Effects of Illumination, Texture, and Motion on Task Performance in Streamtube Visualization of Diffusion Tensor MRI

Devon Penney, Jian Chen, Member, IEEE, and
David H. Laidlaw, Senior Member, IEEE

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-08-10
December 2008

Effects of Illumination, Texture, and Motion on Task Performance in Streamtube Visualization of Diffusion Tensor MRI

Devon Penney, Jian Chen, *Member, IEEE*, and David H. Laidlaw, *Senior Member, IEEE*

Abstract— We present results from a user study comparing user task performance on streamtube visualizations generated from diffusion tensor magnetic resonance imaging (DTI) datasets. The independent variables include illumination model (global illumination or OpenGL local illumination), texture (with and without), motion (with and without), and tasks. The three spatial analysis tasks are: 1) a depth judgment task: determining which of two marked tubes is closer to the user's viewpoint, 2) a visual tracing task: marking the endpoint of a tube, 3) a contact judgment task: analyzing tube-sphere penetration. Our results indicate that global illumination did not improve task completion time for the depth judgment task or the tracing task, contrary to the results of a previous study. Global illumination improved accuracy of participants' answers over local OpenGL-style rendering for the contact judgment task only. Motion contributes to spatial understanding for the visual tracing and contact judgment tasks but at the cost of longer task completion time. A high-frequency texture pattern with more than one directional component led to longer task completion time and higher error rates. These results suggest that black-box visualization techniques are unlikely to be better than task-specific techniques tailored to the data being presented. Lighting design noticeably alters performance, indicating that designers need to pay specific attention to such issues.

Index Terms—Flow Visualization, Human Factors, Multivariate Visualization, Performance Evaluation, Three-Dimensional Graphics and Realism.

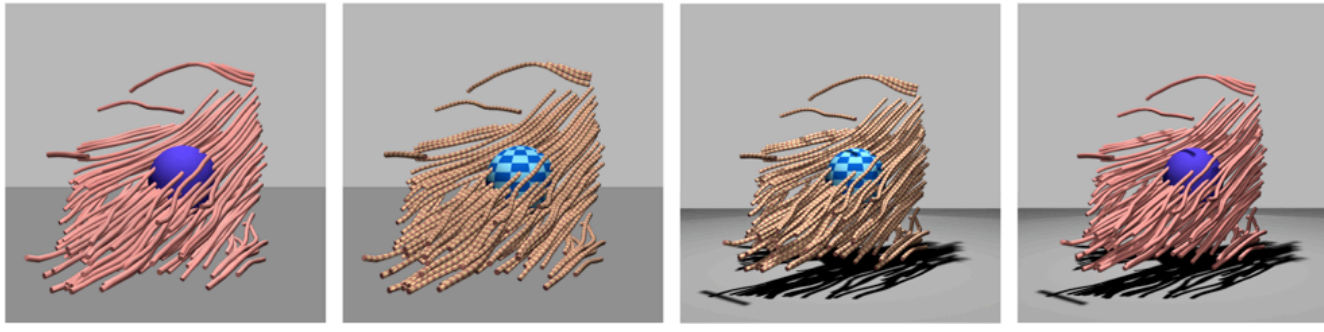

Fig. 1. Four rendering styles studied here, left to right: (1) OpenGL rendering, with intrinsic shadows, no extrinsic shadow or texture; (2) texture added to (1); (3) global illumination rendering, with texture, and (4) global illumination rendering, without texture. The dataset is a subvolume of a diffusion tensor magnetic resonance imaging (DTI) dataset. The sphere is a proxy for a tumor in one of the tasks.

1 Introduction

THREE-DIMENSIONAL (3D) streamtubes are popular for visualizing diffusion tensor magnetic resonance imaging (DTI) to provide valuable tractography information about the human brain [33]. However, dense streamtubes suffer from visual cluttering, which impedes identification of specific tracts of interest and slows down user interaction [1, 33]. Accordingly, efforts have been made to improve spatial structure illumination and rendering by adding specular reflections [18] and increasing visual realism [25]. The illumination methods can enhance shape and depth perception and generate visually appealing interreflections (light reflecting diffusely from one surface onto another) and shadows [3, 35]. While these cues alone can enhance distance judgment and shape understanding [17, 22, 26], it is unclear how such illumination models affect task performance. Ex;loring this question will help us design algorithms that integrate only important cues in the rendering algorithms, thus

• *D. Penney is with the DreamWorks Animation, 1000 Flower St. Glendale, CA 91201, Email: dmp@cs.brown.edu.*
• *J. Chen and D. H. Laidlaw are with the Computer Science Department, Brown University, Providence, RI 02912. E-mail: {jchen, dhl}@cs.brown.edu.*

Manuscript received December 2008.

reducing computational cost while maximizing user performance.

The goal of this work is thus to explore the effect of illumination models on DTI tasks. We compare two illumination models (Fig. 1): a global illumination (GI) and an OpenGL-based local illumination model (GL). GI simulates the complex behaviors of light around the model being viewed, and thus can create realistic renderings [3]. GL is a baseline method that does not allow definition of light sources, although they can be simulated using texture mapping [18]. We hypothesize that GI can improve task completion time and reduce error rate compared to GL because it provides more depth and shape cues.

As a secondary focus, we are interested in understanding the effect of perception cues on DTI tasks. This is because, in real-world use, rendering does not function alone but is combined with perception cues. Here we use motion and texture as perception cues because these two are believed to convey information efficiently in a measured dataset [34]. By combining these factors, results from our study will provide a full picture of how real-world applications work given various cueing effects. Results will also demonstrate the relative contributions of these visual design parameters.

Our experiment has three spatial analysis tasks: a depth judgment task, a visual tracing task, and a contact judgment task, each corresponding to such real-world DTI tasks as analyzing a tube in a DTI bundle and tumor diagnosis. Independent variables include illumination model (GI or GL), motion (with or without), texture (with or without), and task. Dependent variables of performance metrics include task completion time, accuracy, and subjective responses.

Our results indicate that the GI model led to longer task completion time compared to GL, contrary to our hypothesis. GI had no significant impact on accuracy for the depth judgment task, contrary to the findings of a previous study [35], and improved accuracy only in the contact judgment task. Our results also reveal that motion improves accuracy and lowers the error rate, but at the cost of longer task completion time. Participants' reported use of the three measured visual cues (shadow, texture, and motion) varies with the task at hand. Interestingly, the reported uses of other cues, such as context, shading, size, and color, are fairly consistent. Participants reported that GI increased the perceived quality of a visual scene, suggesing that "looking good" (perceived) and "being (measurably) good" (functional) are different. Visual designers must differentiate between the *perceived* and *functional* value of visualization. We providepractical design guidelines based on these experimental results.

The contributions of our work include the following. We offer a deeper understanding of how illumination models, motion, and texture affect task completion time, error rates, and accuracy in interactive 3D environments. We demonstrate empirically that rendering and perceptual parameters are not the same; the perceived value and functional value of visualization are not equivalent. Our study also contributes to a deeper understanding of visual cues, which could impact other kinds of dense tube rendering such as general 3D vector field visualization and other tasks that involve examining curves in 3D space.

2 Background and Related Work

2.1 DTI Visualization

DTI is an imaging method that measures the directional dependence of motion of water molecules in tissues. Because DTI is the only non-invasive technique that shows the internal structure of white matter, it has been widely used in research on brain development, tumor detection, and multiple sclerosis, among other areas. A common way to visualize DTI data is to reconstruct the individual fibers from the tensor information by tracking streamlines [38]. These streamlines are sampled on seed points to show fiber structures, sometimes called fiber bundles. However, the streamline visualization can easily become cluttered because of the complexity of the brain's white matter, and because users often seed at many positions to avoid missing important information. As a result, it can be difficult to get insight into dense datasets.

