

A Clinical User Study Investigating the Benefits of Adaptive Volumetric Illumination Sampling

Valentin Kraft, Christian Schumann, Daniela Salzmann, Dirk Weyhe, Gabriel Zachmann, and Andrea Schenk

Abstract—Accurate and fast understanding of the patient's anatomy is crucial in surgical decision making and particularly important in visceral surgery. Sophisticated visualization techniques such as 3D Volume Rendering can aid the surgeon and potentially lead to a benefit for the patient. Recently, we proposed a novel volume rendering technique called Adaptive Volumetric Illumination Sampling (AVIS) that can generate realistic lighting in real-time, even for high resolution images and volumes but without introducing additional image noise. In order to evaluate this new technique, we conducted a randomized, three-period crossover study comparing AVIS to conventional Direct Volume Rendering (DVR) and Path Tracing (PT). CT datasets from 12 patients were evaluated by 10 visceral surgeons who were either senior physicians or experienced specialists. The time needed for answering clinically relevant questions as well as the correctness of the answers were analyzed for each visualization technique. In addition to that, the perceived workload during these tasks was assessed for each technique, respectively. The results of the study indicate that AVIS has an advantage in terms of both time efficiency and most aspects of the perceived workload, while the average correctness of the given answers was very similar for all three methods. In contrast to that, Path Tracing seems to show particularly high values for mental demand and frustration. We plan to repeat a similar study with a larger participant group to consolidate the results.

Index Terms—Visualization, Medical Imaging, Volume Rendering, User Study, Visceral Surgery



1 INTRODUCTION

ACCURATE and fast understanding of the patient's anatomy with all its details (e.g. the vasculature) is crucial in many fields of general surgery and surgical decision making. Since visualization tools and techniques can help the surgeon to better and/or faster understand the patient's anatomy and thus can lead to more correct and faster decisions, they can have a direct influence on the patient outcome [1], [2].

In recent years, many 3D-based visualization techniques, such as Direct Volume Rendering [3] and Path Tracing [4], [5] have been developed or improved and are used to give insights into the human body based on *computer tomography (CT)* and *magnetic resonance imaging (MRI)*. While standard Volume Rendering usually does not provide realistic lighting, Path Tracing – being a physically-based technique – can generate realistic lighting by tracing light paths and computing absorption and scattering on different tissue types.

Several studies showed that a realistic depiction including important visual cues such as realistic lighting and shadows improve spatial understanding and the ability to recognize anatomical structures [6], [7], [8], [9] which is crucial for many clinical/medical application areas, like e.g. visceral surgery or pre-operative planning of tumour surgery [1]. In visceral surgery, accurate visualization for the spatial understanding of the vascular structures including the main arteries and portal venous system is particularly important, for instance in order to choose an appropriate technique or strategy for tumour resection and/or determine the extent of the resection [1]. As in most other surgical fields, time and precision are crucial also in visceral surgery. Additionally, some studies suggest that realistic visualization has advantages also for other fields such as teaching or education [10].

However, as Path Tracing is a progressive algorithm based on a Monte Carlo method, it only converges to a noise-free image over time and introduces severe random image noise on every user interaction such as camera changes, clipping plane changes, or modifications of the *transfer function (TF)*. This is particularly problematic for augmented reality (AR) or virtual reality (VR) where the camera position changes constantly. Furthermore, high requirements on frame rate and resolution as well as the need for stereo images increase the computational demand in AR/VR applications, which leads to worse performance and even more image noise. To remedy this, a new Volume-Rendering-based visualization technique called *Adaptive Volumetric Illumination Sampling (AVIS)* was introduced [11] that is able to generate realistic lighting & shadowing similar to a path tracer in real-time without introducing any additional image noise and even for high resolutions or stereo

- V. Kraft, C. Schumann and A. Schenk are with the Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany. E-mail: valentin.kraft@mevis.fraunhofer.de, christian.schumann@mevis.fraunhofer.de, andrea.schenk@mevis.fraunhofer.de
- C. Schumann is also with the Techsomed GmbH, Bremen, Germany.
- A. Schenk is also with the Department of Diagnostic and Interventional Radiology Hannover Medical School, Hannover, Germany.
- D. Weyhe and D. Salzmann are with the University Medicine Oldenburg, School of Medicine and Health Sciences, University Clinic for Visceral Surgery, Pius Hospital, Oldenburg, Germany. E-mail: dirk.weyhe@pius-hospital.de, daniela.salzmann@apoqlar.com
- G. Zachmann is with the Computer Graphics and Virtual Reality Department, University of Bremen, Bremen, Germany. E-mail: zach@cs.uni-bremen.de

Manuscript received XXX.



Fig. 1. Rendering of an abdomen CT dataset with the three visualization techniques used in the study (left image: Direct Volume Rendering, middle image: AVIS, right image: fully converged Path Tracing). The transfer function has been optimized to visualize the vascular structures which is often crucial in visceral surgery.

rendering.

