Authors' accepted manuscript
Article published in
IEEE Transactions on Visualization and Computer Graphics, Vol. 26, no 6, pp.2247-2257, 2020.
The final published version available at https://doi.org/10.1109/TVCG.2018.2884940

Interaction driven enhancement of depth perception in angiographic volumes

Simon Drouin, Daniel A. Di Giovanni, Marta Kersten-Oertel, D. Louis Collins

Abstract— User interaction has the potential to greatly facilitate the exploration and understanding of 3D medical images for diagnosis and treatment. However, in certain specialized environments such as in an operating room (OR), technical and physical constraints such as the need to enforce strict sterility rules, make interaction challenging. In this paper, we propose to facilitate the intraoperative exploration of angiographic volumes by leveraging the motion of a tracked surgical pointer, a tool that is already manipulated by the surgeon when using a navigation system in the OR. We designed and implemented three interactive rendering techniques based on this principle. The benefit of each of these techniques is compared to its non-interactive counterpart in a psychophysics experiment where 20 medical imaging experts were asked to perform a reaching/targeting task while visualizing a 3D volume of angiographic data. The study showed a significant improvement of the appreciation of local vascular structure when using dynamic techniques, while not having a negative impact on the appreciation of the global structure and only a marginal impact on the execution speed. A qualitative evaluation of the different techniques showed a preference for dynamic chroma-depth in accordance with the objective metrics but a discrepancy between objective and subjective measures for dynamic aerial perspective and shading.

Index Terms— Image-guided surgery, Volume visualization, Interaction techniques, Depth cues, Evaluation, Angiography

1 Introduction

The human visual system models the environment by performing "a dynamic searching for the best interpretation of the available data" [1]. For this reason, interaction is key, if not essential, to the understanding of 3D data presented on a 2D computer screen where many natural depth cues may be absent. Interaction paradigms used for the exploration of 3D models on a desktop computer are well established. Simple mechanisms that use a mouse (often combined with a keyboard) are almost universally adopted and enable a rapid understanding of the 3D structure of the data being visualized. However, in certain specialized environments, technical and physical constraints make the conventional paradigms impractical. This is the case in operating rooms where surgeons make use of image-guided neurosurgery (IGNS) systems. Such systems enable the visualization of brain scans (e.g. MRI, CT) during surgery to help guide the operation. The location of a 3D tracked pointer is overlaid with the rendering of the scans, allowing the surgeon to understand the correspondence between the images on the screen and the patient on the operating room table as illustrated in Figure 1 (a) and (b).

Existing systems allow for very little interaction on the part of the surgeon which makes it difficult to correctly interpret the 3D structure of the anatomy displayed on the screen. While introducing interaction mechanisms may help to disambiguate the images, conventional interaction paradigms are not appropriate for the task. Bringing a mouse or another physical interaction device in the sterile field of the OR is challenging and typically not ergonomically feasible. In practice, surgeons have to rely exclusively on verbal dictation to a technician or assistant to interact with the system [2]. More important than the interaction device is the paradigm of interaction. The conventional approach, which consists of presenting data from different point of views to facilitate understanding, is not appropriate during surgical navigation. In this phase, surgeons typically point at several parts of the anatomy using the tracked pointer while trying to mentally map the image on the screen to the anatomy of the patient. At this point, preoperative scans need to be presented to the surgeon from a point of view that is aligned with her/his own perspective of the patient. Any relative rotation or translation of the frames of reference makes the mapping task considerably more mentally challenging and error prone [3]–[5].

In this paper, we propose several variants of a new interaction paradigm that overcome both constraints by using the motion of the tracked surgical pointer to enhance depth cues in the image rendered on the screen of the navigation system. This form of interaction uses a piece of equipment already present in the operating room and enables a better understanding of the images rendered on a screen while avoiding viewpoint changes to facilitate the aforementioned mental mapping by the surgeon. We evaluate the proposed interactive rendering techniques in a user study where participants are requested to perform a reaching task to position the pointer close to anatomical landmarks. In the study, we make use of angiographic

- *Simon Drouin, Daniel A. Di Giovanni, and D. Louis Collins are with the Department of Biomedical Engineering at McGill University in the McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal, Canada.*
- *Marta Kersten-Oertel is with the Department of Computer Science and Software Engineering at Concordia University, Montreal, Canada.*
- *Corresponding E-mail: simon.drouin@mail.mcgill.ca.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

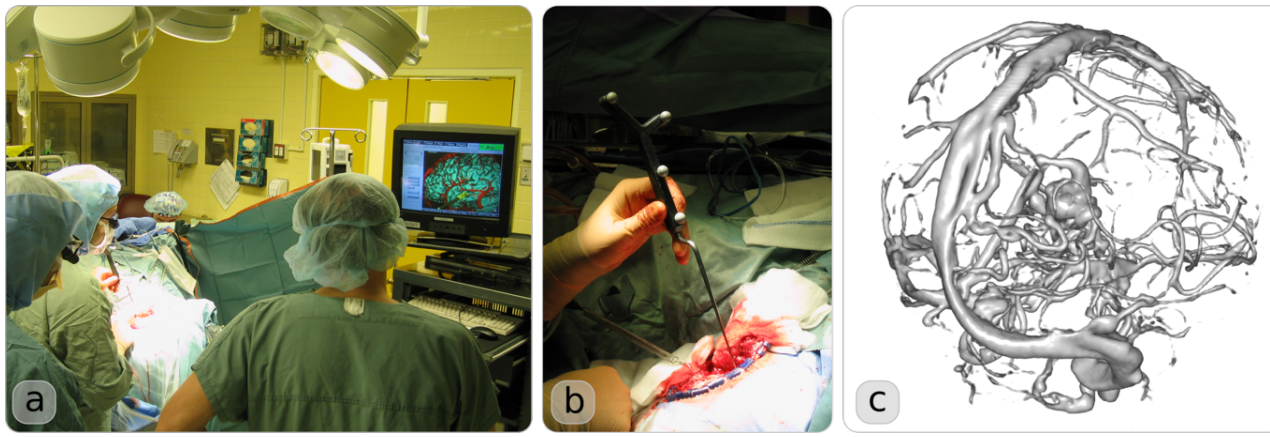


Figure 1 - Illustration of the use of navigation systems during a neurovascular intervention. (a) Overview of the OR setup, showing the surgeon pointing at the anatomy with a tracked tool while looking at the screen of the navigation system. (b) Enlarged view of the tracked pointer manipulated by the surgeon. (c) Example of a 3D rendered angiographic CT of a patient operated for an arteriovenous malformation (AVM).

volumes which are typically used in neurovascular interventions to treat aneurysms and arteriovenous malformations (AVMs), more specifically, we use computed tomography angiograms (CTA) (Figure 1c), which is part of the standard imaging protocol for neurovascular interventions at our institute. The visualization of this type of data is particularly challenging due to the complex three-dimensional structure of vessel trees and the large inter-patient anatomical variability, especially in the abnormal cases that require an operation.

The rest of the article is structured as follows. We first review related work on the rendering of angiographic volumes as well as interactive techniques that have been applied to surgical guidance. Then, we describe each of the interactive rendering methods that we designed and the psychophysics experiment that was used to evaluate the benefit of each method. Finally, we present the results of the study as well as results of a questionnaire filled by participants of the study and draw conclusions.

