significance was found for either transfer function or reconstruction filter on the number of hits in the 2D or 3D experiments.

Regarding correct rejections, there was no effect in the 2D experiment, however, in 3D the transfer function was found to have a significant effect. The results were $F(1,29) = 11.581$, $p = 0.002$, partial $\eta^2 = 0.285$. The increase was 0.333, going from semi-transparent rendering

with the mean accuracy of (4.733 ± 0.150) to opaque with accuracy (5.107 ± 0.130).

A further significant effect was found when assessing the interactions between the transfer function and the reconstruction filter on the correct rejection rate. Statistical significance was found with $F(4,116) = 2.985$, $p = 0.022$, partial $\eta^2 = 0.093$. Following this, simple main effects were run. Accuracy was significantly different between the semi-transparent rendering (4.600 ± 0.218) compared to the opaque rendering (5.367 ± 0.169) for the B-spline approximation filter, $F(1,29) = 15.326$, $p = 0.01$, partial $\eta^2 = 0.346$.

The display had no significant effect on the number of Hits, however, it had a significant effect on the correct rejections with $F(1,29) = 26.464$, $p < 0.001$, partial $\eta^2 = 0.477$ increasing from 2D (3.887 ± 0.131) to 3D (4.940 ± 0.131).

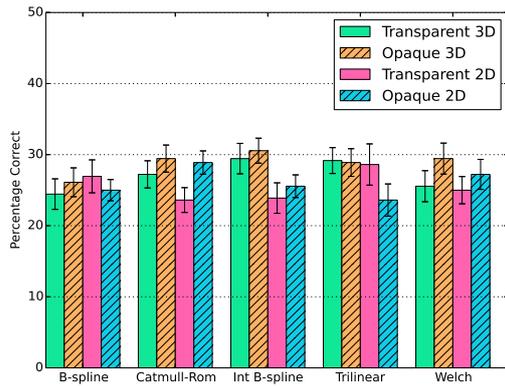
4.2.2. Response Time

Statistical significance was found only on the response times of the misses. In 3D, statistical significance was found with $F(3.126, 215.668) = 2.681$, $p = 0.046$, partial $\eta^2 = 0.037$. Post-hoc analysis revealed that there was no significant differences between individual pairs of reconstruction filters. In 2D, the reconstruction filter was also found to have a significant effect on misses with $F(2.599, 111.760) = 3.720$, $p = 0.018$, partial $\eta^2 = 0.103$. Post-hoc analysis revealed no significant pair-wise interactions.

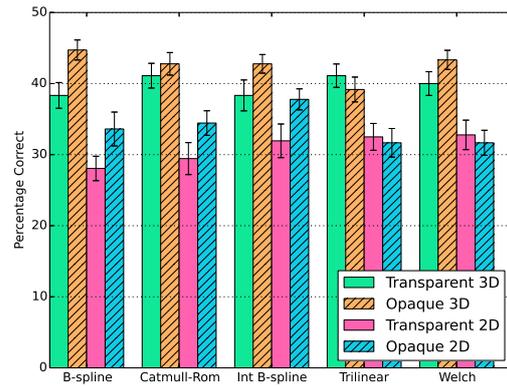
4.3. Signal Detection Theory Analysis

4.3.1. Sensitivity Measure

From the accuracy scores obtained, the sensitivity measure for each participant was calculated. A high sensitivity means that participants had a good ability to detect the



(a) Hits Accuracy (True-Positive)



(b) Correct Rejections Accuracy (True-Negative)

Figure 7: Categorized accuracy rates with standard errors.

path, whereas a low sensitivity value means poor ability. The d' measure used to calculate the sensitivity, is defined as [MD04]:

$$d' = z(H) - z(F) \quad (5)$$

where H is the conditional probability of a hit and F is the conditional probability of a false-alarm. $z(H)$ and $z(F)$ convert the hit and false-alarm probability rates to z-scores that are standard deviation units and have a range of $[0, 4.653]$. Figure 8(a) and Figure 8(b) show boxplots for d' for semi-transparent and opaque renderings, respectively.

In 3D, the transfer function had a significant effect on sensitivity with the results being $F(1, 29) = 4.650$, $p = 0.039$, partial $\eta^2 = 0.138$. Further significance was found when analysing interactions between the independent variables. The results of the ANOVA were $F(4, 116) = 2.487$, $p = 0.047$, partial $\eta^2 = 0.079$ and following this simple main effects were run. Significant results were found for the B-spline filter between semi-transparent rendering (0.879 ± 0.113) compared to the opaque rendering (1.252 ± 0.111) with $F(1, 29) = 9.034$, $p = 0.005$, partial $\eta^2 = 0.238$. Significance was also found for the Welch windowed sinc filter between semi-transparent rendering (0.935 ± 0.121) and opaque (1.272 ± 0.137) with $F(1, 29) = 5.046$, $p = 0.032$, partial $\eta^2 = 0.148$.