To investigate possible measurable benefits of this method and compare it to established methods, we conducted a quantitative study comparing AVIS against standard Direct Volume Rendering and Path Tracing. The study tested for the speed and correctness of given answers for relevant questions specific for selected cases in visceral surgery as well as the perceived workload during these tasks using the NASA TLX questionnaire [12]. The study was split into three sessions with a washout period of at least 7 to maximum 14 days to minimize carry-over effects. The 10 participating visceral surgeons were either senior physicians or experienced specialists from the University Medicine in Oldenburg. Although AVIS is specifically well suited for AR & VR as it does not introduce image noise, due to the additional complexity of evaluation and possible side effects, we decided to conduct the study on usual screens and not on a AR/VR headset.

2 RELATED WORK

Our work builds on previous research about volume rendering algorithms. In this section, we provide an overview of the three volume rendering techniques used in the study, their recent advances and respective studies about clinical applications.

2.1 Volume Rendering Algorithms

Direct Volume Rendering (DVR) [3] is a set of techniques that generate 2D projection images of 3D volumetric datasets. Various techniques have been proposed, however, DVR typically uses Ray Casting to sample the volume at certain intervals or positions on each ray that is traced through the volume. Usually, only primary rays are considered, i.e. no secondary rays as needed for Global Illumination effects such as shadows, reflection or refraction are taken into account due to the involved computational efforts. Several works try to include such effects by employing various approximation schemes or caching [13], [14].

Physically plausible and highly realistic results can be achieved using *Path Tracing (PT)*, a progressive physically-based Monte-Carlo rendering technique that enables Global Illumination phenomena such as indirect light and realistic shadows. It pursues a progressive approach for solving the rendering equation, thus convergence to a noise-free image usually requires several seconds during which no visualization parameters can be changed. Originating from surface rendering, it has been adapted to DVR [4], [5], where the computational demand is even higher, as continuous changes of the density and color through the volume have to be considered instead of discrete light interactions at surfaces. As a result, the generated images suffer from strong image noise during camera interaction, TF changes and clipping. Recently, Path Tracing has been improved by employing temporal reprojection [15]. Even though this greatly improved frame rates, the resulting images are still suffering from noise and blurriness to a certain degree. The adaptation of Path Tracing in the field of medical visualization has been moved forward by the product *Cinematic Rendering (CR)* by "Siemens Healthineers".¹

Adaptive Volumetric Illumination Sampling (AVIS) [11] is a GPU-based DVR method that enables realistic lighting at high framerates and with high resolutions. It reduces the number of illumination calculations adaptively during ray casting. *Voxel cone tracing* [16] is adapted to compute *Ambient Occlusion (AO)* [17] in combination with *image-based lighting* [18] to generate a realistic lighting approximation for the determined samples. The resulting rendering method allows to compute noise-free images with realistic lighting effects at very high resolutions and frame rates while supporting both interactive transfer function updates and clipping of the visualized data with only minimal precomputation. It thus is particularly well suited for AR and VR.

1. <https://www.siemens-healthineers.com/medical-imaging-it/syngo-carbon-products/cinematic-rendering>

2.2 Volume Rendering Studies

Volume Rendering techniques were tested in clinical use cases on several occasions, and Cinematic Rendering is particularly well studied.

Binder et al. [10] compared CR and conventional 2D CT imaging in terms of speed and comprehension of anatomy using a two-period crossover study design and with medical students as the participants. The result of the study yielded good results for CR, however, they compared the 3D-based CR methodology with advanced lighting to a 2D-based technique without lighting or colouring and did not include other 3D-based visualization techniques.

Using a similar randomized 2-sequence crossover study design, Elshafei et al. [19] compared conventional 2D CT visualization and CR with regard to anatomic understanding, preoperative planning and intraoperative strategies. The participants were resident and attending surgeons and the results suggest that CR is beneficial for both efficiency and correctness when answering clinically relevant questions. Elshafei et al. did not include another 3D-based visualization technique, such as DVR, into their study, as well.

Li et al. [20] conducted a study to assess the value of CR for evaluating the relationship between deep soft tissue sarcomas and adjacent vessels using two experienced radiologists. They compared CR against conventional 3D Volume Rendering with the result that CR showed lower accuracy, sensitivity, specificity, positive and negative predictive values for vascular invasion diagnosis than the traditional methods, although the results were not statistically significant.

Wollschlaeger et al. [21] compared CR to standard volume rendering with regard to the preoperative visualization of multifragmentary intraarticular lower extremity fractures, but only using two experts for evaluation. The results suggested that CR demonstrated a higher image quality, a higher anatomical accuracy and provided a more detailed visualization of the fracture than DVR.

Fukumoto et al. [22] compared standard Volume Rendering to a Volume Renderer with global illumination in the context of forensic evaluation of stab wounds by three radiologists with the result that the global illumination renderings got higher image-quality scores and the ability to assess the stab wounds was significantly better as well.

3 MATERIAL & METHODS

This preclinical study with a randomized three-period crossover design was conducted over a time span of two months at the University Clinic for Visceral Surgery, Pius Hospital in Oldenburg. The three methods that should be compared against each other were our novel AVIS rendering technique [11] and the two most established methodologies in 3D medical visualization: standard Volume Rendering (with shading, but without shadowing) and Path Tracing (implementation based on Kroes et al. [4] with a KNN-based denoiser) (see also Fig. 1). While the standard DVR uses a Blinn-Phong shading model and a mixture of ambient light and directional lights, AVIS was set to use only ambient light and its ambient-occlusion based shadows. The Path Tracer was configured similarly with ambient lighting only, a matte material and no indirect light bounces.