2 Background and Related Work

Rendering of angiographic volumes has received a lot of attention due to the complexity of vascular structures and the difficulty to interpret them given the large inter-patient anatomical variability. In addition to the usual requirements for the rendering of 3D medical images, the visualization method employed should enable the perception of distance between vessels occluding each other as well as between vessels and structures of interest (e.g. tumour). Additionally, when information is available, the visualization should allow to distinguish between arterial and venous systems [6].

Ropinski et al. [7] have designed a series of depth enhancing direct volume rendering methods including three different edge enhancement methods, color encoded depth (chroma-depth and its two-color variant, pseudo chroma-depth), stereo (autostereoscopic screen) and depth of focus. They compared each method with simple direct volume rendering in a user study. Participants performed the task more accurately using the "overlaid edge" method but were faster using pseudo chroma-depth. Kersten et al. [8] ran a similar study comparing edge enhancement, pseudo chroma-depth, aerial perspective (fog), kinetic depth and stereo (alternating eye active glasses). Pseudo chroma-depth yielded the highest accuracy for both novices and experts taking part in the study. Both groups were faster with the aerial perspective cue and while experts qualitatively preferred aerial perspective, novices had a slight preference for pseudo chroma-depth. Abhari et al. [9] proposed a contour-enhancement volume rendering method that is found to be more effective than simple volume rendering in a user study, where each of the 10 participants needed to determine whether two vessel segments are connected. Joshi et al. [10] compared simple volume rendering with and without lighting, distance color blending, tone shading and halo-enhanced volumes in a systematic analysis of the preference of experts and found tone shading and distance color blending were prefered for conveying depth. Ritter et al. [11] introduced a surface-based non-photorealistic vessel rendering method where simulated shadows encode relative depth and hatching is used to enhance curvature of the vessels. Lawonn et al. [12] build on the depth encoded shadows idea of Ritter et al. and added depth-dependent halo contours, supporting lines that connect user-defined vessels to a plane and shape enhancing streamlines [13]. Both Ritter and Lawonn evaluate their method in a user study based on a web questionnaire where marked vessels need to be ordered in depth. Lawonn et al. compared the proposed method with pseudo chroma-depth as well as Phong shading. Lawonn et al. extended their supporting line idea to propose an automatic method to position supporting lines that are connected to a user-defined cylindrical region used to carve out other anatomical elements and reveal vessels [14]. The idea of enhancing depth with supporting structural elements is taken a step further by Kreiser et al. who fill the void between vessels with a smoothly interpolated surface encoding depth information, leaving the surface of the vessels themselves empty and thus allowing the encoding of application-specific information[15]. In a different but related type of work, Wang et al. [16] proposed different methods to

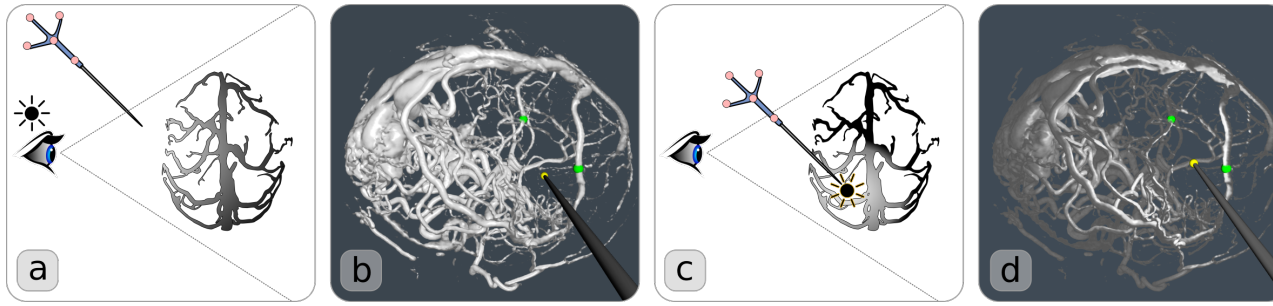


Figure 2 - Illustration of static and dynamic shading. (a) In static shading, the virtual light source is located at the same position as the virtual camera. (b) shows an example of static shading. With dynamic shading (c) the virtual light source is tracked by the tip of the pointer manipulated by the surgeon. (d) shows an example of dynamic shading.

transfer depth information from low-resolution CTA to a high-resolution 2D x-ray image. Both color-coded depth and edge enhancement with corrected occlusion were found to outperform prior methods.

Dynamic depth cues have received little attention in the field of medical image visualization. The previously mentioned work of Kersten et al. [8] has looked at the cue of Kinetic Depth where the visualized volume's rotation is automatically animated, but the performance of users with this cue wasn't improved. This may be due to the fact that motion is generated automatically instead of being the result of user interaction. The human visual perception literature has considered the interaction between human visual and motors systems (see [17] for an extensive review on this topic). Anticipatory mechanisms in the brain are thought to foresee the sensory consequences of motor actions and thus facilitate understanding of the 3D structure of the environment. For this synergy to happen, motion must be generated by the motor system of the observer.

Other groups have looked at user interaction as a way to improve perception of medical data. Bichlmeier et al. introduced the concept of the virtual mirror [18]. It relies on the motion of a tracked surgical tool to allow the surgeon to interact by rendering a virtual dentist's mirror tracked by the tool. Joshi et al. [19] proposed an interactive clipping method for volume rendering. Their method allows clipping away cubic, spherical and cylindrical regions of a volume to preserve as much contextual information as possible. The concept of a virtual window or magic lens [20] has been used by several groups in different contexts to facilitate perception. The idea is to allow users to interactively define a region of interest where the external surface of the rendered data is hidden, cropped, carved or made semi-transparent to reveal underlying structures. Such a paradigm creates powerful occlusion cues between the external surface and internal structures which helps users understand the geometry of internal structure. Mendez et al. [21] define a cuboid region around a tracked tool manipulated by the user to cut the surface of an organ and reveal vessels underneath. Bichlmeier et al. [22] create a semi-transparent region in the patient skin inside a conical channel aligned with the line of sight of a head-mounted display (HMD). Users may look through the skin anywhere by moving their head. Kalkofen et al. [23] go beyond the simple cut-out paradigm of the virtual window by allowing different compositing patterns inside the region of interest to be manipulated by the user. In Gras and al. [24], the virtual window is produced by sensing the force applied to the patient's tissues by a tracked probe manipulated by the user. Force sensing information is also used to deform virtual models of the underlying anatomy which is producing a motion cue that further enhances depth perception.

In this paper, our goal is to combine the benefits of some of the static depth enhancements methods that have been proposed for the visualization of angiographic volumes with the benefits of interactive methods. We hypothesize that the modification of the parameters of depth enhancement methods with user-generated motion will provide additional information and enable better depth perception.

3 Dynamic Rendering Methods

We propose to leverage the motion of the tracked pointer to dynamically modify the rendering parameters of three previously studied depth cues: shading, pseudo chromadepth and aerial perspective. In this section, we describe each of the three dynamic rendering methods developed, together with their non-interactive counterparts that will serve as a basis of comparison in the psychophysics experiment described in the remainder of the article.