There are at least two ways to improve visualization of these dense datasets: to declutter by clustering fiber bundles, or to convey spatial structures via perceptual principles. Using clustering lets us reduce the enormous quantity of individual fibers to a limited number of logical fibers that still convey anatomical meaning and are visually understandable. By improving spatial structure, a visualization system often constructs an illumination model to show shadows of the fibers or allows users to query their data interactively [1, 20, 39]. While Moberts et al. [21] compare different clustering algorithms, we study structure from the perspective of perception.

2.2 Illumination model and relevant cues

Two illumination techniques are common for DTI rendering. GL involves per-vertex lighting calculations and interpolation to fill each polygon within the mesh. Ambient lighting effects are indicated by a global color shift. A known drawback of GL is that it does not allow for physically based light sources or reflectance models and thus, does not support shadow and shading of the tubes, though texture-mapping algorithms can be used to show lighting of lines [18]. The other popular rendering technique is GI, which involves simulating the behavior of light throughout the scene in order to increase visual realism [15]. For example, the photon-mapping algorithm can produce interreflections and soft shadows [3].

Two types of shadows, extrinsic and intrinsic [16], can be generated via illumination models. Intrinsic shadows, also called shading, are the shadows an object makes on itself (Fig. 1). It has long been used by artists [7] and understood by psychologists to provide information on the convex or concave shapes of objects and the direction of illumination in a scene [16]. When combined with glyph texture, intrinsic shadows also improve inter-surface distance judgment [31]. Extrinsic shadows (Figs. 1, 3, and 4), on the other hand, are those cast on one object by another,

and provide particularly salient cues to relative position, such as depth, distance, and orientations of objects [29, 30].

It is generally agreed that shadows also increase reported task accuracy when motion is absent [36] and improve depth perception as well [30]. For example, Thompson et al. analyzed how shadow and interreflection affect depth perception [30]. Their hypothesis was that both interreflections and shadows, whether fine or crude approximations and whether alone or in combination, "glue" objects to the surfaces they touch and hence improve spatial structure. They tested their participants' abilities to determine whether a box was touching a ground plane and found that both shadow and interreflection cues and even one alone significantly improved depth perception. In our study, we test illumination models that can produce these subtle visual cues, with more complex geometry from DTI.

Even though algorithms for calculating shadows exist, they are rarely used in rendering systems that involve highly complex geometry. It could be that shadow is so subtle that it sometimes does not provide strong 3D depth cue. Therefore, the shape of individual structure is difficult to perceive, especially when the scene is evenly illuminated [8]. In this work, we study whether people make use of these rich cues and whether they could help perform visualization tasks more efficiently and effectively.

Unlike OpenGL rendering, GI supports interreflection. The light that ultimately reaches the eye and affects the image bounces more than once, so that surfaces not directly facing the light can be illuminated by other surfaces. Banks introduced the idea of maximizing the reflected light over the perimeter of an infinitesimally thin cylinder, treating diffuse and specular reflection separately [2]. Hardware solutions to maximize reflection illumination modes have been applied to diffusion tensor imaging by Wenger et al. [37]. Obert et al. addressed the aesthetic drawbacks of the inflexibility of GI by creating a relighting tool for CG filmmaking [24]. While many studies have focused on algorithm design, our goal here is to understand the impact of image quality on task performance. Results from our work can guide visual designers towards important design features.

Anecdotal evidence suggests that cues generated from GI allow more accurate discrimination of shapes by revealing spatial structure and orientation among surfaces [5, 9, 10, 28, 35]. For example, Weigle and Banks showed that GI was beneficial in visualizing highly dense tubes for detecting the boundary shapes and reduced error rates in depth judgment tasks, especially when a perspective view is used [35]. Our work is in a similar spirit in comparing illumination models, but uses experimental settings that are closer to real-world uses of DTI when motion and texture are presented.

Besides rendering with visual realism, nonphotorealistic rendering (NPR) has also been used to enhance features. For example, Ritter et al. design a "shadowlike" hatching algorithm to improve depth cues for vascular tree visualization [27]. Wenger et al. use threads and halos to clarify depth relationships within dense DTI rendering [37]. Ebert et al. present numerous NPR techniques, such as silhouette enhancement, fading, sketching lines, opaque, intensity depth-cueing, and toning, that can potentially be applied to DTI imaging to improve scene understanding [8].

2.3 Texture, motion, and context

Texture is also an important factor in structural understanding of DTI imaging. It has been suggested that "lines that follow the form" convey shape effectively [12]. Here we use results from Jianu et al. to construct diamond textures that follow the tube geometry [14].

Motion is known to improve spatial understanding [34]. However, its effect on task completion time is unclear. Most previous studies examine the effects of texture and motion alone. Our study, instead, supplements these variables with the rendering model.

Increases in scene complexity could increase information access costs and thus degrade task performance. For example, Hubona et al. study depth visualization by having users perform two tasks related to positioning and sizing 3D objects in a scene under conditions of varying lighting, stereo or mono viewing, and background types [11]. They found that a stair-step background degraded task performance due to increased scene complexity, and attribute this to increased information access cost. In addition, scene understanding depends on the surrounding environment and the observer's viewpoint [11, 23, 26].

2.4 Comparison to Psychophysics Studies

Most related work comes from psychophysics or psychology literature that measures the use of individual dimensions of visual features, such as movement, shadow, shading, color, contrast, and brightness. However, trying to generalize these results for use in computer graphics and visualization could be problematic, because these features alone could be pre-attentive in simple settings, but be weakened by other cues [4, 32]. It is still uncertain how people use a combination of visual cues to perform visual tasks. In contrast, our study measures task performance in practical application settings.

2.5 Comparison to a Previous Study

The most recent study by Weigle and Banks [35] is in the same spirit as ours in that its main purpose was to compare GI and GL. However, there are several important differences between these two studies. First, the datasets differ: our scenes are taken from DTI subvolumes that have more curvature, while the previous study uses abstract tensor field data. Second, visual realism differs: current GI provides extrinsic shadows cast on the ground plane, but the version used in [35] includes only shadows between tubes. Third, independent variables differ: our experimental design includes texture and motion, as opposed to the static images in the previous study. Fourth, performance metrics differ: our metrics includes not only error rate but also task completion time and accuracy, which are important parameters for 3D visual analytic tasks. We also did not limit the task execution time because these tasks were not pre-attentive. For similar depth judgment tasks, participants in the current study spent

about four times longer than in [35]. Fifth and probably most important, the tasks differ: our tasks cover a richer set for DTI uses. Finally, our subject pool (26 subjects) is larger than theirs (5 subjects).

3 Visualization Synthesis

3.1 Geometry and texture

The basis of our task geometry was a high-resolution streamtube model generated from DTI data. A subvolume of this dataset has constant size and tube density, calculated heuristically based on the total number of triangles in this subvolume. The underlying geometry was used to generate texture coordinates to render a diamond pattern on the tubes for both GL and GI. The diamond-shape texture follows the tube geometry [14].

3.2 Illumination model

Two rendering algorithms are used for the study: GL and GI. Fixed-pipeline OpenGL-style is rendered with per-vertex lighting and Gouraud shading using Maya's hardware renderer. The GI rendering is produced with the Maya Mental Ray plugin. A Maya lighting rig was created based on a traditional three-point lighting scheme with several additional fill lights. Rim and key light placements were calculated according to a pre-set camera using a 35 mm focal length. The scene was lit from overhead, a built-in bias in the human visual system for looking at a scene [10]. Three million simulated photons were used in each image. The resolution of an image was 1600x1600, rendered with a perspective view.

In constructing the scene, we carefully chose light placement and intensity to generate images with contrast and lighting properties appropriate for the study. For example, shadows must be achieved by lighting the scene in a manner appropriate for the data, because GI does not have "perceptually good" shadows. The lighting process is particularly difficult due to the numerous parameters for all the renderer's components, each of which has a visual impact on the final image. Unfortunately, small changes in light placement and intensity can yield wildly differing images.