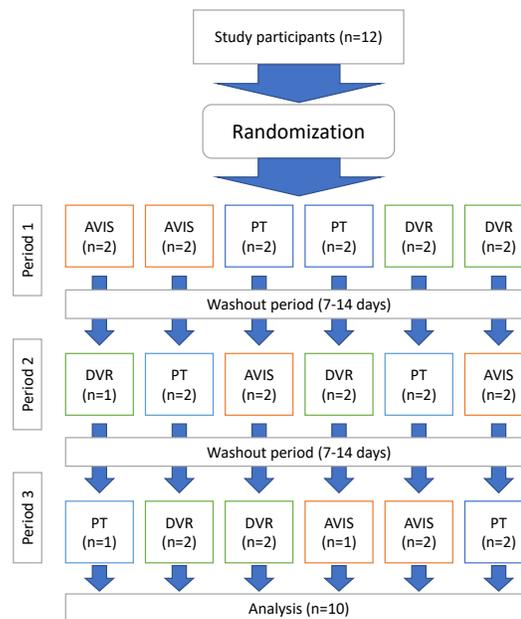


Fig. 2. A schematic overview over the study procedure for the selected participants ($n = 12$). During the study, two participants had to be excluded because they could not comply with the 7 to 14 days washout period.

3.1 Participants

12 volunteers agreed to participate in this study and 10 successfully finished it, since two participants had to be excluded due to exceeding the maximum 14 days washout period between the sessions. The participants were all visceral surgeons and either senior physicians or experienced specialists from the University Clinic for Visceral Surgery, Pius Hospital in Oldenburg (8 male and 2 female, mean age: 45.8 ± 9.2 years, median age: 43.5 years). They had 18.2 ± 8.3 years of professional experience on average (median: 15.0 years). Their prior knowledge with 3D software was assessed with a 3-point likert scale (0: no prior knowledge, 1: little prior knowledge, 2: much prior knowledge) and was rather low (mean: 0.5). Their technical affinity was assessed with a 3-point likert scale as well (0: low affinity, 1: medium affinity, 2: high affinity) and was average (mean: 1.1). None of the study participants had seen visualizations based on AVIS before.

3.2 Study Design & Procedure

Based on the three methods that should be compared, six designs were created; each for every permutation of the three methods (D01-D06, see Fig. 2). For each design D , two participants were randomly assigned, fitting the number of initial participants very well. For every participant, the technical system (desktop PC or laptop) was chosen for the whole experiment and, of course, could not be switched after starting the experiment due to the different hardware specifications of the systems. Now, each participant had to attend three sessions - each for one of the respective randomly chosen method in the respective chosen order. For each session, a random permutation of the 12 anonymized

datasets was generated (without repetition). The participants did not know which method they were assessing and had not seen the study software before. Furthermore, also the UI and interaction did not differ between the methods. A washout period of minimum 7 days and maximum 14 days between the assessments was chosen to avoid carry-over effects. If a participant did not manage to finish the session on time, he/she was excluded from the experiment, which happened to two participants, thus 10 participants successfully finished the study.

Our co-primary hypotheses for the study were that for the new AVIS method

- The probability of a correct answer is non-inferior compared to the other methods (H1) AND
- Time spent answering is superior (lower) compared to the other methods (H2)

Each session then followed the given schema (see Fig. 2): First, the participant had to answer the demographic questionnaire (only for first session, see supplemental content). Then, to familiarize with the program, a training dataset was shown, which was identical for all participants, and the rules, task and functionalities of the application were explained in detail to the participant. In addition to that, the participant had the possibility to ask questions anytime during the whole experiment. Then, the actual study started: For each of the 12 datasets, the participant had to answer three clinically relevant, visceral-specific questions. The questions were binary yes/no-questions and stayed the same for each dataset. They were specifically designed from expert visceral surgeons that did not participate in the experiment to be clinically relevant and tailored to the respective dataset. The questions were for instance: "The hepatica communis arises with the superior mesenteric artery?" or "The superior mesenteric artery crosses the venous confluence?" (translated from German, see appendix for full question catalogue). In order to answer these questions, they had to use all of the interaction possibilities (change the TF, clipping planes, 3D interaction), since the initial TF was not optimal and no clipping planes were set. Hence, only the skin of the patient was shown on the start of each case. In addition to that, the participants always had to answer the following questions: "How hard were the questions in this specific case?" and "How much did the visualization help in answering these questions?" (likert scale with 4 options). We recorded every answer together with the required timings to later determine the correctness and speed of the answers as well as other relevant parameters, such as the used transfer function, clipping parameters, screenshots, etc. Then, we asked the participant to answer the NASA TLX questionnaire with respect to the last dataset of each session to see which influence the visualization technique had on the perceived workload after getting familiar with the application.