The goal of the techniques presented here is to identify concrete mechanisms by which current intraoperative visualization techniques can be improved rather than to demonstrate some properties of the visual system of the user. For this reason, we developed all our rendering methods on top of the currently accepted standard for these types of visualization rather than trying to render each depth perception cue in isolation. We have based all our methods on a ray-casting approach where each sample is shaded using the Blinn-Phong model [25]. Shading is known to provide important surface orientation information to the human visual system [26], information which may be integrated to compute depth. However, all methods we implement should be equally influenced by this cue and its presence will allow a direct comparison of our proposed rendering methods with the ones currently used in medical practice.

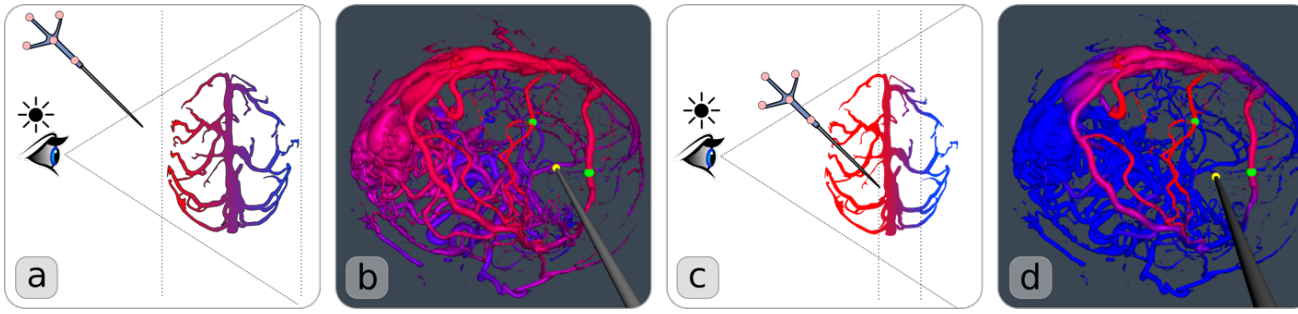


Figure 3 – Illustration of the static and dynamic chroma-depth methods. (a) In static chroma-depth, the color gradient from red to blue spans over the range of depth covered by the volume. (b) shows an example of static chroma-depth. With dynamic chroma-depth (c), the color gradient starts at a plane defined by the position of the pointer and perpendicular to the viewing axis and spans over a predefined range. (d) shows an example of dynamic chroma-depth.

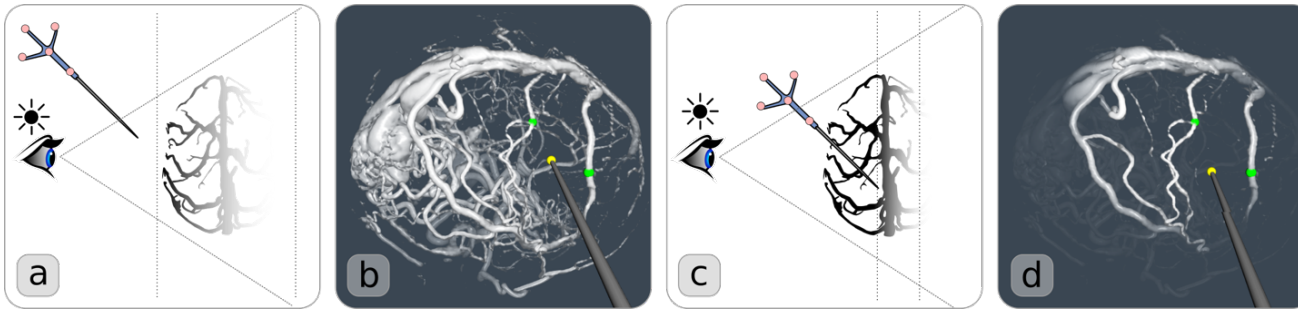


Figure 4 – Illustration of the static and dynamic aerial perspective methods. (a) In static aerial perspective, the anatomy dissolves into the background over the depth range of the entire volume. (b) shows an example of static aerial perspective. With dynamic aerial perspective (c), the dissolution gradient starts at a plane defined by the position of the pointer and perpendicular to the viewing axis and spans over a predefined distance.

3.1 Shading

The static shading condition is our baseline rendering method. In this condition, each sample of the ray-casting loop is shaded using the Blinn-Phong model. The light source that is used to compute the shading terms is located at the position of the virtual camera and the pointer's motion has no influence on the rendering. The method is illustrated in Figure 2(a) and an example is shown in Figure 2(b).

In the dynamic shading method, we attach the virtual light source to the tip of the tracked pointer manipulated by the user as illustrated in Figure 2(c). The user can then use the pointer as a digital lantern and gather information about the vascular structure based on reflection patterns on the vessels. To further help localize the pointer in relation to the vasculature in the dynamic shading method, the light intensity falls off linearly with distance from the pointer. The dynamic rendering method is illustrated in Figure 2(c) and a dynamic shading example is shown in Figure 2(d).

3.2 Chroma-Depth

Chroma-depth is a depth cue that encodes distance from the observer with color. Different color schemes are possible but the most typical one, often called pseudo chroma-depth, encodes depth using a gradient between red (close) and blue (far). This color scheme choice is motivated by the chromostereopsis effect which causes the brain to perceive red elements of an image as closer than blue ones. Elements of different colors located at the same distance from the observer produce slightly different binocular disparity because of different refraction angles in the cornea and lens of the eye, which results in a slightly different perception of depth [27]. Both [7] and [8] have found chroma-depth to be the most efficient depth cue amongst the ones they studied.

Our approach to render chroma-depth differs from both Ropinski et al.[7] and Kersten et al.[8] in that it is based on a shaded rendering model. The color of each sample of the ray-casting loop is computed according to equation [1].

$$C_{sample} = phong(p) * mix(\, C_{red}, C_{blue}, r\,) \qquad [1]$$

Where C_{sample} is the output RGB color, p is the position in the volume that is being sampled, *phong(p)* is the result of computing the Blinn-Phong shading model at position p, *mix(a,b,r)* interpolates linearly between colors a and b with ratio r and C_{red} and C_{blue} are constant RGB triplets representing pure red and blue. The choice of interpolating colors in the RGB space is deliberate. While certain color spaces such as CIE LAB provide a way to interpolate colors to obtain a linear progression of perceptual qualities like saturation and lightness, no such model exists for the perception of depth due to the chromostereopsis effect. Bailey and Clark interpolate in the HSV space [28], Wang et al. use a divergent color map where the transition between red and blue goes through an intermediate point of desaturation (white) [29], while others employ linear interpolation in RGB space [7], [8]. In the absence of a vali-

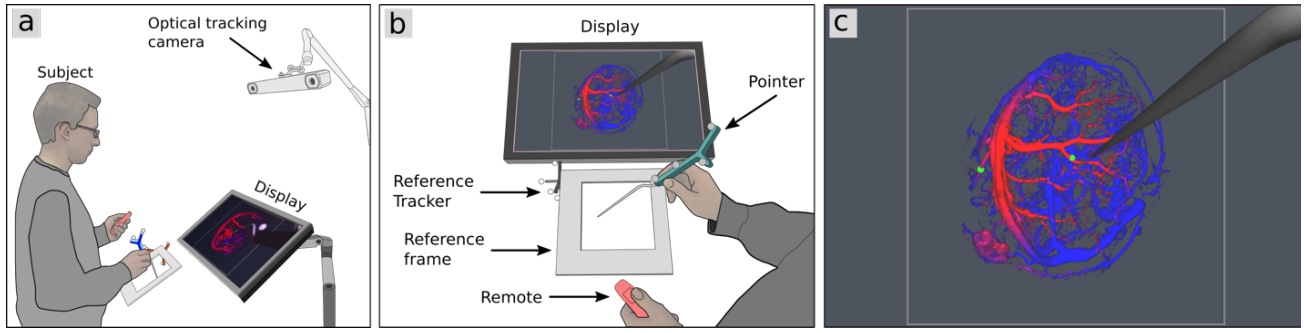


Figure 5 – Illustration of the physical setup of the psychophysics experiment used to compare rendering methods. (a) General overview of the setup showing the position of the subject relative to the display and tracking camera. (b) A closer look at the working area showing the tracked pointer, the remote control and the frame of reference that indicates the general location of the angiographic volume in physical space. (c) The image rendered on the screen using the dynamic chroma-depth method and showing a 3D representation of the tracked pointer.

dated model, we simply interpolate in RGB space.