3.3 Motion

We employed passive rotation for the tube tracking task and rocking of the dataset for the depth and contact judgment tasks when motion is presented. Motion was synthesized using a sequence of 23 images, each image sequence accounting for ±5 degrees of rotation in Y (vertical in the image plane) from the axis-aligned DTI data. We did not allow free-form interaction because the high computational costs associated with GI prohibited real-time viewpoint-based rendering.

4 Experimental Design

4.1 Design

The primary purpose of this study was to explore the effects on task performance of illumination model, motion, and texture. Our hypothesis was that use of GI, texture, and motion would all improve task performance by reducing task completion time and error rates and improving accuracy. We used a 2x2x2x3 within-participant design with the following independent variables: illumination model (GL and GI), motion (presence or absence), texture (presence or absence), and task (depth judgment, visual tracing, and contact judgment). Each participant performed 48 tasks, 16 for each of the three tasks, following a Latin-square design. The order of the tasks was randomized. Task completion time, accuracy for the tube tracing, error rate, and subjects' comments were recorded.

TABLE 1
Task Types

Type	Example
Depth judgment	Which tube is closer, the blue or the green tube?
Visual tracing	Where is the endpoint of the tube?
Contact judgment	Does the blue sphere intersect, touch, or have no contact with the tubes?

4.2 Participants

Twenty-six volunteers (11 males and 15 females) participated in this study. All were Brown University undergraduate and graduate students. Two had prior exposure to medical imaging applications.

4.3 Tasks and user interface

Table 1 summarizes the three tasks used in our study. All tasks involve judgments related to the underlying geometrical tube structures drawn from the DTI data. All tasks were performed using a Dell 3007WFP monitor running at a resolution of 2560x1600.

Depth judgment task

Participants were shown a blue and a green tube embedded in a dataset (Fig. 2) and were asked to report which tube was closer to their viewpoint. The tubes were selected at random with the constraint that they not occlude each other from any sampled viewpoint. Internally, the distance between the two tubes was computed by sampling on a fixed interval across the dataset rotation range. We call this task a global visualization task because it may require visual scanning of neighboring tubes in order to make a correct judgment.

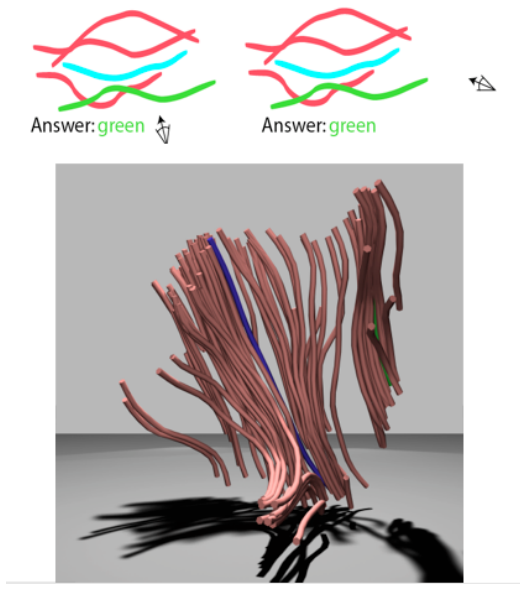

Fig. 2. Depth judgment task. Participants were asked to report which of the two tubes was closer to their viewpoint. The answer is green. The dataset rotates automatically when motion is present.

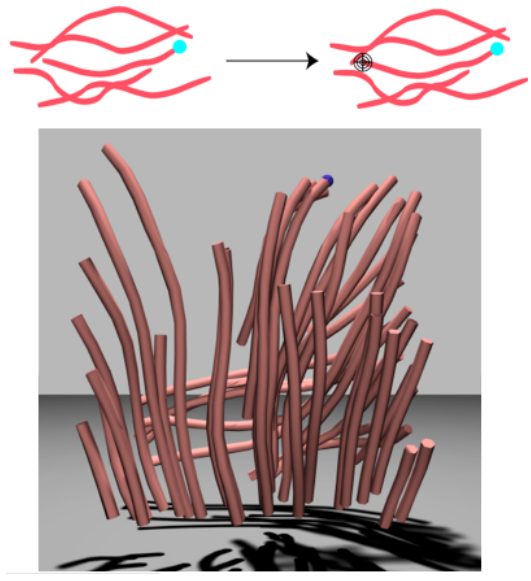

Fig. 3. Visual tracing task. Participants were asked to trace a tube and mark its endpoint. The user should mark the location at the red dot. When motion was enabled, participants could rotate the dataset using the keyboard.

Visual tracing task (following task)

This task required participants to trace a tube marked with a blue sphere on one end and to mark the other endpoint (Fig. 3). The tube was randomly selected. The constraint for the selection was that the two endpoints must be visible from all possible viewpoints. Once the unknown endpoint was found, the participant clicked on the projected point on the image plane (screen). As a result, task accuracy could be measured as the distance between the true location and the marked location on the image plane. This was the only task in which the user could rotate the dataset interactively using the keyboard. We did not use motion by rocking, because tracking a tube in a dynamic scene could be difficult and impractical. We call this task a global visualization task because it also requires visual scanning of neighboring tubes, similar to the depth judgment task.

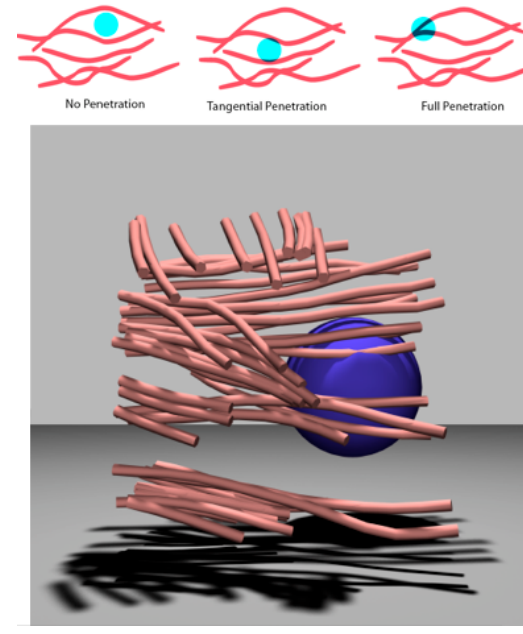

Fig. 4. Tumor task. Participants were asked to judge the relationship between the tumor (blue sphere) and its surrounding tubes. There are three possible answers: no penetration, tangential penetration and full penetration. The answer here is (3), full penetration.

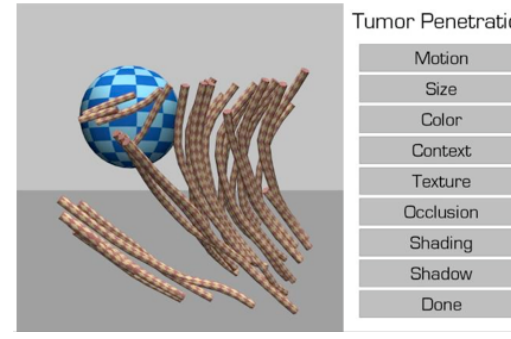

Fig. 5. User interface for cue choice. This UI was shown after each task to ask about the cues used in answering the task questions. This is a multiple-choice question.