The display had a significant effect on sensitivity with $F(1, 29) = 42.216$, $p < 0.001$, partial $\eta^2 = 0.593$ increasing from 2D (0.731 ± 0.026) to 3D (1.150 ± 0.059).

4.3.2. Bias Response Measure

There was an overall preference towards a participant deciding that there is no connection between two nodes with the mean of No answers being 59.862%. The response bias

can be measured using the Criterion Location measurement, defined in [MD04] as:

$$c = -\frac{1}{2}[z(H) + z(F)] \quad (6)$$

where H , F and z are as Equation 5 and the range of c is ± 2.325 . A value of 0 is the neutral point and reflects an ideal participant who prefers neither response. A negative value reflects a *liberal* participant who is more likely to say yes. A positive value reflects a *conservative* participant who is more likely to err towards saying no. Figure 9(a) and Figure 9(b) show the boxplots for c for semi-transparent and opaque renderings respectively.

Only the display had a significant effect on bias, with the bias response measure increasing from 2D (0.191 ± 0.052) to 3D (0.371 ± 0.042) with $F(1, 29) = 7.670$, $p = 0.010$, partial $\eta^2 = 0.209$.

5. Discussion

5.1. Overall Accuracy and Response Time

Firstly, we discuss the overall results and compare to other studies. Due to other studies only reporting the overall accuracy and not hits or correct rejections, we are limited to just comparing findings on overall accuracy values and response times.

In 3D, the average accuracy rate of 69.195% is below that of similar, but not DVR, path-searching experiments with comparable complexities with approximately 85% being reported in [WF96], 90% in [WM05] and more recently 88.245% in [HHL10]. The results are however within the range of depth discrimination tasks using stereoscopic DVR with angiography-like datasets. [CWD*14] reports 67.59%, [RSH06] reports approximately 72% and [KOCC14] reports

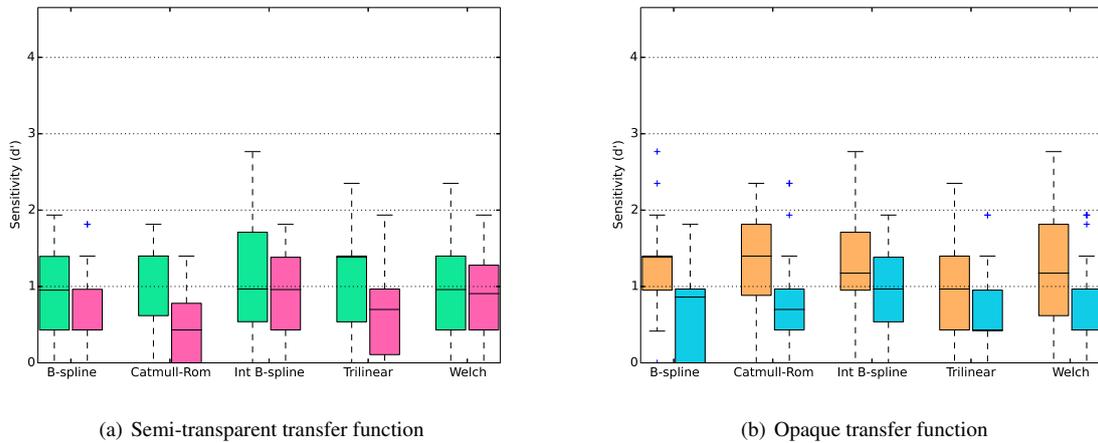


Figure 8: Sensitivity measure for both transfer functions used.