In the static version of the chroma-depth method, the mixing ratio r is based on the depth range of the whole volume as shown in equation [2]:

$$r_{static} = \frac{d_p - d_{min}}{d_{max} - d_{min}} \qquad [2]$$

where d_p is the distance of the sampled point to the camera and (d_{min}, d_{max}) define the depth range of the whole volume. In the absence of user input, spreading the gradient over the complete depth range is the best way to maximize depth discrimination without making assumptions about the location of the target. In the dynamic version of the chroma-depth rendering method however, we are able to limit the range further as the user is free to move the pointer to any area of interest. We limit the range to a pre-defined distance and start the red-blue gradient from a depth plane located at the tip of the tracked pointer. The predefined range was determined by gathering the feedback of users during an informal pilot study. The mixing ratio is computed using equation [3]:

$$r_{dynamic} = clamp\left(\frac{d_p - d_{pointer}}{g}, 0, 1\right) \qquad [3]$$

where the clamp function is used to restrict the value between [0,1], $d_{pointer}$ is the distance between the camera and the tracked pointer tip projected on the camera axis and g is the pre-defined range of the gradient. The algorithm is illustrated in Figure 3(c) and a dynamic chroma-depth example is shown in Figure 3(d).

3.3 Aerial Perspective

Aerial perspective refers to the fading contrast and color of distant objects caused by the scattering of light on particles floating in the air (fog, smoke, humidity, etc.). In computer graphics, this cue is usually rendered by progressively mixing distant object's color with the background color and is most often referred to simply as "fog". In this paper, we render the fog effect by simply mixing the result of the Blinn-Phong shading model with the background color (equation [4]).

$$C_{sample} = mix(\, phong(p), C_{back}, r\,) \qquad [4]$$

Where C_{back} is the RGB color of the background. The mixing ratio r is obtained with the same strategy as for chroma-depth: equation [2] for the static fog rendering method and equation [3] for the dynamic equivalent. Both the static and dynamic fog rendering methods are illustrated in Figure 4.

4 Experiment

We implemented a psychophysics experiment to validate the static and dynamic cues. The goal of the experiment is to evaluate whether the dynamic rendering techniques proposed provide a better understanding of the vasculature than their static counterpart.

4.1 The physical setup

The setup for the study is illustrated in Figure 5. It is meant to simulate the OR context in which angiographic volumes are used to guide a surgery. The system tracks a surgical pointer manipulated by the user. A plexiglass frame maintained fixed above the table serves as a mockup for a craniotomy and indicates the location of the operating field for the user. The rendered angiographic volumes and a 3D representation of the pointer and frame are presented to the user on a screen located beyond the simulated operative field. The volumes presented on the screen have been registered to the plexiglass frame so that during the experiment, the position and orientation of the pointer relative to the volume is known at all time.

The design of the task is based on a depth measurement scheme called "perceptual matching" which has been used extensively in the field of human visual perception to evaluate the relative perception of depth by an observer of both real and virtual objects [30]–[37].

4.2 The task

To start each trial in the study, users must touch a black circular marker on the reference frame. This allows users to control the start of the trial and the experimenter to make sure all trials are started with the pointer in the same position. Users are then presented with an angio-

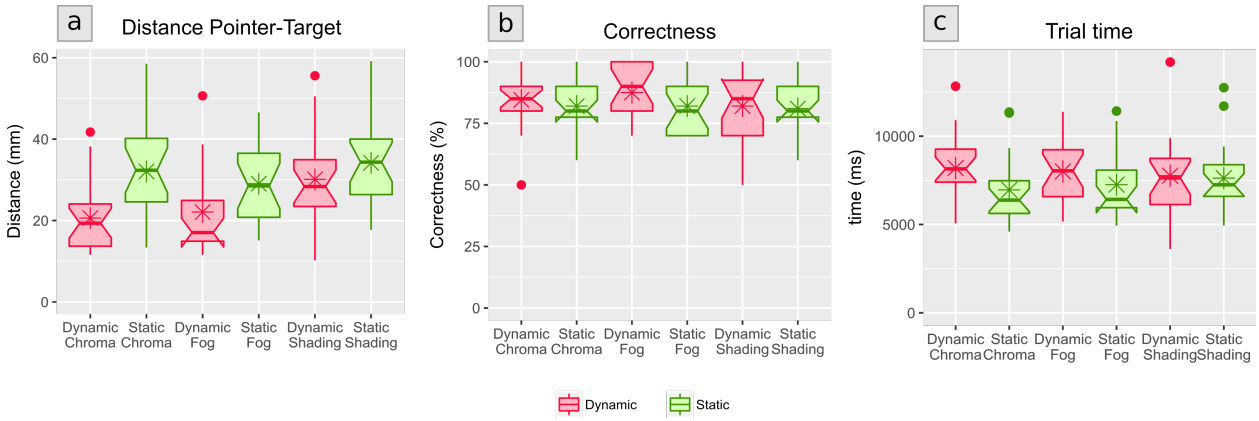


Figure 6 – Box plot (center line: median, vertical line: (min, max), box: (25th, 75th) percentile, notch: 95% confidence interval, star: mean) of metrics of the study for the different rendering methods evaluated: (a) Distance between pointer and chosen target at the end of the trial, (b) Percentage of correctly identified closest target, (c) Duration of a trial.

graphic volume where two vessels have been labelled with a green sphere (as illustrated in Figure 5c). The task consists of positioning the tip of the tracked pointer as close as possible to the closest of the two targets without touching it. This procedure is meant to force users to consider both the global structure of the vasculature when making a choice about the closest target and the local structure when approaching the chosen one. If one of the targets is touched, the trial is immediately terminated, and the data is discarded. This is done to make sure that subjects use the examined depth cues to perform the task rather than mutual occlusion of the pointer representation and target vessels. The target spheres have sizes that are relative to the diameter of their corresponding vessel segments so that this property can't be used to infer the distance of the target. When users have positioned the pointer as close to the chosen target as they can judge from the rendered vasculature, they end the trial by pressing a button on a presentation remote they hold in their other hand. To force the participants to act quickly, each trial is timed. If the remote is not pressed within 15 seconds, the system beeps and the data is discarded.