3.3 Study Data

In the study, we used 12 fully anonymized, standard CT abdomen datasets from the Pius Hospital with a focus on pancreas. The resolutions varied between 512x512 and 849x512 voxel with a minimum of 64 slices and a maximum



Fig. 3. A screenshot from the study application, as used by a study participant, showing a volume rendering of a CT dataset with the AVIS visualization technique and a clipping plane in front of the dataset.

of 265 slices with 1.6 to 3.0 mm slice thickness. The images originated from different scanner models and the arterial phase was used in most cases, while some few were taken in the venous phase.

The study was conducted within the VIVATOP project and data processing was approved by the Medical Ethics Committee of the University of Oldenburg (No. 2021-013).

3.4 Technical Equipment

Two systems were used for the study: A workstation PC with Windows 10 operating system, Intel(R) Core(TM) i7-9800X CPU @ 3.80GHz, Nvidia GeForce RTX 2080 Ti and a laptop with Windows 10 operating system, Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz and Nvidia GeForce GTX 1050.

For the study, a custom software application based on MeVisLab [23], a powerful software development framework for medical imaging, was built. The application offered multiple relevant functionalities that are often also available in established medical products to make the handling as intuitive as possible (see Fig. 3). It offered means to clip the volume on each dimensional axis, that is, via a clipping cube. In addition to that, the user was able to change the transfer function via dragging the mouse cursor while holding down the right mouse button, thus changing the level and contrast of the transfer function. This enabled the user to efficiently hide or show different tissue structures, such as muscles, bones or vessels. The rotation of the camera was free in all directions and followed a common 3D camera interaction schema. In addition to that, the user had the possibility (and was told to) pause the application once he/she was distracted (e.g. by a phone call) so that we get timings which were as precise as possible. Since each participant was supervised for the whole study, we thus could ensure that the timings were consistent.

The software recorded all relevant parameters (used method & dataset, transfer function parameters, timings, given answers, etc.) during the experiment in the background in JSON-encoded files to later being able to reconstruct a specific situation, if necessary, and to make the data analysis as flexible and easy as possible. In addition to the JSON files, screenshots of the User UI when

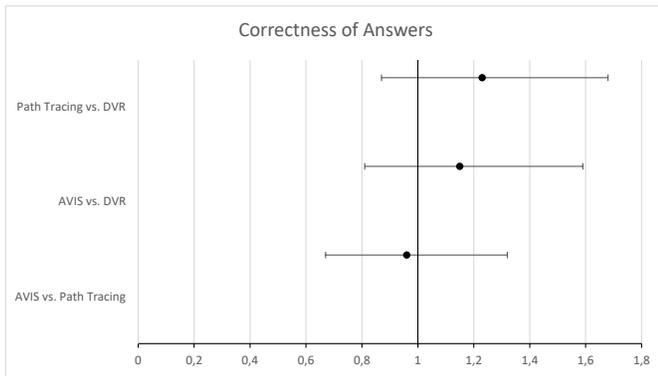


Fig. 4. Visualization of odds ratios (A vs. B: $odds(A)/odds(B)$) and corresponding 95% confidence intervals for effect contrasts with regard to the correctness of the given answers.

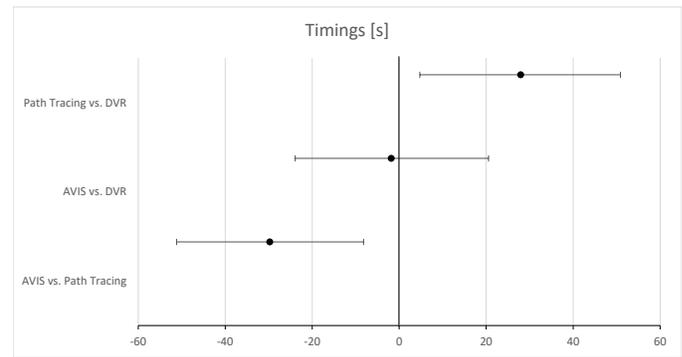


Fig. 5. Visualization of the overall mean time differences for the given answers in seconds (A vs. B: $time(A) - time(B)$) with 95% confidence intervals. It is obvious that AVIS and DVR are performing quite similar, while AVIS and DVR enabled the participants to give significantly faster answers on average in comparison to PT.

answering the questions were taken as well (see Fig. 3).

3.5 Statistical Analysis

The main statistical analysis was planned in accordance with the two pre-specified hypotheses regarding the endpoints correctness and time (compare section 3.2).

For the analysis of the proportion of correct answers, a generalized linear mixed model (GLMM) with outcome correctness (per individual question) was fitted. A Bernoulli link function was employed to model the binary outcome. The different methods (AVIS, DVR, PT) entered the model as a categorical covariate. Furthermore, the analysis was adjusted for sequence, period (fixed effects) as well as study participant and question (random effects).

For the second, continuous outcome time (in seconds, per set of questions per case), a (gaussian/normal) mixed linear model with similar covariates as mentioned before was utilized. The overall null hypothesis can be rejected if AVIS is non-inferior to DVR and PT with respect to correctness (non-inferior margin: 10%) and in addition superior compared to DVR and PT with respect to time. The specification as co-primary endpoints allows to conduct each analysis independently at significance level 5% without adjustment for multiplicity [24].

The analysis was performed with R [25] (version 4.1.2) with help of the brms package [26], [27].