Tumor task (contact judgment task)

Participants were asked to judge whether and how the blue sphere intersected the tubes. A blue sphere embedded in the dataset represents the location of a tumor. The judgment should be based on the closest distance between the two. There are three possible answers: (1) no penetration, (2) tangential penetration, and (3) full penetration (Fig. 4). Here no penetration means that tumor was free-floating, with no tubes touching it; tangential interaction means that the tumor grazed the tube(s) but did not fully intersect; and full intersection means that the tumor intersected the tube(s). The sphere was randomly placed in the dataset with equal distribution among the three cases. We call this a local visualization task because

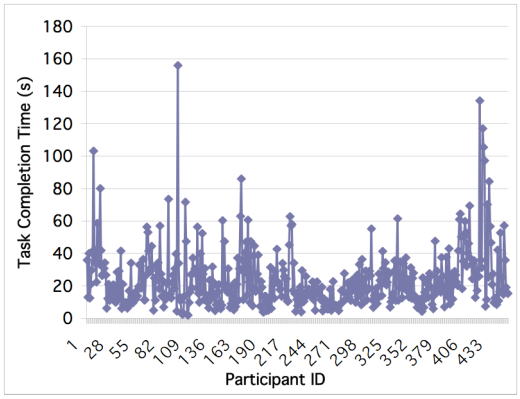

Fig. 6. Task completion time for all observations.

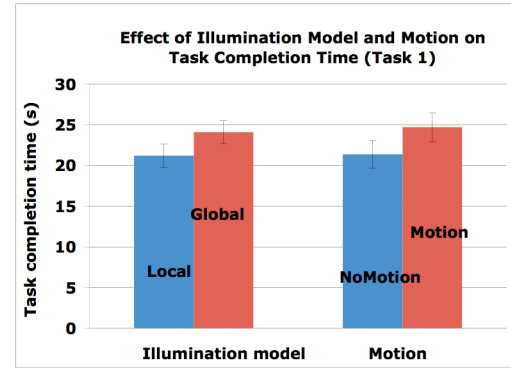

Fig. 7. Effect of illumination model and motion on task completion time for the depth judgment task. The error bar is one standard deviation. Both the illumination model and the motion had significant impact on task performance.

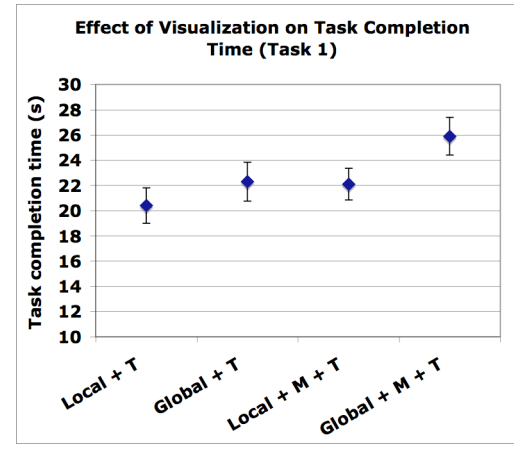

Fig. 8. Effect of visualization on task completion time. The HSD test showed that the "local + T" and "global + M + T" belong to different groups.

participants need to examine only the tubes around the tumor.

User interface

After each task, participants were required to identify the cues that were useful in performing the task (Fig. 5). The purpose of this step is to determine the usefulness of cues and to find correlations between the perceived usefulness and task completion time.

4.4 Procedure

The experiment included three sections. First, participants answered a questionnaire about previous exposure to medical imaging, art, and graphics. They were then guided through a training session on the task conditions, datasets, and user interface. They were provided with a training document that listed all cues and tasks, and were told how to examine the visual cues and were allowed to iterate as much as they liked. They were also asked to remember the cues and were told that this document would not be provided during the testing phase. The training datasets were generated in the same fashion as the actual study but with different data.

Next was the testing phase. Participants conducted the three tasks and answered the cue question after each task. Finally, subjects' responses were collected in a post-questionnaire relating the perceived usefulness of these techniques to DTI, the difficulty of the tasks, and the aesthetics of the rendering schemes. Subjects were also interviewed for additional comments.

RESULTS

5.1 Main effects

We performed four-factor ANOVA on illumination models (local GL and global GI), texture (presence and absence), motion (presence and absence), and task type (depth judgment, visual tracing, and tumor). We found a significant effect of motion on task completion time ($p<0.0001$, $F(15, 1184)=26.36$), with mean=26.04s with motion vs. mean=20.56s without motion. Another significant main effect was task type ($p=0.0005$, $F(15, 1184)=12.07$). There were strong two-way interactions between task type and motion ($p=0.0063$, $F(15,1184)=7.49$) and between task type and texture ($p=0.0032$, $F(15, 1184)=8.71$). No other two-way or three-way interaction was significant. We report our results by task type because it had a strong influence on task completion time.

5.2 Depth judgment task

Figure 6 shows performance data for all participants. We removed five outliers at 99% percentile, leaving 446 observations. The cutoff line was 103 seconds. Only one answer of the five outliers was correct. Texture was always present in this condition due to missing data in the experimental design (we had miscoded the case of no texture). As a result, no effect of texture can be measured. Correctness had no significant impact on task performance.

Figure 7 shows the main effects of motion and illumination model on task completion time. The illumination model had significant impact on task performance ($p=0.02$, $F(1, 377)=5.4$), with GL showing shorter task completion time (mean= 21.4s vs. 24.1s). Motion had strong impact on task performance ($p=0.038$, $F(1, 377)=4.75$), with no motion showing better task performance (mean=21.4s vs. 24.1s). Subject also had strong effect on task performance ($p<0.0001$, $F(24, 377)=7.33$).

The effects of the four combinations on task performance was borderline significant ($p=0.054$, $F(3, 442)=2.56$),

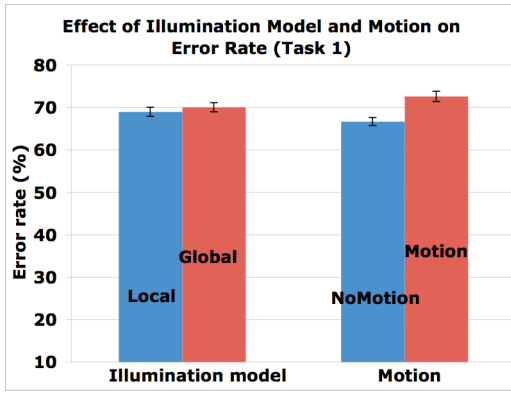

Fig. 9. Effect of illumination model and motion on error rate.

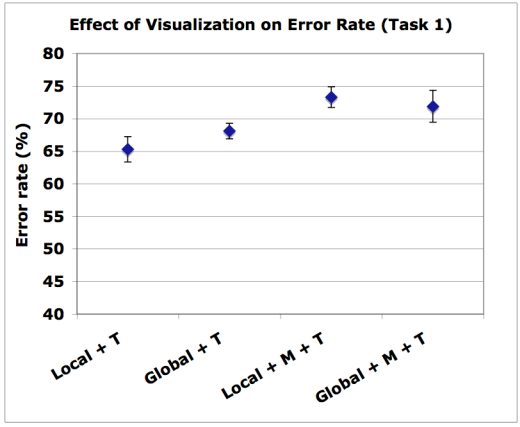

Fig. 10. Effect of visualization on error rate. The GL local illumination model without motion led to the lowest error rate.

by which we mean that the effect was significant if the threshold of the statistical α value = 0.1 but not if α = 0.05 (Fig. 8). The Tukey Studentized Range Test (HSD) procedure shows that the pair (Global + M + T) and (local + T) are in different groups.

The main effects of motion and illumination model on error rate are shown in Fig. 9. To our surprise, motion increased the error rate (72.6% vs. 66.7%). The global illumination model led to a slightly higher error rate (70.0% vs. 68.95%). The Pearson chi-square statistics on the motion and error rate suggest that there is no significant evidence of an association between motion and error rate (p=0.17). There is also no association between rendering style and error rate (p=0.8). The local illumination model when combined with motion and texture (local + M + T) led to the highest error rate (Fig. 10).