4.3 Study Details

During the experiment, the CTAs of 10 different patients are presented to the user. Each volume is presented 8 times. Each presentation is from a different point of view and with different target vessels, for a total of 80 trials for each subject. Each of the trials is rendered with one of the three different rendering methods in one of the two modes: static or dynamic shading, static or dynamic chroma-depth or static or dynamic aerial perspective. The association of rendering method with each of the 80 trials and the order in which each of the trials is presented is randomized for each subject. Before the study starts, each subject is guided through a training phase. First, the subject watches a 10 minutes PowerPoint presentation narrated by the experimenter where the task is explained and each of the rendering methods employed during the study is described. The experimenter then runs a demonstration mode of the study that simply cycles through the 6 rendering methods with the same volume but is otherwise identical to the real experiment. After, the subject is trained in the demonstration mode until she/he feels comfortable with the system before running the study.

To gather qualitative information about user's experience with the different rendering methods, we developed an online questionnaire that was filled out by every subject after running the study. The goal of the questionnaire was to collect subjective feedback on each of the rendering methods they experimented with, and to determine how effective they thought each method was to determine the relative depth of the targets and correctly approach the chosen one. We also provided a free form text box to enter general comments about the relative effectiveness of different rendering methods. A copy of the questionnaire is provided in the supplementary material published with this paper.

4.4 The implementation

The user study implementation uses the open source navigation system called IBIS Neuronav [38]. The IBIS software captures the 3D pose of the tracked pointer and reference frame with a Polaris optical tracking system (NDI, Waterloo, ON, Canada). The volumetric CTA data is rendered using the PRISM framework[39] which is integrated in IBIS Neuronav. The framework implements a GPU ray-casting method while providing the possibility to replace key parts of the ray integration method with custom code. Each of the rendering methods presented to the user during the study is implemented as a simple modification of the ray-casting method that takes into account the position of the tip of the pointer. The workflow of the study is controlled by a custom extension to the IBIS software that was created for this study. It guides the user through the study, randomizes trial order and records all the data required for the analysis of the user's performance.

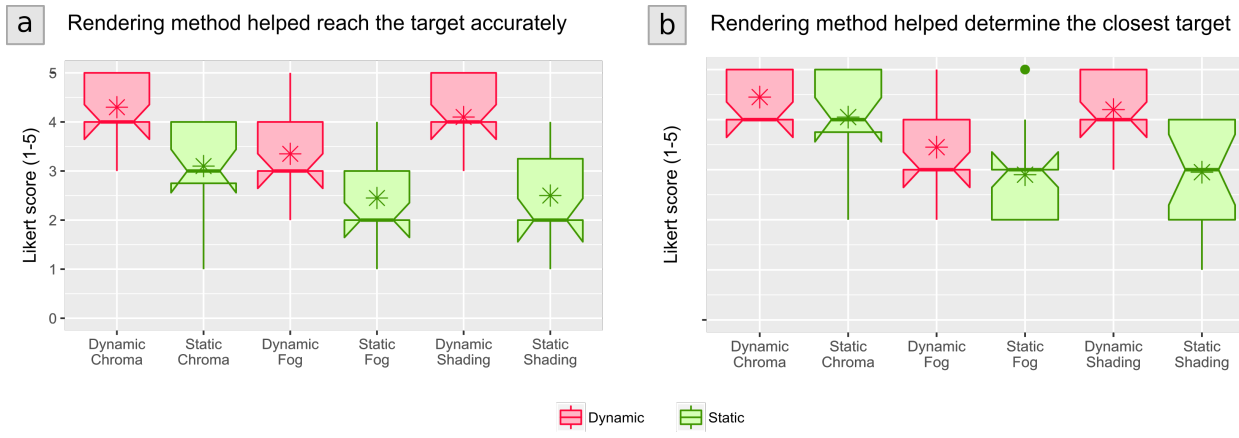


Figure 7 – Box plot (center line: median, vertical line: (min, max), box: (25th, 75th) percentile, notch: 95% confidence interval, star: mean) of the subjective evaluation (Likert scale 1-5, where 1 = strongly disagree and 5 = strongly agree) of the rendering methods. (a) The rendering method helped reach the target accurately, (b) The rendering method helped determine the closest target.

During our study, the IBIS Neuronav software ran on an Ubuntu Linux 14.04 PC equipped with an Intel Quad Core i7 processor, 32Gb of RAM and an NVidia GeForce GTX 670 graphics card with 4Gb of video memory. The PC is installed on a cart and controlled by the experimenter and drives a second monitor (shown in Figure 5a) that presents images to the subject.

5 Results

The study was run with 20 subjects (9 female, 11 male, age 23-53y). They were all professors or graduate students who are working on the analysis, processing or visualization of brain scans on a daily basis. All subjects had normal or corrected to normal vision and correct perception of color (i.e., no color-blind subject). Most subjects (18) went through the training round only once while the other two requested a second round of training before feeling comfortable enough to run the study.

5.1 Psychophysics experiment

During the experiment, we recorded:

1. The distance between the closest target and the tip of the pointer at the end of the trial (accuracy).
2. Whether the user correctly identified the closest target (correctness).
3. The time taken by the user to execute one trial (trial time).

The graphs of Figure 6 shows the mean and standard deviation of those three metrics (accuracy, correctness and trial time) for each of the rendering methods examined.

For each of the metrics, we computed a two-way repeated measures ANOVA where the first factor indicates the rendering method (Shading, Chroma-depth or Fog) and the second factor indicates whether the dynamic or the static version of the method was used (dynamics).

5.1.1 *Effect on targeting accuracy*

For the interaction between dynamics and rendering method, Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2(2) = 9.47$, p = .009). After applying a Greenhouse-Geisser correction, there was no statistically significant two-way interaction between the factors ($F(2, 38) = 3.26$, p = .069). However, the main effect of dynamics showed a statistically significant difference between dynamic and static conditions ($F(1, 19) = 18.71$, $p < .001$). When using a dynamic rendering method, participants were able to get on average 7.5 mm closer to the target than with static conditions (95% Confidence interval (3.85,11.07)). For the main effect of the rendering method, the ANOVA showed a significant difference of accuracy between rendering methods ($F(2, 38) = 9.22$, $p < .001$). Post-hoc pairwise t-tests (Bonferoni corrected for multiple comparisons) revealed a significant difference between chroma-depth and shading and between fog and shading, but not between chroma-depth and fog. Statistically significant results are summarized in Table 1.

Methods	Diff (mm)	95% CI	p
Chroma/Shading	5.8	(1.07,10.48)	0.014
Fog/Shading	6.6	(1.95,11.19)	0.004

Table 1 – Statistically significant pairwise comparison of mean accuracy between rendering methods. We show mean difference between methods in the left column (Diff), 95% confidence intervals (95% CI) and Bonferoni-corrected p-value of the t-test.

5.1.2 *Effect on correctness*

Despite the higher mean selection correctness obtained for dynamic chroma-depth and fog, The ANOVA didn't reveal a significant interaction between dynamics and rendering method nor a significant main effect for one of the factors.

5.1.3 *Effect on time*

For the interaction between dynamics and rendering method, there was a statistically significant two-way interaction between factors ($F(2, 38) = 7.76$, $p = .001$). There was a simple main effect of dynamics on chroma-depth and fog, but not on shading. Detailed results of the simple main effect are shown in Table 2.

Methods	Diff (ms)	95% CI	F(1,19)	p
Chroma	1284	(712, 1856)	22.09	< 0.001
Fog	748	(274, 1223)	10.88	0.004

Table 2 – Statistically significant simple main effect of dynamics on rendering method. We show the time difference between means of static and dynamic conditions (Diff), the 95% confidence interval (95% CI), the F statistics and the corresponding p-value.