The analysis of the NASA TLX questionnaire was mostly done in Microsoft Excel and was mainly focused on mean and median values and the respective standard deviations.

4 RESULTS

During the whole study, 10 experienced surgeons completed a total of 360 case evaluations (120 per method). We tested for the proportion of correct answers, the time needed to give them and the perceived workload.

4.1 Correctness & Timings

The results show that no method is significantly non-inferior with respect to the proportion of correct answers when

the non-inferiority margin is 10% or less (cf. Fig. 4). Also the mean correctness of the given answers show only very subtle differences between all three methods (DVR: 69.4%, AVIS: 71.4%, PT: 72.5%).

The mean timings over all participants and sessions (DVR: 182.4s, AVIS: 176.3s, PT: 206.2s) are lowest for AVIS with a difference of 29.9s (14.5%) to PT which had the highest mean times of the methods (cf. Fig. 5). These results are confirmed in sensitivity analyses (exponentially modified Gaussian instead of Gaussian outcome distribution) but with reduced effect sizes.

The questions were perceived as not too hard and not too easy on the 4-point Likert scale (relative frequencies over all questions and participants: "easy": 6.4%, "rather easy": 35.0%, "rather hard": 48.9%, "hard": 9.7%).

Based on the results, it thus can be stated that users are significantly faster in answering the questions with DVR and AVIS than with Path Tracing, however, there is no evidence for a difference between AVIS and DVR in this regard. The non-inferiority regarding the proportion of correct answers could not be proven since the results are not statistically significant, although the mean correctness suggests that all methods are very similar in this aspect. Thus, no method fulfills both hypotheses H1 and H2, since our first hypothesis (H1) cannot be answered with certainty and the second one (H2) is only true for AVIS/DVR and Path Tracing.

4.2 Perceived Workload

The perceived workload of the participants during the assessments was measured with the NASA TLX questionnaire. A box plot of the results can be seen in Fig. 6.

The results illustrate that, on average, Path Tracing shows high values for mental and physical demand as well as for frustration and effort. The values for mental demand and frustration are particularly high. In contrast, AVIS has the best results for physical and temporal demand as well as for the frustration, although the difference to DVR is very little in the latter. AVIS only performs worst for the section self performance, although the mean values for all three methods are very close here. Remarkably, AVIS outperforms

PT in every other aspect. DVR seems to have values in between and leads in the effort section only. However, the results are rather close to the AVIS results for mental demand, performance and frustration.

The NASA TLX scores were normalized to a scale from -10 to 10 to simplify the interpretation.

4.3 Sub-group analysis

When taking a closer look at the results for the participants who used the laptop compared to the participants who used the desktop PC, there are some differences that stand out. As for the used overall time, the laptop users were generally substantially faster than the overall group (DVR: 146.8s (-19.5%), AVIS: 148.8s (-15.6%), PT: 131.3s (-36.3%)). Also the proportion of correct answers is different for AVIS and PT compared to the overall group (DVR: 69.4% (+0.0%), AVIS: 68.5% (-4.0%), PT: 66.7% (-8.0%)). There are also differences to the overall group in terms of the times in which users interacted with the camera (DVR: -16.7%, AVIS: -6.1%, PT: -41.9%).

5 DISCUSSION

Although the study did not yield statistically significant differences regarding the proportion of correct answers between the methods, it shows that DVR and AVIS outperform PT when it comes to the time needed for answering the clinically relevant questions. This finding is consistent with the fact that the user does not have to wait for the noise reduction when using PT. AVIS demonstrates even slightly better time efficiency than DVR on average, while still delivering realistic shadows and lighting as crucial depth cues, unlike DVR. The speed-up of 14.5% on average in our investigation compared to PT is remarkable, given the fact that time efficiency is a highly relevant topic in the medical field and that PT seems to be even significantly more time efficient than 2D slicing [10], [19].

The results of the NASA TLX questionnaire (cf. Fig. 6) are also quite remarkable: It is obvious that Path Tracing arguably has the worst results with quite high effort and physical demand and particularly high values for frustration and mental demand. This could relate to the fact that the image noise and the need to wait for it to disappear increases the frustration and/or mental demand of the user. This is especially interesting since we already used a Path Tracer with a KNN-based denoiser that should reduce the image noise but has the disadvantage of blurring the image and potentially losing important image details, which might be particularly critical in a medical context. We therefore argue that even a denoising does not solve the problems that occur due to additional image noise and that AVIS thus has a clear advantage over Path Tracing with respect to image noise. We further speculate that the very high value for mental demand when using Path Tracing (cf. Fig. 6) could be based on the fact that the brain already starts to decipher the image while it is still noisy or blurry. The results for temporal demand align with our findings on average timings, indicating that AVIS is the fastest method. Interestingly, DVR shows the highest average value for temporal demand although the average timings were quite similar to those of AVIS. We

could imagine that this is connected to the missing depth cues of DVR, however, one would probably expect higher average timings for DVR then. Also the great differences in the physical demand section seem to be strange, given the fact that the physical demand should have been rather equal for each method since the physical interaction was identical for all methods. Interestingly, the proportions between the average values seem to be similar to the ones for the frustration, which could lead to the conclusion that those two might be connected. It is evident, however, that AVIS shows best results in physical and temporal demand and performs approximately equally good as DVR for frustration, mental demand and self performance. The only exception where AVIS is not best or equally good as DVR is for the effort. Here, DVR performs best, which could potentially hint to the fact that the shadows may be also distracting to some of the participants. This, however, would somewhat contradict with other findings, e.g. with the results from Lindemann et al. [7]. One possible reason could be rooted in the lighting settings we used, as other studies have found out that global illumination methods are less helpful if the light sources are not placed properly [8]. On the other hand, Li et al. [20] also could not find positive effects when using Cinematic Rendering instead of standard DVR and their results were not statistically significant as well.