Participants' uses of a visual cue were measured using the percentage rate when the cue was present. 62.5% of participants made use of motion cues when there was motion and 32.7% participants used shadow cues with GI. The reported overall use of other cues ranked from high to low is: context 57.0%, occlusion 54.3%, shading 32.5%, color 17.1%, and size 14.8%.

5.3 Visual tracing task (tube following)

There is a strong effect of error rates on task performance ($p<0.0001$, $F(1, 249)=92.15$), with the shorter time for the correct answers (mean=22.4s vs. 42.13s). We therefore based the analysis on two groups: all observations and the correct observations. For both conditions, we removed outliers beyond the 99% percentile, five outliers for both conditions..

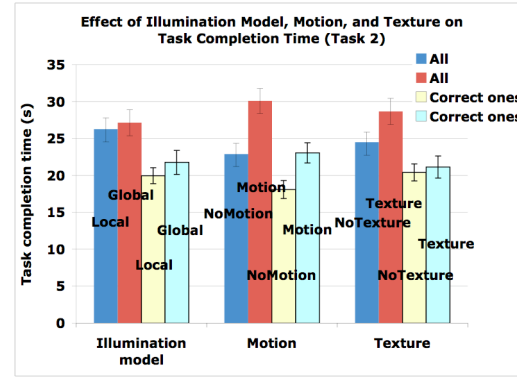

Fig. 11. Effect of illumination model, motion, and texture on task completion time with all and with correct observations. The error bar is one standard deviation. Motion had a significant effect on task performance. Texture was also significant with all observations, but not with only the correct ones. Illumination model was not significant.

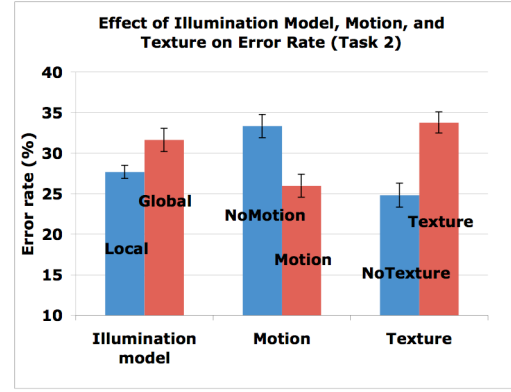

Fig. 12. Error rate for the visual tracing task.

The main effect of participants was significant ($p<0.0001$, $F(24, 249)=3.5$) (Fig. 11). There was also a significant main effect of motion ($p=0.002$, $F(1, 249)=9.41$), with no motion showing better task performance, mean = 22.9s vs. 30.1s. The HSD test for time showed they belonged to different groups. Texture was also significant ($p=0.017$, $F(1, 249)=5.79$), with no texture having better task performance (mean = 24.47s vs. 28.7s). The illumination model was not significant: GL showed slightly better task performance (mean=26.3s vs. 27.1s). When only correct answers were used for statistical measurement, texture became insignificant ($p=0.75$, $F(1, 154)=0.09$). Motion remained a significant main effect.

The effect of illumination model, texture, and motion on error rate is shown in Fig. 12. All main effects are significant, and GI, motion, and no texture led to a higher error rate.

The effects of visualization on task performance, taking into account all observations and all correct ones, are shown in Fig.s 13. We found that local illumination with-

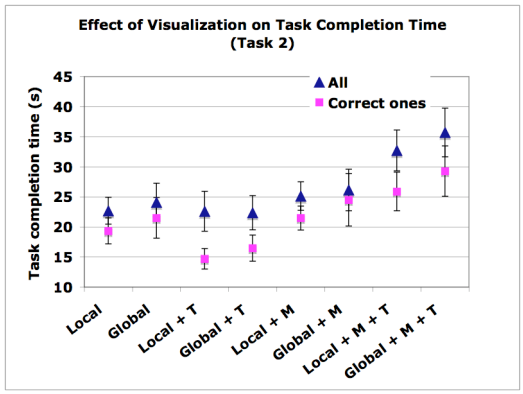

Fig.13. Effect of visualization on task completion time. The error bar is one standard deviation.

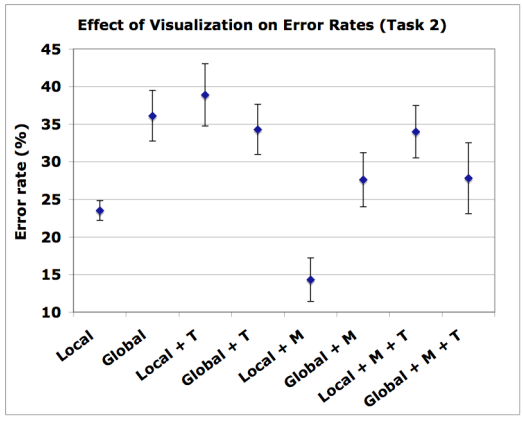

Fig. 14. Effect of visualization on error rates. The error bar is one standard deviation.

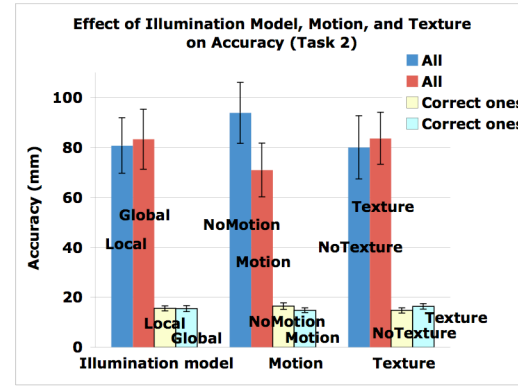

Fig. 15. Effect of illumination model, motion, and texture on accuracy. The error bar is one standard deviation.

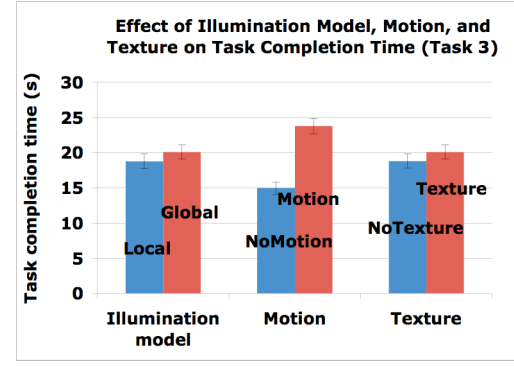

Fig. 16. Effect of illumination model, motion, and texture on task performance. The error bar is one standard deviation. Motion had significant impact on task performance; texture and illumination method did not.

out motion and with texture (Local + T) led to the shortest task completion time, followed by Global + T, though the HSD test did not find pairs in different groups. Local illumination with motion and without texture (Local + M) led to the lowest error rates, followed by local illumination without texture and motion (Local in Fig. 14).

Accuracy was measured by the distance between the pointer and the true target location on the image plane: the smaller the distance, the more accurate the subject for the task response. We found no significant main effect (motion, rendering style, or texture) on accuracy. However, there was a strong correlation between accuracy and task completion time ($r=0.25$, $p<0.0001$) for all observations: longer task completion time results in lower accuracy. This trend did not hold merely for the correct observations: there is no association between accuracy and time. Also, motion had significant impact on accuracy, leading to higher accuracy when all observations were used (Fig. 15); however, this trend became not significant when only correct observations were used.

All observations showed a strong association between the use of motion cues and the presentation of motion ($p<0.0001$), with 56.5% participants reporting that they used motion cues when available. Reported texture cue use also had a strong association with the presentation of texture, with 47.4% participants reporting used of texture when presented. The use of shadow cueing was low: only 4.41% used it when presented with global illumination. The other reported cue uses, from high to low, are context 67.8%, occlusion 50.5%, shading 43.4%, size 14.6%, and color 3%. The same trend is seen when all correct answers are used.