There was no significant difference between dynamic rendering methods ($F(2, 38) = 1.01$, $p = .374$), but there was a significant difference between static rendering methods ($F(2, 40) = 4.97$, $p = .012$). A post-hoc pairwise t-test reveals there is only a significant difference between static shading and static chroma-depth (shading was faster by 665 *ms*, 95% confidence interval (178, 1153), p = 0.006).

5.2 Questionnaire

The core of the form filled by participants consisted of questions where users were required to subjectively rate, on a Likert scale (1-5), whether each of the rendering methods:

1. Helped reach target more accurately (accuracy)
2. Helped determine the closest of the two targets (correctness)

Figure 7 presents the means and standard deviations of Likert scores obtained for each of the rendering methods and for both aspects.

We computed a 2-way repeated measures ANOVA for both aspects (correctness and accuracy) using again the rendering method as the first factor and dynamics as the second factor.

5.2.1 Effect on accuracy

There was a statistically significant interaction in user perception between dynamics and rendering method ($F(2, 38) = 4.60$, $p = .016$). Dynamics had a significant simple main effect for all three rendering methods, with a systematically higher score for dynamic over static as shown Table 3.

Method	Dynamic – Static	95% CI	F(1,19)	p
Chroma	1.2	(.71, 1.69)	25.81	< .001
Fog	.9	(.42, 1.38)	15.55	.001
Shading	1.6	(1.13, 2.07)	51.75	< .001

Table 3 – Simple main effect of dynamics over rendering method. We show the mean Likert score difference between dynamic and static (Dynamic – Static), the 95% confidence interval (95% CI), the F statistics and the p-value for each method.

There was a statistically significant difference in user perception between both static rendering methods ($F(2, 38) = 9.85$, $p < .001$) and dynamic rendering methods ($F(2, 38) = 10.23$, $p < .001$). Table 4 shows the significant pairwise comparisons (t-test with Bonferoni correction for multiple comparisons) between rendering methods for both static and dynamic conditions.

	Rendering Method	Diff	95% CI	p
Static	Chroma/Fog	.65	(.31, .10)	< .001
	Chroma/Shading	.60	(.16, 1.04)	.006
Dynamic	Chroma/Fog	.95	(.40, 1.50)	.001
	Shading/Fog	.75	(.09, 1.41)	.022

Table 4 – Significant pairwise comparison between all static rendering methods and all dynamic rendering methods. We show the mean different of the Likert Score (Diff), the 95% confidence interval (95% CI) and the p-value of the t-test (p).

5.2.2 Effect on correctness

There was a statistically significant interaction in user perception between dynamics and rendering method ($F(2, 38) = 7.66$, $p = .002$). Dynamics had a significant simple main effect for all three rendering methods, with a systematically higher score for dynamic over static as shown Table 5.

Method	Dynamic – Static	95% CI	F(1,19)	P
Chroma	.40	(.05, .75)	5.63	.028
Fog	.55	(.20, .91)	10.50	.004
Shading	1.25	(.80, 1.70)	33.45	< .001

Table 5 – Simple main effect of dynamics over rendering method. We show the mean score difference between static and dynamic (Dynamic – Static), the 95% confidence interval (95% CI), the F statistics and the p-value for each method.

There was a statistically significant difference in user perception between both static rendering methods ($F(2, 38) = 9.85$, $p < .001$) and dynamic rendering methods ($F(2, 38) = 10.23$, $p < .001$, after applying a Greenhouse–Geisser correction for violating sphericity). Table 6 shows the significant pairwise comparisons (t-test with Bonferoni correction for multiple comparisons) between rendering methods for both static and dynamic conditions.

	Rendering Method	Diff	95% CI	p
Static	Chroma/Fog	1.15	(.43, 1.87)	< .001
	Chroma/Shading	1.10	(.50, 1.70)	< .001
Dynamic	Chroma/Fog	1.00	(.31, 1.69)	.003
	Shading/Fog	.75	(.02, 1.49)	.044

Table 6 – Significant pairwise comparison between all static rendering methods and all dynamic rendering methods. We show the mean different of the Likert Score (Diff), the 95% confidence interval (95% CI) and the p-value of the t-test (p).

6 Discussion

The results of the psychophysics experiment suggest that the proposed dynamic rendering methods provide an improved appreciation of the local structures of the vasculature as users were able to get significantly closer to a

target vessel without touching it, compared to the static methods. This could be explained by the fact that dynamic methods provide information that is inherently more local than their static counterpart. For example, the dynamic chroma-depth method implements a depth gradient over a much shorter range than its static equivalent, which enables a finer representation of local information given the limited range of hues that can be represented by the medium and perceived by the human visual system. Similar observations can be made for aerial perspective and shading.

Although dynamic methods provide a better appreciation of the local structure of vessels, our experiments did not find this to be true for the appreciation of the global structure as there was no significant difference between static and dynamic methods in terms of determining the closest of two targets in each trial of the study (i.e. correctness). Given the local aspect of the information depicted by the dynamic methods discussed above, this is not a surprising result.

Generally speaking, dynamic rendering methods have a negative impact on the time required to complete a trial. This could be explained by the fact that with dynamic methods, a longer exploration of the vasculature provides additional information about the structure while with static methods, all the information is available from the start. This assumption was confirmed by comments from several participants of the study who reported having a tendency to spend more time exploring with dynamic methods. This assumption is also consistent with the simple main effect of dynamics on different rendering methods. There was a significant trial time difference between static and dynamic methods for chroma-depth and fog while there was none for shading. We can hypothesize that shading provided less information that is relevant to the completion of the task and hence led the participants to spend less time exploring the volume before completing a trial, which is consistent with the results for accuracy, where chroma-depth and fog outperformed shading.

The results of the questionnaire show that participants estimate that their performance, both for accuracy and correctness, was improved when using dynamic rendering methods. Furthermore, their preferred method in both cases is chroma-depth. These results are consistent with the performance in the psychophysics study, to one difference: participants appear overly confident about the benefit of dynamic methods to determine the closest target while their performance in the task wasn't significantly better.

The most surprizing discrepancy between objective metrics and the results of the questionnaire is with the shading and fog dynamic methods. While participants performed significantly worst with dynamic shading on the objective measures, they didn't judge their performance significantly worse than with chroma-depth (for both correctness and accuracy) and estimated shading benefited the most from interaction (mean Likert score difference between static and dynamic, accuracy: 1.6 against 1.2 for chroma-depth and 0.9 for fog, correctness: 1.25 against 0.4 for chroma-depth and 0.55 for fog). We speculate that participants may have been positively biased by the natural aspect of the visuo-motor scheme involved in dynamic shading. Everybody has experimented with a similar gesture when, for example, examining the interior of an object, machine or vehicle using a flashlight. Another aspect that may explain the results for dynamic shading is the fact that it provides a richer type of information, inherently three-dimensional, compared to chroma-depth and fog, which provide information only along the z-axis. As depth information is sufficient for the execution of the task involved in our psychophysics experiment, it is not surprising that the methods from which this information is obtained most directly (chroma-depth and fog) outperforms shading. In a clinical context however, the additional information provided by dynamic shading may be important. Further investigation using more complex, task-relevant metrics are needed in order to verify this hypothesis.