Although we could not prove both our hypotheses, we believe that the study results indicate that AVIS offers the "best of both worlds": The advantages of DVR, including enhanced time efficiency, no additional image noise and reduced frustration and lower mental demand, while still producing realistic lighting similar to Path Tracing that can improve spatial understanding. Although the resulting positive effects from enhanced spatial understanding were not directly represented in the data we collected, these advantages are evident and already proven in other studies. We think that in contrast to that, the results furthermore suggest that for Path Tracing the introduction of noise may result in higher frustration and mental demand for the user and worse timings in decision making. We propose to investigate this in more depth in further studies.

From the study participants, we also got mostly positive feedback for AVIS. Most physicians saw a clear advantage in viewing the data in 3D and also in the advanced lighting and shadowing capabilities. One participant for instance stated: "I could better concentrate on the vessels" when using AVIS. In addition, we also noticed some general things during our study, e.g. that a majority of the participants initially configured the clipping planes & TF in a similar fashion as used in conventional 2D slicing tools, i.e. with a very "hard" (high-contrast) TF and viewing along the dimensional axes. Later, however, they tended to extensively use the 3D view & "softer" TF with more transparency to really "segment" out relevant structures and vessels (see Fig. 3). This hints to the fact that the capabilities of 3D visualization software may be unfamiliar to many at the beginning but seem to have advantages over the 2D view for most participants. Furthermore, we noticed that the learning effect seemed to be very strong and also the used time per session vastly decreased for the later sessions. We therefore also incorporated this effect in our statistical model.

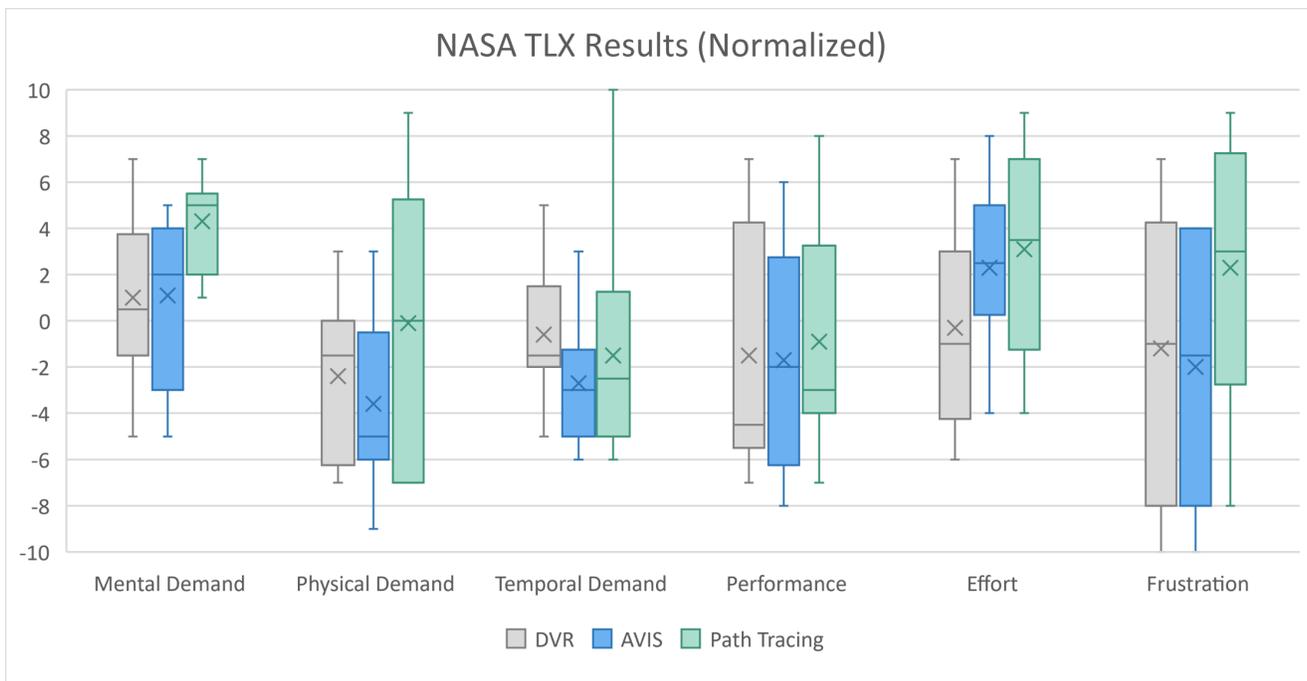


Fig. 6. Box plot of the results of the NASA TLX questionnaire indicating the perceived workload of the participants during the study for each methodology. Lower values are better except for the (self)-performance.