5.4 Contact judgment task (tumor task)

Six outliers were removed from the data, among which four gave the correct answers. Since the correctness of participants' answer did not have a significant impact on task performance ($p=0.18$, $F(1,447)=1.77$), we analyze the data based on all responses.

The only significant main effect was motion ($p<0.0001$, $F(1, 384)=47.22$): again, no motion led to better task performance (mean=15.0s vs. 23.8s) (Fig. 16). Illumination model (mean=18.7s vs. 20.0s) and texture (mean=18.80s vs. 20.08s with and without texture, respectively) were not significant.

We also measured the effect of visualization method on task performance (Fig. 18) and error rate (Fig. 19). The test cases were significant ($p<0.0001$, $F(7, 435)=8.21$). As expected, the full model (Global + M + T) had the lowest error rate, although task completion time was longer. A well-balanced visualization method was (Local + T), with shorter task completion time and relatively low error rate.

Participants were asked to rate on a post-questionnaire

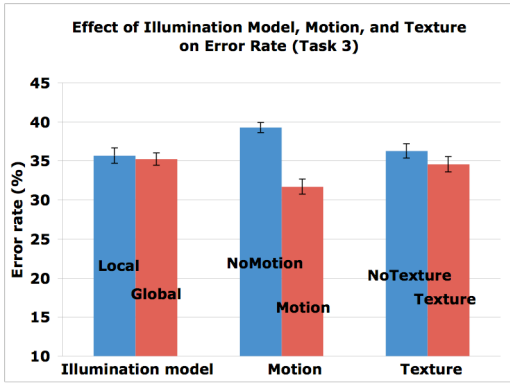

Fig. 17. Effect of illumination model, motion, and texture on error rate for contact judgment task. The error bar is one standard deviation.

the usefulness of the illumination model, aesthetics, and the difficulty of the tasks. A scale from 1 to 7 was used, 1 being the worst and 7 the best. The perceived usefulness of the image in DTI increased with the realism of the rendering, i.e., Local (3.27) < Local + Texture (3.78) < Global (4.29) < GIobal + Texture (5.27). The scores for aesthetic beauty were the reverse of the texture categories: Local + Texture (3.57) < Local (3.62) < Global + Texture (4.19) < Global (4.88). Participants rated the visual tracing task the most difficult (score = 5.11) and the contact judgment task the easiest (score=3.73). The difficulty rating for the depth judgment was in the middle range (score=4.77). Participants had a strong preference (5.52 on the 1-7 scale) for rotating the dataset (with motion) over using static frames (without motion).

6 Discussion

6.1 Effect of illumination model on performance metric is task dependent

The effect of the illumination model on task performance was task-dependent: global illumination led to longer task completion time for the depth judgment task, similar to the results in Weigle and Banks [35]. Our results show a significant main effect compared to their non-significant one probably because our study has more participants (26 vs. 5). The illumination model had no effect on the tube tracing and contact judgment tasks with regard to task completion time.

We had expected that GI would be important for the contact task because interreflection could suggest orientations and distances between the tumor and the tube [3]. However, this effect was not supported in our study, with GI showing a slightly lower error rate (Fig. 17). One possible explanation is that the tubes are so dense that subtle cues showing inter-object relationships disappear in the visualization. Our results suggest that a far less dense tube rendering is desirable in the contact judgment task to make it easier to detect the mapping between the tube and the shadow cues. Our results also suggest the need for careful task-oriented illumination models. A task-specific rendering could be more useful here than any general illumination approach.

GI had approximately the same error rate as GL in the depth judgment task, contrary to the results of Weigle and Banks's study. A possible explanation is that the size cue (the closer the tube, the larger the tube diameter) is more pronounced in their perspective viewing, Global illumination led to higher error rates for the tube tracing task. Since only 4.4% of the participants reported using shadow cues, we suspect that GI increased the visual complexity of the scene by adding extra cues, thus increasing task completion time. These results suggest that the visual complexity caused by interreflections and shadow cues could be large compared to GL.

In general, GI adds overall brightness to the image and hence reduces the effects of individual colors, while at the same time increasing the apparent contrast of the scene. Of the tasks for which people rarely use shadow cues, tasks that required higher contrast are likely to benefit from GI. Also, the density of the scene cannot be too high; adding a mask to increase local and global contrast might be useful.

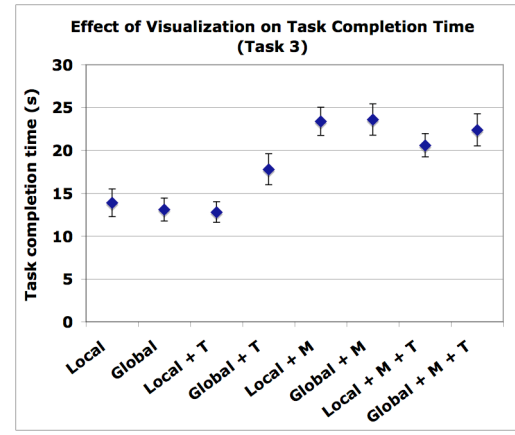

Fig. 18. Effect of visualization on task completion time for contact judgment task. The error bar is one standard deviation. (Local + T) showed the best results, and the results for global illumination were no better, even without texture.

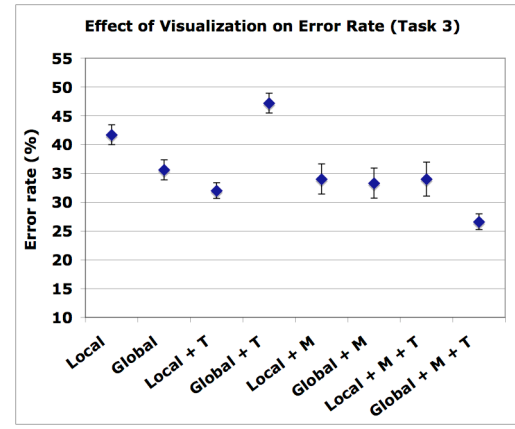

Fig. 19. Effect of visualization on error rate for contact judgment task. The full cue condition (Global + M + T) had the lowest error rate.

6.2 Motion increases task accuracy at the cost of longer task completion time

Motion had a negative effect on task completion time for

all tasks, even though it increased task accuracy for the contact judgment task. Motion improved task accuracy for the tumor task because that task required participants to make use of local visualization, so that less visual integration from multiple views is required. For the depth judgment and visual tracing tasks, motion had negative impact on task completion time. This result indicates that motion-introduced benefits come at a cost of cognitive workload for visual tracking. Participants had to understand the image from multiple views and had to track the views that did not represent continuous motion (±5 degrees).

A practical design suggestion to balance task completion time and accuracy is to make motion available at the beginning of a trial, then turn it off when subjects understand the layout. Ignoring motion completely to shorten task completion time might not be practical because the scene could appear flattened and two-dimensional [13].

6.3 Texture should reflect underlying task and data pattern

The presence of texture caused higher error rates in the depth and tube tracing tasks, but not in the tumor task. Again, we attribute this result to the difference between local and global tasks. Texture did not help the visual tracking task because when a DTI bundle has many tubes the entire scene is flattened to a single mass of diamond-textured shapes. Because of spatial masking from context, participants would have great difficulty in detecting a tube's direction when the context has similar frequency content. In addition, the diamond shape could have been poor at conveying orientation information because it included more than one edge orientation, making it hard for the human visual system to detect the continuity. A way to reduce this difficulty is to make the geometry long and thin, to distinguish the two textures by at least 30° [6]. Another way would be to orient the texture to near-horizontal or near-vertical lines [19] or to follow the principal direction.

Texture must be carefully generated in highly dense environments. Participants also reported confusion about spatial relationships due to texture. The texture design may also be task-dependent. For the occlusion task, a useful texture could be one that presents the relative depth of components. When textures were presented in the visual tracing and contact judgment tasks, subjects examined them in greater detail. Texturing methods such as dots [27], as proposed for the vascular structure, are worth exploring within the context of different rendering methods, as they have been proven to improve task performance.