In the case of dynamic fog, we find the opposite discrepancy between objective and subjective metrics. Dynamic fog obtains a significantly lower Likert score for both correctness and accuracy despite the fact that participants performed better in terms of accuracy. The results for fog could be explained by the fact that objectively, this cue contains the same information as chroma-depth, which results in a similar performance on an objective task, but is less visually striking, which may subjectively bias users against it.

7 Conclusion and Future Work

The goal of this paper was to investigate the benefits of 3D rendering techniques that rely on the manipulation of a positioning device available in operating rooms to improve the perception of angiographic volumes used to guide a neurovascular surgical intervention. We have presented three different interactive rendering techniques based on this principle and compared the performance of users with each of those techniques to their performance with the comparable non-interactive counterpart. The quantitative results of the study have shown the potential of some interactive techniques based on the cues of chroma-depth and aerial perspective (fog) to significantly improve the ability of users to navigate the local structure of the vasculature while not being detrimental to the appreciation of the global structure. Dynamic methods did have a negative impact on the speed of execution. However, delays on the order of a few seconds are negligible during a surgical operation and are an acceptable price to pay if it results in an improvement of accuracy.

Further work is necessary to refine the interactive techniques we developed. In particular, it is important to measure the effect of the chosen depth range over which we compute the color mixing gradients for dynamic

chroma-depth and aerial perspective. It should also be determined whether the fact that vessels outside the predefined range are represented with a single depth level has a negative impact on the user's perception.

Another important aspect that requires attention is the perceptual linearity of color encoding schemes used in dynamic methods. The accuracy of depth judgement is highly dependent on the use of a system where equal variations of physical depth produce equal variations in perceived depth. As mentioned earlier, no such perceptually uniform system has been defined for the chroma-depth effect. The depth perception in chroma-depth has been shown to result from a combination of factors including chromatic aberration[27] and luminance differences between red and blue [40], which makes it difficult to rely on any existing color system for interpolation. To account for these factors, it is necessary to conduct thorough psychophysics study that will assess the relative perceived depth of different colors.

The discrepancy between subjective appreciation and objective measures of the different dynamic techniques could be due to the limitations of the metrics we have used to evaluate the method. Some interactive methods such as dynamic shading may provide spatial information to the user that is not critical to the execution of the task we proposed, but that may help to better appreciate the vascular structure and thus be beneficial for a surgeon in a real operation. New evaluation metrics are needed for interactive rendering methods that would seek to evaluate the perceptibility of more abstract domain-specific information and better evaluate the usefulness of the method.

In the future, we will adapt the techniques developed in this paper to the field of surgical augmented reality (AR). This technology can greatly simplify the mapping of brain scans with the patient on an operating room table but is associated with a diminished perception of depth. The methods developed in this paper are compatible with the constraints of surgical AR and could potentially help mitigate the depth perception problem associated with this technology.

We believe that the work presented in this paper is relevant beyond the field of medical image visualization. It is now well established that preparation and execution of motor actions have an important impact on the ability for an observer to visually perceive three-dimensional structures[17]. Until recently, there was a lack of affordable devices and displays to take full advantage of this phenomenon when designing visualization programs. In the past couple of years however, we have seen a multiplication of the number of affordable motion sensing devices that enable more natural interaction with computers (e.g. motion sensors, cameras, hand and eye tracking systems). Simultaneously, the variety of commonly used displays for which the traditional mouse and keyboard paradigm is impractical has also greatly increased (augmented and virtual reality headsets, immersive projection screens, mobile devices). A sensible display and interaction paradigm still has to be defined for several combinations of those new tools. Our work demonstrates that user-generated motion can be leveraged to encode important local information in 3D rendered images. This encoding pattern could find applications in various fields such as computer gaming, design, architecture and data visualization. It would be particularly useful in cases where virtual or augmented head-mounted displays are combined with embedded hand trackers or simple 3D tracked remote controls. This is a promising direction for future work.

8 REFERENCES

[1] R. L. Gregory, *Eye and brain, the psychology of seeing. 3rd edition*. London: Weidenfeld and Nicolson, 1977.

[2] D. Onceanu and A. J. Stewart, "Direct Surgeon Control of the Computer in the Operating Room," in *Medical Image Computing and Computer-Assisted Intervention*, 2011, pp. 121–128.

[3] C. S. Harris, "Perceptual adaptation to inverted, reversed, and displaced vision.," *Psychol. Rev.*, vol. 72, no. 6, pp. 419–444, 1965.

[4] S. Abeele and O. Bock, "Mechanisms for sensorimotor adaptation to rotated visual input," *Exp. Brain Res.*, vol. 139, no. 2, pp. 248–253, Jul. 2001.

[5] L. Masia, M. Casadio, G. Sandini, and P. Morasso, "Eye-hand coordination during dynamic visuomotor rotations.," *PLoS One*, vol. 4, no. 9, p. e7004, Sep. 2009.

[6] B. Preim, A. Baer, D. Cunningham, T. A. Isenberg, and T. Ropinski, "A Survey of Perceptually Motivated 3D Visualization of Medical Image Data," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 501–525, Jun. 2016.

[7] T. Ropinski, F. Steinicke, and K. Hinrichs, "Visually supporting depth perception in angiography imaging," in *Smart Graphics*, 2006, pp. 93–104.

[8] M. Kersten-Oertel, S. J. Chen, and D. L. Collins, "An Evaluation of Depth Enhancing Perceptual Cues for Vascular Volume Visualization in Neurosurgery," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 3, pp. 391–403, 2014.

[9] K. Abhari, J. S. H. Baxter, R. A. Eagleson, T. Peters, and S. de Ribaupierre, "Perceptual enhancement of arteriovenous malformation in MRI angiography displays," in *SPIE Medical Imaging*, 2012.

[10] A. Joshi *et al.*, "Effective visualization of complex vascular structures using a non-parametric vessel detection method," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1603–1610, Nov. 2008.

[11] F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, and H.-O. Peitgen, "Real-time illustration of vascular structures.," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 877–84, 2006.

[12] K. Lawonn, M. Luz, B. Preim, and C. Hansen, "Illustrative Visualization of Vascular Models for Static 2D Representations," in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, 2015, pp. 399–406.

[13] K. Lawonn, T. Moench, and B. Preim, "Streamlines for Illustrative Real-Time Rendering," *Comput. Graph. Forum*, vol. 32, no. 3pt3, pp. 321–330, Jun. 2013.

[14] K. Lawonn, M. Luz, and C. Hansen, "Improving spatial perception of vascular models using supporting anchors and illustrative visualization," *Comput. Graph.*, vol. 63, pp. 37–49, Apr. 2017.

[15] J. Kreiser, P. Hermosilla, and T. Ropinski, "Void Space Surfaces to Convey Depth in Vessel Visualizations," *arXiv*, 2018.

[16] J. Wang, M. Kreiser, L. Wang, N. Navab, and P. Fallavollita, "Augmented Depth Perception Visualization in 2D/3D Image Fusion," *Comput. Med. Imaging Graph.*, vol. 38, no. 8, pp. 744–752, Jul. 2014.

[17] M. Wexler and J. J. A. van Boxtel, "Depth perception by the active observer.," *Trends Cogn. Sci.*, vol. 9, no. 9, pp. 431–8, Sep. 2005.