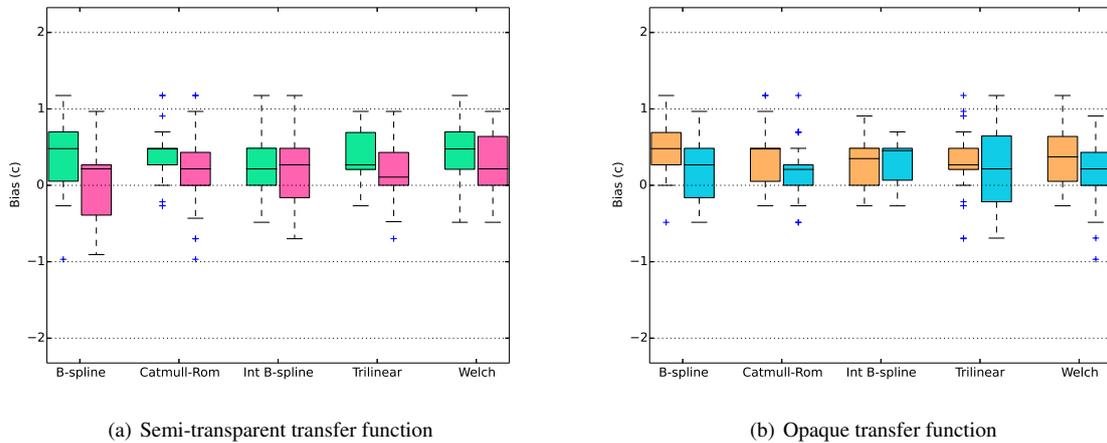


Figure 9: Bias response measures.

64.8%. The differences suggest that surface-based renderings, as used in the previous path-searching tasks, produce less ambiguity and therefore the use of low opacity must be carefully considered when using DVR. This result follows that of [JEH03] where stereoscopic displays combined with transparent surfaces led to ambiguous depth perception when viewing anatomical data. It also provides evidence for the results of [LBS14] where opaque iso-surface rendering was found to benefit from stereoscopic displays, further confirming our hypothesis on opacity.

Contrary to the hypothesis, and previous research [RIH14, MN88], the choice of reconstruction filter was found to not have any impact on raw task accuracy for path-searching tasks. Instead, we find a more complex relationship between smoothing and post-aliasing on the one hand and participant

responses on the other. We suggest that the differences between the results found in this experiment and the earlier experiment can be attributed to the introduction of noise and randomised line diameters used when generating the data sets. This may have impacted the geometry of the volume data, in particular the shape compactness of the rendered images, which has been shown to be correlated with the perceived quality of depth [RI16].

Regarding the display, we found that there was a higher task accuracy when the stimulus was presented stereoscopically. That aligns well with results from previous studies [KSTE06, CWD*14, LSSB12].

5.2. Hits and Correct Rejections

While the transfer function had a significant effect on correct rejections, increasing with opacity for both 2D and 3D, there was no significant effect on hits.

In 3D, an interaction effect was found between the transfer function and the reconstruction filter on correct rejections for the B-spline and interpolating B-spline filters with low accuracy values found when combined with semi-transparent rendering. According to the Marschner and Lobb metrics, the filters have considerably different smoothing properties, yet their degrees of post-aliasing is comparable. It may then be suggested that low post-aliasing combined with semi-transparent rendering can lead to a decrease in the accuracy when assessing if two nodes are not connected.

Regarding the display, we found a significant effect on correct rejections but not on hits, meaning that the overall higher accuracy rates of the 3D display can be attributed to the higher number of correct rejections. In turn, the higher number of correct rejections in 3D can be partially attributed to the higher bias, that is, to the higher number of No answers given by the participants.

5.3. Signal Detection Theory

When analysing the sensitivity of participants to finding a path the opaque transfer function produced the highest sensitivity values. In 3D an interaction effect was also found between the transfer function and the reconstruction filter with the B-spline and Welch windowed sinc filters producing the lowest sensitivity values. From the frequency domain metrics, the B-spline has the highest smoothing and Welch windowed sinc has the highest post-aliasing property. From this we can suggest that the sensitivity of participants to being able to detect the presence of a path is lowest when there is either a high amount of smoothing or post-aliasing.

From the bias response measure it was found that in 3D the reconstruction filter had a weak effect with the interpolating B-spline producing the least conservative participants and the B-spline producing the most conservative responses. From Figure 6 we can see that the B-spline filter has a considerable amount of smoothing but similar post-aliasing to the interpolating B-spline filter. We can therefore suggest that a filter that exhibits a large amount of smoothing with minimal post-aliasing may make participants more likely to be conservative for this type of task.

Regarding the display, we found that it had a significant effect on both sensitivity and bias. Notice that the simultaneous increase of overall accuracy, sensitivity and bias is a well understood phenomenon related to the shape of the ROC curves the participants operate on, see [MD04].

6. Conclusion

This study evaluated the effect of five reconstruction filters and two transfer functions on a path-searching task with stereoscopic and monoscopic DVR. We found that there was a considerable amount of conservative bias in the responses of the participants, which then was taken into account in the further analysis of the results.