Shading is useful in conveying shape information. Since the only shape here is the tube and its path, participants reported they used shading (32.5%, 43.4%, and 25%). We would expect shading or intrinsic shadow to be used for the tumor task because both convey shapes. We did not observe this effect, probably because the tumor had a large round shape that did not require detailed visual examination. Another explanation was that the scene complexity caused less clear interobject relationships.

6.4 Scene complexity has strong impact on stimuli effectiveness

As in Thompson et al. [30], scene complexity has probably introduced higher cognitive load. The geometry in [30] consisted of a ground plane and a box, and the depth cues afforded by GI were sufficient to identify ground plane contact. However, with the added complexity of streamtubes, this advantage is not as clear, even though most participants reported using context (57%, 67.8%, and 60.5% for the three tasks).

One solution to this problem might be to use semi-transparency, or show task-relevant tubes with clear conjunction to produce partial occlusion effects. Another solution is to design seeding strategies to distribute the streamlines maximizing task-related cues. Though numerous studies suggest algorithms for line distribution, whether they facilitate cue use beyond reducing density is unclear.

Our informal observations also suggest that the size of the display may increase task complexity. The tube-following task is simple when no occlusion occurs and when visual scan of the environment requires only foveal vision. In these conditions, eyeball movement is sufficient to accomplish the task. When head movement is required and the tube is long, the mental effort becomes much greater because of the low resolution of human peripheral vision. The task complexity can be very high when the number of tubes associated with the task-relevant one is more than human visual capacity (handling 3 to 4 items at once). Enhanced techniques would be useful in order to improve depth judgment.

6.5 Render only task-relevant information for effective visualization

Our results suggest that GL often shortens task execution time. It might thus be sufficient for most visualization tasks (Figs. 8, 13, 18), but it might also lead to a relatively high error rate (Figs. 9, 14, 19). It is worth noting that the full cue conditions (GI+T+M) led to the lowest error rate for the contact judgment task (Fig. 19), though the effect was not caused by illumination model.

6.6 Subjective choice of cues is relatively consistent

The order of reported cue use was consistent across task conditions, with the most used cues being context followed by occlusion. The least used cues were size and color. This agrees with the literature where occlusion is known to be the strongest visualization cue to represent depth information [34]. It was not surprising that participants did not make use of color because the hues were the same. Size cue was not particularly useful even though we used perspective viewing. This could be because the camera did not have a wide enough view to make the size cues pronounced.

6.7 Functional value and perceived value of visualization

Our results demonstrated no significant increase in task performance with the increase in visual realism (GI), despite the high rank of the perceived usefulness of realistic

rendering. This result suggests that visual design should differentiate between the *functional* and *perceived* values of visualization. A rendering that "looks good" might not be suitable for scientific visualization if it does not help users gain insight from the data. In contrast, renderings that "look good" are the goal of many graphics applications. It would be interesting to study some perceptual or attention-based rendering approaches to see if more insights can be obtained with lower error rates and shorter task completion times for scientific visualization tasks.

6.8 Lighting design

An astute participant who identified herself as a graphic designer said of the intersecting geometry rendered with GI: "Usually to show contact, you put a dark area under the part that touches", similar to the effects presented in "visual glue" [30]. The shadow in question was probably too dim to be useful as a cue due to an overuse of fill lights and ambient illumination calculated from GI, which is a byproduct of insufficient lighting design. Considerable pains are taken in the visual effects industry to design lighting that complements the complexity of the rendering algorithms, which suggests that visualization applications could benefit from further exploration of lighting paradigms.

In addition, computer graphics artists achieve "fake" lighting with shading operations in cases where the visual complexity and target exceed the capabilities of physically based rendering techniques. This type of methodology could be used to accentuate important aspects of the visualization data, as is commonly done in NPR treatments such as research on vascular structures [27]. Visualization methods that mix rendering approaches could provide an interesting balance by using physically based algorithms such as GI as a starting point and tailoring the output on the basis of prior knowledge of the data being visualized, as in NPR techniques.

6.9 Comments on our application-oriented study and cue-oriented psychophysical study method

There is a tradeoff between psychological studies on simple scenes and our studies using real applications. We evaluated scenes that resembled actual DTI datasets with large visual complexity. While psychological studies are good at isolating causes and effects, application-oriented studies allow us to understand the combinational effects of cues. We like to study multiple cue uses also because humans have a potential for visual gestalt, i.e., our visual perception system lets us develop an understanding of a whole that could be more than the sum of the parts. The visual gestalt could add a new twist to visualization design that makes integrated visual access appealing.

6.10 Limitations of our study

Our study is preliminary and has several limitations. We sought scenes from subvolumes of DTI data and thus did not explicitly control the tube densities. This caused difficulties in reporting details about what visualization conditions work the best. Also, placing the blue sphere randomly in a randomly selected tube subvolume for the contact judgment task may not make be a "good" task for testing visualization techniques, where we define "good" as sufficiently difficult (i.e., a clearly free-floating tumor is trivially easy) and sufficiently constrained (i.e., the tumor is not hidden by all the tubes). There are similar constraint issues with generating geometry for the other task types.

Conclusions

We have presented the results of a user study comparing task completion time and error rates for streamtube DTI rendering tasks using global illumination, texture, and motion. Major results of this study are the following:

- The effects of global illumination, motion, and texture on task completion time and error rate are task dependent. It is important to control scene complexity in order for the illumination model to be useful.
- Motion enhanced accuracy for tasks requiring longer execution times. Use motion with caution: it is important before participants understand the scene, but may not be needed afterward.
- Motion reduces error rates and increases accuracy for tasks that require local visual scans (contact judgment task). For tasks requiring global scans across the screen (depth judgment and tube tracing), motion led to higher error rates.
- When motion and texture are presented, people learn to use them. The use of shadow cueing is task dependent. The use of other visual cues such as context, shading, and size is independent of task conditions for the tasks we measured.
- For complex scenes, the GI with high visual fidelity does not lead to better orientation and depth judgment performance, suggesting that a less dense rendering algorithm or better depth representation is preferable.
- A range of visualization methods can offer similar task performance for certain tasks, so that designers can have a range of choices for their tasks at hand.
- When making design decisions, it is important to differentiate the functional and perceived value of visualization.

In conclusion, there are no black-box solutions in visualization, and any rendering technique must be modified depending on the data involved to maximize task performance. Studies such as this provide a mechanism for decomposing a visualization according the usefulness of individual cues.

Acknowledgment

The authors wish to thank the participants for their time and effort. This work was supported in part by a grant from NSF-CCF-1785542 and by Brown University Center for Vision Research Fellowship. The authors would also like to thank David Banks and Chris Wiegle for their discussion about the tasks and data analysis. Specifial thanks to Katrina Avery for her excellent editorial support.

REFERENCES

[1] D. Akers, A. Sherbondy, R. Mackenzie, R. Dougherty, & B. Wandell, "Exploration of the Brain's White Matter Pathways with Dynamic Queries", *Proceedings of the IEEE Visualization*, pp. 377-384, 2004.

[2] D.C. Banks, "Illumination in Diverse Codimensions", *SIGGRAPH*, pp. 327-334, 1994.

[3] D.C. Banks and C.-F. Westin, "Global Illumination of White Matter Fibers from Dt-Mri Data", *Visualization in Medicine and Life Sciene*: Springer 2007.

[4] J.R. Bergen and B. Julesz, "Parallel Versus Serial Processing in Rapid Pattern Discrimination", *Nature*, vol. 303(1983.

[5] A. Blake and H. Bulthoff, "Shape from Specularities: Computation and Psychophysics", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 331(1260), pp. 237-252, 1991.