[18] C. Bichlmeier, S. M. Heining, M. Feuerstein, and N. Navab, "The virtual mirror: a new interaction paradigm for augmented reality environments.," *IEEE Trans. Med. Imaging*, vol. 28, no. 9, pp. 1498–510, Sep. 2009.

[19] A. Joshi, D. Scheinost, K. P. Vives, D. D. Spencer, L. H. Staib, and X. Papademetris, "Novel interaction techniques for neurosurgical planning and stereotactic navigation.," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1587–94, 2008.

[20] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose, "Toolglass and magic lenses," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93*, 1993, pp. 73–80.

[21] E. Mendez, D. Kalkofen, and D. Schmalstieg, "Interactive context-driven visualization tools for augmented reality," in *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2006, pp. 209–218.

[22] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual Anatomic Mimesis Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality," in *IEEE/ACM International Symposium on Mixed and Augmented Reality*, 2007, pp. 129–138.

[23] D. Kalkofen, E. Mendez, and D. Schmalstieg, "Comprehensible Visualization for Augmented Reality," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 2, pp. 193–204, Mar. 2009.

[24] G. Gras, H. J. Marcus, C. J. Payne, and P. Pratt, "Visual Force Feedback for Hand-Held Microsurgical Instruments," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, vol. 9349, pp. 480–487.

[25] J. F. Blinn, "Models of light reflection for computer synthesized pictures," in *Proceedings of the 4th annual conference on Computer graphics and interactive techniques - SIGGRAPH '77*, 1977, vol. 11, no. 2, pp. 192–198.

[26] J. J. Koenderink, A. J. Van Doorn, and A. M. L. Kappers, "Surface perception in pictures," *Percept. Psychophys.*, vol. 52, no. 5, pp. 487–496, Sep. 1992.

[27] R. C. Allen and M. L. Rubin, "Chromostereopsis," *Surv. Ophthalmol.*, vol. 26, no. 1, pp. 22–27, Jul. 1981.

[28] M. Bailey and D. Clark, "Using ChromaDepth to obtain inexpensive single-image stereovision for scientific visualization," *J. Graph. Tools*, vol. 3, no. 3, pp. 1–9, 1998.

[29] X. Wang, C. Schulte Zu Berge, S. Demirci, P. Fallavollita, and N. Navab, "Improved interventional X-ray appearance," in *ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality*, 2014, pp. 237–242.

[30] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis, "Depth judgment measures and occluding surfaces in near-field augmented reality," *Proc. 7th Symp. Appl. Percept. Graph. Vis. - APGV '10*, vol. 1, no. 212, p. 149, 2010.

[31] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci, "Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays," *Presence Teleoperators Virtual Environ.*, vol. 11, no. 6, pp. 610–625, 2002.

[32] J. P. Rolland, W. Gibson, and D. Ariely, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence Teleoperators Virtual Environ.*, vol. 4, no. 1, pp. 24–49, 1995.

[33] J. W. McCandless, S. R. Ellis, and B. D. Adelstein, "Localization of a time-delayed, monocular virtual object superimposed on a real environment," *Presence Teleoperators Virtual Environ.*, vol. 9, no. 1, pp. 15–24, 2000.

[34] S. R. Ellis and B. M. Menges, "Localization of Virtual Objects in the Near Visual Field," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 40, no. 3, pp. 415–431, 1998.

[35] G. P. Bingham, A. Bradley, M. Bailey, and R. Vinner, "Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 27, no. 6, pp. 1314–1334, 2001.

[36] C. Prablanc, J. F. Echallier, E. Komilis, and M. Jeannerod, "Optimal response of eye and hand motor systems in pointing at a visual target - I. Spatio-temporal characteristics of eye and hand movements and their relationships when varying the amount of visual information," *Biol. Cybern.*, vol. 35, no. 2, pp. 113–124, 1979.

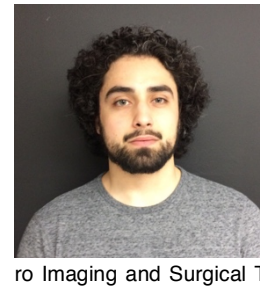
[37] P. J. Edwards *et al.*, "Design and evaluation of a system for microscope-assisted guided interventions (MAGI).," *IEEE Trans. Med. Imaging*, vol. 19, no. 11, pp. 1082–93, Nov. 2000.

[38] S. Drouin *et al.*, "IBIS: an OR ready open-source platform for image-guided neurosurgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 12, no. 3, pp. 363–378, 2016.

[39] S. Drouin and D. L. Collins, "PRISM: An open source framework for the interactive design of GPU volume rendering shaders," *PLoS One*, vol. 13, no. 3, p. e0193636, Mar. 2018.

[40] P. Thompson, K. May, and R. Stone, "Chromostereopsis: a multicomponent depth effect?," *Displays*, vol. 14, no. 4, pp. 227–234, 1993.

Simon Drouin is a PhD candidate in the Neuroimaging and Surgical technologies laboratory of the McConnell Brain Imaging Centre at the Montreal Neurological Institute. He received a bachelor's degree in computer engineering from École Polytechnique de Montréal in 2000 and completed a MSc in Computer Science at McGill University in 2008, where he focused on applications of computer vision. Between different study periods, he worked as a software developer at Ubisoft Entertainment inc., Janro Imaging Labs inc., the National Film Board of Canada (NFB) and the Montreal Neurological Institute. His research focusses on human-computer interaction involving real-time graphics rendering.



Daniel A. DiGiovanni received his BA (honours) degree in Linguistics and Religious Studies from Carleton University (Ottawa) in 2011. After working as an English Teacher overseas, he completed a M.Cog.Sci. in Cognitive Science at Carleton University in 2016 and began his PhD in Neuroscience at McGill University (Montreal) in the same year. He is a member of the Neuro Imaging and Surgical Technologies Lab (NIST) at the McConnell Brain Institute at the Neuro (MNI). His research is focused on the development of functional imaging techniques for pre-surgical planning.

Marta Kersten-Oertel received the BSc (Honours) degree in Computer Science and the BA degree in Art History from Queen's University (Kingston) in 2002. In 2005 she completed her MSc in Computer Science at Queen's University. After working as a research assistant at the GRaphisch-Interaktive Systeme at the University of Tübingen (Germany), in 2015 she received the PhD degree in Biomedical Engineering at McGill University (Montreal). Since 2016, she is an Assistant Professor of Computer Science

and Software Engineering at Concordia University and head of the Applied Perception lab. Her research is focused on developing and evaluating new visualization techniques, and display and interaction methods in the context of image-guided surgery.

D. Louis Collins is a scientist in the McConnell Brain Imaging Centre at the Montreal Neurological Institute (MNI) and a Professor in the Departments of Neurology and Neurosurgery, and Biomedical Engineering at McGill University (Montreal). He obtained the PhD degree in Biomedical Engineering from McGill University in 1994. He conducted his postdoctoral research at the Université de Rennes, France in the Laboratoire Signaux et Images en Médecine. He returned to Montreal in 1996 as a Research Associate at the MNI, and he became a faculty member in 1999. His research interests include automated anatomical segmentation and atlasing in a neuroanatomical context for the analysis of normal aging, the effect of neurological diseases on the brain as well as modelling and visualization for planning and guidance of neurosurgery.