One of the limitations of the experiments presented here is that for each condition they support the computation of one only point on the ROC curve of the average participant. In the future, we plan a repetition of the two experiments, this time asking the participants to rate their confidence on each Yes-No answer they give. Methods from signal detection theory will then be used to process this additional confidence information to obtain multiple points on each ROC curve.

In the future, we plan a meta-analysis of previously published papers on the effect of stereoscopic visualisation on path-searching tasks. Even though the experimental results of such path-searching tasks are customarily reported in a way that makes bias computations impossible, nevertheless, approximations of the most common sensitivity measure d' are possible [MD04] and could yield new interesting insights to their results.

7. Acknowledgements

The blood vessel data set is publicly available at <http://www.volvis.org>.

References

- [ABG*09] AGUS M., BETTIO F., GIACHETTI A., GOBBETTI E., IGLESIAS GUTIÁN J., MARTON F., NILSSON J., PINTORE G.: An interactive 3D medical visualization system based on a light field display. *The Visual Computer* (2009), 883–893. 3, 4
- [ABKP15] ABHARI K., BAXTER J. S. H., KHAN A. L. I. R., PETERS T. M.: Visual enhancement of MR angiography images to facilitate planning of arteriovenous malformation Interventions. *ACM Transactions on Applied Perception* 12, 1 (2015). 3, 4
- [BBD*09] BOUCHENY C., BONNEAU G., DROULEZ J., THIBAUT G., PLOIX S.: A perceptive evaluation of volume rendering techniques. *ACM Transactions on Applied Perception* 5 (2009), 1–24. 3, 4
- [BG07] BRUCKNER S., GRÖLLER E.: Enhancing depth-perception with flexible volumetric halos. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1344–51. 3
- [BTU99] BLU T., THEVENAZ P., UNSER M.: Generalized interpolation: Higher quality at no additional cost. *Proceedings of IEEE International Conference on Image Processing* (1999), 667–671. 5
- [Car93] CARLBOM I.: Optimal filter design for volume reconstruction and visualization. *Proceedings of IEEE Visualization 1993* (1993), 54–61. 5
- [CCAL12] CHEN J., CAI H., AUCHUS A. P., LAIDLAW D. H.:

- Effects of stereo and screen size on the legibility of three-dimensional streamtube visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2130–2139. 4
- [CDW*12] CHO I., DOU W., WARTELL Z., RIBARSKY W., WANG X.: Evaluating depth perception of volumetric data in semi-immersive VR. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2012), pp. 266–269. 3
- [CMHV10] COSTA M., MOREIRA S., HAMER R., VENTURA D.: Effects of age and optical blur on real depth stereoacuity. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians* 30 (2010), 660–6. 1
- [Csé08] CSÉBFALVI B.: An evaluation of prefiltered reconstruction schemes for volume rendering. *IEEE Transactions on Visualization and Computer Graphics* 14 (2008), 289–301. 5
- [CWD*14] CHO I., WARTELL Z., DOU W., WANG X., RIBARSKY W.: Stereo and motion cues effect on depth perception of volumetric data. *Proceedings of SPIE on Stereoscopic Displays and Applications* (2014). 3, 4, 9, 10
- [DV10] DÍAZ J., VÁZQUEZ P.: Depth-enhanced maximum intensity projection. *IEEE/EG International Symposium on Volume Graphics* (2010), 1–8. 4
- [EJH*13] ESCOBAR M., JUHNKE B., HISLEY K., ELIOT D., WINER E.: Assessment of visual-spatial skills in medical context tasks when using monoscopic and stereoscopic visualization. *Proceedings of the 2013 SPIE Medical Imaging Conference 8673* (2013). 3
- [GS13] GROSSET P., SCHOTT M.: Evaluation of depth of field for depth perception in DVR. *Proceedings of IEEE Pacific Visualization 2013* (2013), 1–8. 3, 4
- [HHL10] HASSAINE D., HOLLIMAN N., LIVERSEGE S.: Investigating the performance of path-searching tasks in depth on multiview displays. *ACM Transactions on Applied Perception* 8 (2010), 1–18. 1, 4, 9
- [HHM97] HUBBOLD R., HANCOCK D., MOORE C.: Autostereoscopic display for radiotherapy planning. *Stereoscopic Displays and Virtual Reality Systems 3012* (1997), 16–27. 3
- [JEH03] JOHNSON L. G., EDWARDS P., HAWKES D.: Surface transparency makes stereo overlays unpredictable: the implications for augmented reality. *Studies in Health Technology and Informatics* 94 (2003), 131–136. 10
- [KOCC14] KERSTEN-OERTEL M., CHEN S. J.-S., COLLINS D. L.: An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (Mar. 2014), 391–403. 3, 4, 9
- [KSTE06] KERSTEN M. A., STEWART A. J., TROJE N., ELLIS R.: Enhancing depth perception in translucent volumes. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (2006), 1117–1124. 3, 4, 10
- [LBS14] LAHA B., BOWMAN D. A., SOCHA J. J.: Effects of VR system fidelity on analyzing isosurface visualization of volume datasets. *IEEE Transactions on Visualization and Computer Graphics* 20, 4 (2014), 513–522. 3, 6, 10
- [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 (Dec. 2011), 1922–31. 3, 4
- [LSSB12] LAHA B., SENSHARMA K., SCHIFFBAUER J., BOWMAN D.: Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics* 18 (2012), 597–606. 3, 6, 10
- [MD04] MACMILLAN N. A., DOUGLAS CREELMAN C.: *Detection Theory: A User's Guide*, 2, revised ed. Taylor & Francis, 2004. 4, 9, 11
- [ML94] MARSCHNER S., LOBB R.: An evaluation of reconstruction filters for volume rendering. *Proceedings of IEEE Visualization* (1994), 100–107. 2, 5
- [MN88] MITCHELL D., NETRAVALI A.: Reconstruction filters in computer-graphics. *ACM SIGGRAPH Computer Graphics* 22 (1988), 221–228. 2, 10
- [MSRH08] MEYER-SPRADOW J., ROPINSKI T., HINRICHS K.: Supporting depth and motion perception in medical volume data. *Visualization in Medicine and Life Sciences D* (2008). 4
- [NDS10] NOVOTNY P., DIMITROV L., SRAMEK M.: Enhanced voxelization and representation of objects with sharp details in truncated distance fields. *IEEE Transactions on Visualization and Computer Graphics* 16 (2010), 484–498. 4
- [Pfa00] PFAUTZ J.: *Depth Perception in Computer Graphics*. PhD thesis, 2000. 1
- [RDRS10] ROPINSKI T., DORING C., REZK-SALAMA C.: Interactive volumetric lighting simulating scattering and shadowing. *Proceedings of IEEE Pacific Visualization* (2010), 169–176. 3
- [RI16] ROBERTS D., IVRIȘȘIMȚZIS I.: Quality measures of reconstruction filters for stereoscopic volume rendering. *Computational Visual Media* 2, 1 (2016), 19–30. 10
- [RIH14] ROBERTS D., IVRIȘȘIMȚZIS I., HOLLIMAN N.: An evaluation of reconstruction filters for a path-searching task in 3d. In *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on* (Sept 2014), pp. 165–170. 3, 10
- [RKS13] RAGAN E. D., KOPPER R., SCHUCHARDT P., BOWMAN D. A.: Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics* 19, 5 (2013), 886–896. 4
- [RMSD*08] ROPINSKI T., MEYER-SPRADOW J., DIEPENBROCK S., MENSMAJN J., HINRICHS K.: Interactive volume rendering with dynamic ambient occlusion and color bleeding. *Computer Graphics Forum* 27, 2 (Apr. 2008), 567–576. 3
- [RSH06] ROPINSKI T., STEINICKE F., HINRICHS K.: Visually supporting depth perception in angiography imaging. *Smart Graphics* (2006), 93–104. 3, 4, 9
- [SE03] SVAKHINE N., EBERT D.: Interactive volume illustration and feature halos. *Proceedings of the 11th Pacific Conference on Computer Graphics and Applications III* (2003). 3
- [TBU00] THÉVENAZ P., BLU T., UNSER M.: Interpolation revisited. *IEEE Transactions on Medical Imaging* 19 (2000), 739–58. 2
- [vSvDZS*10] VAN SCHOOTEN B. W., VAN DIJK E. M. A. G., ZUDILOVA-SEINSTRAS E., SUINESIAPUTRA A., REIBER J. H. C.: The effect of stereoscopy and motion cues on 3D interpretation task performance. *Proceedings of the International Conference on Advanced Visual Interfaces - AVI '10* (2010), 167. 4
- [WF96] WARE C., FRANCK G.: Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics* (1996), 1–18. 1, 4, 9
- [WM05] WARE C., MITCHELL P.: Reevaluating stereo and motion cues for visualizing graphs in three dimensions. *Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization* (2005), 51. 1, 4, 9