5.1 Limitations

As stated before, the study did not yield statistically significant differences regarding the proportion of correct answers between the methods. We speculate that the reason is that the data variability is too large, which could be rooted in the small size of the participant group. However, as we designed and conducted the study with highly experienced specialists that are hard to acquire, this limitation was difficult to overcome. The fact that all our participants were experts might also be a factor for the small differences in the proportions of correct answers between the methods, since they may rely more on their experience as a surgeon than on the visualization itself. We would therefore like to suggest for future work to repeat a similar study with more participants. We would advocate for not widening the target group to really prove clinical relevance. Furthermore, as in all crossover studies, carryover effects by memorization are potential source of bias and cannot be excluded for this study, either. As described in section 4.3, there also might be some important differences between the laptop and desktop user subgroups. We would like to address this interesting hints in future work as well. However, given the interaction times, which differ greatly from the overall group, especially for PT, we believe it is most likely that users generally enjoyed working with the laptop less, and particularly when using PT. We speculate that one reason could be that the even severer image noise caused by the weaker laptop hardware has demotivated users to engage further with the system, although speculations like this should be investigated in the future studies and we found no evidence in the NASA TLX results of the laptop subgroup to support this hypothesis.

6 CONCLUSION

We conducted a three-period crossover study with highly experienced specialists in visceral surgery to investigate the possible benefits of the new Volume-Rendering-based method AVIS by comparing it to the established methods standard Volume Rendering and Path Tracing. The study investigated the time needed to give answers for relevant questions in visceral surgery and their correctness as well as the perceived workload during these tasks using the NASA TLX questionnaire.

The results from the NASA TLX questionnaire suggest that AVIS outperforms the other methods for the physical and temporal demand and is as good as DVR when it comes to frustration and mental demand. In contrast to that, these values are particularly high for Path Tracing which may be caused by the fact that the user has to wait for the image noise to reduce for every interaction. These results are in line with the fact that the timings for giving a correct answer are worst for Path Tracing. AVIS has the best results regarding the used time and offered an average speed-up in user performance of 14.5% in our experiment. Regarding the correctness of the given answers, the differences between the methods are very subtle, meaning that no method is significantly inferior or superior to the other methods.

To our knowledge, we have included more expert participants in our user study than most other similar studies; however, we were still not able to reach statistical significance, since it is extremely difficult to acquire enough clinical experts, and experts can draw on a lot more pre-existing knowledge than what is conveyed in any visualization. For future work, we would still suggest to repeat a similar study with a larger participant group – if possible – to consolidate

the results, since AVIS seems to be a promising new method with potential benefits for accelerating the decision making process and to further investigate the influence of image noise on the perceived workload of the user. As there may be substantial differences between the laptop and desktop subgroups on some elements of the study, we also plan to investigate these possible differences in more detail in future work. We would furthermore encourage to extent the study to other relevant clinical fields. In addition to that, it would be also interesting to do a similar comparison using AR or VR headsets (e.g. Microsoft HoloLens) instead of conventional screens.

ACKNOWLEDGMENTS

We thank our partners, especially Prof. Dr. Weyhe and his team at University Clinic for Visceral Surgery, Pius Hospital in Oldenburg. This research has been funded under grant number 16SV8078 by the German Federal Ministry of Education and Research (BMBF) in the project VIVATOP. We would also like to thank all participants and especially Max Westphal, who significantly helped with building the statistical model and analyzing the results.