[6] R. Blake and K. Holopigian, "Orientation Selectivity in Cats and Human Assessed by Masking", *Vision research*, vol. 25(10), pp. 1459-1467, 1985.

[7] L. da Vinci, *The Notebooks of Leonardo Da Vinci*. New York: Dover, 1970.

[8] D. Ebert and P. Rheingans, "Volume Illustration: Non-Photorealistic Rendering of Volume Models", *Proceedings of the IEEE Visualization*, pp. 195-202, 2000.

[9] J.A. Ferwerda, S.H. Westin, R.C. Smith, & R. Pawlicki, "Effects of Rendering on Shape Perception in Automobile Design", *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, pp. 107-114, 2004.

[10] J.J. Gibson, *The Ecological Approach to Visual Perception*, 1986.

[11] G.S. Hubona, P.N. Wheeler, G.W. Shirah, & M. Brandt, "The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3d Depth Visualization", *Transactions on Human-Computer Interaction*, vol. 6(3), pp. 214-242, 1999.

[12] V. Interrante, H. Fuchs, & S.M. Pizer, "Conveying the 3d Shape of Smoothly Curving Transparent Surfaces Via Texture", *IEEE Transactions on Visualization and Computer Graphics*, vol. 3(2), pp. 98-117, 1997.

[13] V. Interrante and C. Grosch, "Strategies for Effectively Visualizing 3d Flow with Volume Lic", *Proceedings of the IEEE Visualization*, pp. 421-424, 1997.

[14] D. Jianu, W. Zhou, C. Demiralp, & D.H. Laidlaw, "Visualizing Spatial Relations between 3d-Dti Integral Curves Using Texture Patterns", *Proceedings of the IEEE Visualization (poster compendium)*2007.

[15] J.T. Kajiya, "The Rendering Equation", *SIGGRAPH*, vol. 20(4), pp. 143-150, 1986.

[16] D.C. Knill, P. Mamassian, & D. Kersten, "Geometry of Shadows", *Journal of the Optical Society of America A*, vol. 14(12), pp. 3216-3232, 1997.

[17] C. Madison, W. Thompson, D. Kersten, P. Shirley, & B. Smits, "Use of Interreflection and Shadow for Surface Contact", *Perception and psychophysics*, vol. 63(2), pp. 187-194, 2001.

[18] O. Mallo, R. Peikert, C. Sigg, & F. Sadlo, "Illuminated Lines Revisited", *Proceedings of the IEEE Visualization*, pp. 19-26, 2005.

[19] R.J. Mansfield, "Neural Basis of Orientation Perception in Primate Vision", *Science*, vol. 186(1133-1135), 1974.

[20] Z. Melek, D. Mayerich, C. Yuksel, & J. Keyser, "Visualization of Fibrous and Thread-Like Data", *Transaction on Visualization and Computer Graphics*, vol. 12(5), pp. 1165-1172, 2006.

[21] B. Moberts, A. Vilanova, & J.J. van Wijk, "Evaluation of Fiber Clustering Methods for Diffusion Tensor Imaging", *IEEE Visualization*, pp. 65–72, 2005.

[22] R. Ni, M. Braunstein, & G. Andersen, "Perception of Scene Layout from Optical Contact, Shadows, and Motion", *Perception*, vol. 33(11), pp. 1305-1318, 2004.

[23] J.F. Norman, J.T. Todd, & F. Phillips, "The Perception of Surface Orientation from Multiple Sources of Optical Information", *PERCEPTION AND PSYCHOPHYSICS*, vol. 57(pp. 629-629, 1995.

[24] J. Obert, J. Krivanek, D. Sykora, & S. Pattanaik, "Interactive Light Transport Editing for Flexible Global Illumination", *Proceedings of the SIGGRAPH (Sketch)*2007.

[25] V. Petrovic, J. Fallon, & F. Kuester, "Visualizing Whole-Brain Dti Tractography with Gpu-Based Tuboids and Lod Management", *IEEE Transactions on visualization and computer graphics (also in IEEE visualization conference)*, vol. 13(6), pp. 1488-1495, 2007.

[26] V.S. Ramachandran, "Perception of Shape from Shading", *Nature*, vol. 331(6152), pp. 163-166, 1988.

[27] F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, & H.O. Peitgen, "Real-Time Illustration of Vascular Structures", *Transaction on Visualization and Computer Graphics*, pp. 877-884, 2006.

[28] J.C. Rodger, "Choosing Rendering Parameters for Effective Communication of 3d Shape", *Computer Graphcis and Applications*, pp. 20-28, 2000.

[29] S.A. Shafer and T. Kanade, "Using Shadows in Finding Surface Orientations", *Computer vision graphics image processing*, vol. 22(pp. 145-176, 1983.

[30] W.B. Thompson, P. Shirley, B. Smits, D.J. Kersten, & C. Madison, "Visual Glue", *University of Utah Technical Report UUCS-98-007, March*, vol. 12(1998.

[31] J.T. Todd and F.D. Reichel, "Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces", *Psychological Review*, vol. 96(4), pp. 643-657, 1989.

[32] A. Treisman and G. Gelade, "A Feature-Integration Theory of Attention", *Cognitive Psychology*, vol. 12(1), pp. 97-136, 1980.

[33] A. Vilanova, S. Zhang, G. Kindlmann, & D.H. Laidlaw, "An Introduction to Visualization of Diffusion Tensor Imaging and Its Applications", *Visualization and Processing of Tensor Fields*: Springer-Verlag, pp. 121–153, 2006.

[34] C. Ware, *Information Visualization: Perception for Design*: Morgan Kaufmann, 2004.

[35] C. Weigle and D.C. Banks, "A Comparison of the Perceptual Benefits of Linear Perspective and Physically-Based Illumination for Display of Dense 3d Streamtubes", *Transaction on Visualization and Computer Graphics*, vol. 14(6), pp. 1723-1730, 2008.

[36] C. Weigle and R.M. Taylor, "Visualizing Intersecting Surfaces with Nested-Surface Techniques", *IEEE Visualization*, pp. 503-510, 2005.

[37] A. Wenger, D. Keefe, S. Zhang, & D.H. Laidlaw, "Interactive Volume Rendering of Thin Thread Structures within Multivalued Scientific Datasets", *IEEE Transaction on Visualization and Computer Graphics*, vol. 10(6), pp. 664-672, 2004.

[38] S. Zhang, Ç. Demiralp, & D.H. Laidlaw, "Visualizing Diffusion Tensor Mr Images Using Streamtubes and Streamsurfaces", *IEEE Transaction on Visualization and Computer Graphics*, vol. 9(4), pp. 454-462, 2003.

[39] M. Zöckler, D. Stalling, & H.-C. Hege, "Interactive Visualization of 3d-Vector Fields Using Illuminated Stream Lines", *Proceedings of the IEEE Visualization*, pp. 434-445, 2004.

Devon Penney obtained the BS degree in computer science from Brown University. He is now with DreamWorks Animation, CA. His research centers lighting design, animation, and computer graphics.

Jian Chen obtained her PhD degree in computer science from Virginia Tech in 2006. She is currently a postodocal research associate in computer science at Brown University. Her research interests include human-centered computing, scientific visualization, computational modeling, and 3D interaction. Her paper on information visualization got a best paper award from HFES annual meeting in 2006. She had several other papers nominated for best paper awards. She has published more than 10 journal and conference papers. She is a member of IEEE, ACM and SIGCHI.

David H. Laidlaw received the PhD degree in computer science from the California Institute of Technology, where he also did postdoctoral work in the Division of Biology. He is a professor of computer science at Brown University, Providence, RI. His research centers applying concepts from visualization, modeling, computer graphics, and computer science to other scientific disciplines. He is a senior member of the IEEE and a member of the IEEE Computer Society.