REFERENCES

- [1] A. A. Javed, R. W. Young, J. R. Habib, B. Kinny-Köster, S. M. Cohen, E. K. Fishman, and C. L. Wolfgang, "Cinematic Rendering: Novel Tool for Improving Pancreatic Cancer Surgical Planning," *Current Problems in Diagnostic Radiology*, vol. 51, no. 6, pp. 878–883, 11 2022.
- [2] H. Cheng, J. W. Clymer, B. Po-Han Chen, B. Sadeghirad, N. C. Ferko, C. G. Cameron, and P. Hinoul, "Prolonged operative duration is associated with complications: a systematic review and meta-analysis," *Journal of Surgical Research*, vol. 229, pp. 134–144, 9 2018.
- [3] R. A. Drebin, L. Carpenter, and P. Hanrahan, "Volume rendering," *Computer Graphics (ACM)*, vol. 22, no. 4, pp. 65–74, 8 1988.
- [4] T. Kroes, F. H. Post, and C. P. Botha, "Exposure render: An interactive photo-realistic volume rendering framework," *PLoS ONE*, vol. 7, no. 7, 2012.
- [5] B. Csebfalvi and L. Szirmay-Kalos, "Monte Carlo volume rendering," in *VIS'03 Proceedings of the 14th IEEE Visualization*. IEEE, 2003, pp. 449–456.
- [6] M. S. Langer and H. H. Bülthoff, "Depth discrimination from shading under diffuse lighting," *Perception*, vol. 29, no. 6, pp. 649–660, 6 2000.
- [7] F. Lindemann and T. Ropinski, "About the influence of illumination models on image comprehension in direct volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 1922–1931, 12 2011.
- [8] L. Zheng, A. J. Chaudhari, R. D. Badawi, and K. L. Ma, "Using global illumination in volume visualization of rheumatoid arthritis CT data," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 16–23, 11 2014.
- [9] J. Díaz, T. Ropinski, I. Navazo, E. Gobbetti, and P.-P. Vázquez, "An experimental study on the effects of shading in 3D perception of volumetric models," *The Visual Computer*, vol. 33, no. 1, pp. 47–61, 1 2017.
- [10] J. S. Binder, M. Scholz, S. Ellmann, M. Uder, R. Grützmann, G. F. Weber, and C. Krautz, "Cinematic Rendering in Anatomy: A Crossover Study Comparing a Novel 3D Reconstruction Technique to Conventional Computed Tomography," *Anatomical Sciences Education*, vol. 14, no. 1, pp. 22–31, 1 2021.
- [11] V. Kraft, F. Link, A. Schenk, and C. Schumann, "Adaptive Illumination Sampling for Direct Volume Rendering," in *Advances in Computer Graphics*, 12221st ed., N. Magnenat-Thalmann, C. Stephanidis, E. Wu, D. Thalmann, B. Sheng, J. Kim, G. Papa- giannakis, and M. Gavrilova, Eds. Springer, Cham, 2020, vol. 12221 LNCS, pp. 107–118.
- [12] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," 1988, pp. 139–183.
- [13] F. Hernel, P. Ljung, and A. Ynnerman, "Interactive global light propagation in direct volume rendering using local piecewise integration," *SPBG'08 Proceedings of the Fifth Eurographics / IEEE VGTC conference on Point-Based Graphics*, pp. 105–112, 2008.
- [14] T. Ropinski, C. Döring, and C. Rezk-Salama, "Interactive volumetric lighting simulating scattering and shadowing," in *IEEE Pacific Visualization Symposium 2010, PacificVis 2010 - Proceedings*. IEEE, 3 2010, pp. 169–176.
- [15] J. Martschinke, S. Hartnagel, B. Keinert, K. Engel, and M. Stamminger, "Adaptive Temporal Sampling for Volumetric Path Tracing of Medical Data," *Computer Graphics Forum*, vol. 38, no. 4, pp. 67–76, 2019.
- [16] C. Crassin, F. Neyret, M. Sainz, S. Green, and E. Eisemann, "Interactive indirect illumination using voxel cone tracing," *Computer Graphics Forum*, vol. 30, no. 7, pp. 1921–1930, 9 2011.
- [17] S. Zhukov, A. Iones, and G. Kronin, "An ambient light illumination model," in *Rendering Techniques '98*, 1998, pp. 45–55.
- [18] H. Landis, "Production-ready global illumination," in *Siggraph course notes*, 2002, pp. 87–101.
- [19] M. Elshafei, J. Binder, J. Baecker, M. Brunner, M. Uder, G. F. Weber, R. Grützmann, and C. Krautz, "Comparison of Cinematic Rendering and Computed Tomography for Speed and Comprehension of Surgical Anatomy," *JAMA Surgery*, vol. 154, no. 8, p. 738, 8 2019.
- [20] K. Li, R. Yan, H. Ma, D.-f. Zhang, Y. Ding, and Z.-h. Li, "Value of the Cinematic Rendering From Volumetric Computed Tomography Data in Evaluating the Relationship Between Deep Soft Tissue Sarcomas of the Extremities and Adjacent Major Vessels," *Journal of Computer Assisted Tomography*, vol. 43, no. 3, pp. 386–391, 2019.
- [21] L. M. Wollschlaeger, J. Boos, P. Jungbluth, J.-P. Grassmann, C. Schleich, D. Latz, P. Kroepil, G. Antoch, J. Windolf, and B. M. Schaarschmidt, "Is CT-based cinematic rendering superior to volume rendering technique in the preoperative evaluation of multifragmentary intraarticular lower extremity fractures?" *European Journal of Radiology*, vol. 126, p. 108911, 5 2020.
- [22] W. Fukumoto, N. Kitera, H. Mitani, T. Sueoka, S. Kondo, I. Kawashita, Y. Nakamura, M. Nagao, and K. Awai, "Global illumination rendering versus volume rendering for the forensic evaluation of stab wounds using computed tomography," *Scientific Reports*, vol. 12, no. 1, p. 2452, 2 2022.
- [23] F. Ritter, T. Boskamp, A. Homeyer, H. Laue, M. Schwier, F. Link, and H.-O. Peitgen, "Medical Image Analysis," *IEEE Pulse*, vol. 2, no. 6, pp. 60–70, 11 2011.
- [24] A. Dmitrienko and R. B. D'Agostino, "Multiplicity Considerations in Clinical Trials," *New England Journal of Medicine*, vol. 378, no. 22, pp. 2115–2122, 5 2018.
- [25] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2022.
- [26] P.-C. Bürkner, "brms: An R Package for Bayesian Multilevel Models Using Stan," *Journal of Statistical Software*, vol. 80, no. 1, 2017.
- [27] P.-C. Buerkner, "Advanced Bayesian Multilevel Modeling with the R Package brms," *The R Journal*, vol. 10, no. 1, p. 395, 2018.