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

Category: Research

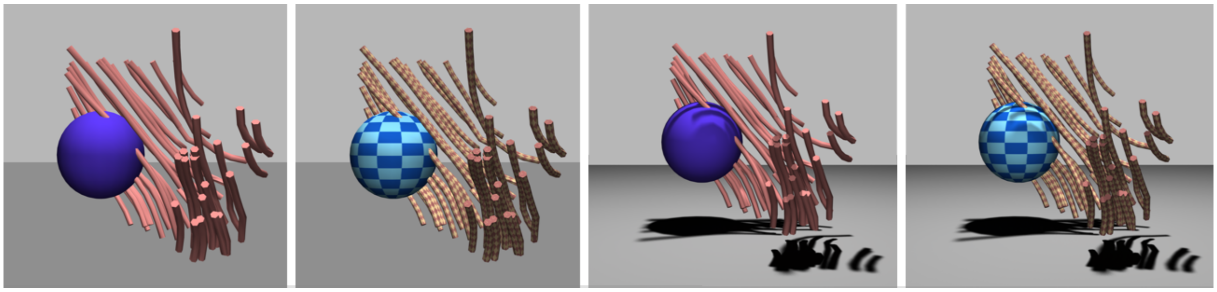


Fig. 1. The four rendering styles we studied, left to right: (1) OpenGL-style rendering; (2) OpenGL with texture; (3) global illumination rendering, and (4) global illumination rendering with texture. The dataset is a subvolume of a diffusion tensor magnetic resonance imaging (DTI) dataset. The sphere is a proxy for a tumor in one of the tasks.The scene is also one of the small datasets used in the study.

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 and OpenGL style local illumination), texture (with and without), motion (with and without), and task. 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 tasks we measured. Global illumination reduce the error rates of participants' answers over local OpenGL-style rendering when motion was present for the visual-tracing task. Motion contributes to spatial understanding for all tasks, but at the cost of longer task completion time. A high-frequency texture pattern led to longer task completion times and higher error rates. These results suggest that generic 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 it specific attention.

Index Terms—Flow visualization, human factors, multivariate visualization, performance evaluation, three-dimensional graphics and realism.

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 [32]. However, dense streamtubes suffer from visual cluttering, which impedes identification of specific tracts of interest and slows down user interaction [1, 32]. Accordingly, efforts have been made to improve spatial structure illumination and rendering by adding specular reflections [17], and increasing visual realism [24]. 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 [16, 21, 25], it is unclear how such illumination models affect task performance, and, particularly, how combined cues depict 3D scenes effectively. Exploring this question will help us design visualization methods that integrate only important cues in the rendering algorithms, thus reducing computational cost while maximizing user performance.

The goal of this work is to explore the effect of illumination models, motion, and texture on DTI tasks. We compare two illumination models (Fig. 1): global illumination (GI) and an OpenGL-based local illumination model (GL). GI simulates the complex behaviors of light around the model being viewed and can create realistic renderings [3]. GL is a baseline method for comparison purpose. Here we also measure motion and texture, because these two are believed to convey information effectively [33]. By combining these factors, results from our study can provide a picture of how real-world visualizations work under various display conditions. 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 corresponds to a real-world DTI task such as analyzing a tube among DTI bundles or evaluating a tumor. Independent variables include illumination model (GI and GL), motion (with and without), texture (with and without), and task. completion time, accuracy, and subjective responses. We hypothesize that GI, motion, and texture can reduce task completion time and error rates because these factors support better depth and shape cues, thus aid DTI data understanding.

Our results indicated 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]. Our results also revealed that motion improves accuracy and lowers error rates, but at the cost of longer task completion time. Scene complexity had strong impact on task performance expect for the task completion time in the contact-judgment task conditions. Participants reported that GI increased the perceived quality of a visual scene, suggesting that *looking good* (perceived) and *being (measurably) good* (functional) are different. Visual designers must differentiate between the perceived and functional value of visualization.

Our work has several key contributions. 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 empirically demonstrate that 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 diffusion rate of water molecules in tissue. 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 many individual fibers from the tensor information using streamline algorithms [39]. These streamlines or streamtubes are initiated at 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: decluttering by clustering fiber bundles, and conveying spatial structures via perceptual principles. Clustering reduces 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, 40, 19]. While Moberts et al. [20] 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 render each polygon. Ambient lighting effects are indicated by a global color shift. The other common rendering technique is GI, which involves simulating the behavior of light throughout the scene in order to increase visual realism [14]. For example, a photon-mapping algorithm can produce inter-reflections and soft shadows [3].

Numerous evidence suggests that cues generated from GI allow more accurate discrimination of shapes by revealing spatial structure and orientation among surfaces [4, 8, 9, 28, 35]. For example, Weigle and Banks [35] 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 advances the previous study by using experimental settings that are closer to real-world uses of DTI when motion and texture are presented. The scenes are taken from DTI subvolumes with three-level visual complexities; These subvolumes produce tubes with more curvature than the tubes used in [35] thus were more visually challenging. In addition, more tasks are used in the current study which required both local and global shape understanding.

2.2.1 Shadow and inter-reflection

Two types of shadows, extrinsic and intrinsic [15], can be generated via illumination models. Intrinsic shadows, also called shading, are the shadows an object makes on itself Fig. 1. They have 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 [15]. Such convex and concave shapes are also much more recognizable by human observers [26, 22]. Extrinsic shadows (Figs. 1, 3, 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, 31]. GL and GI can both produce shading effect. Only GI can produce extrinsic shadows between tubes.

It is generally agreed that shadows increase reported task accuracy [31, 36]. For example, Thompson et al. analyzed how shadow and inter-reflection affect depth perception [31]. They found that both inter-reflections 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. However, Hubona et al. found the effects of shadow were task-dependent. Shadows enhanced the accuracy but not the speed of object positioning; they did not enhance either the accuracy or the speed of object resizing [10]. Shadows are also subtle enough to distort the perception of 3D shapes [22].

Another difference between GI and GL is that GI implementations support inter-reflection. 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 film making [23]. While many studies have focused on algorithm design, our goal here is to understand the impact of illumination method on task performance. Results from our work can guide visual designers towards important design features.

2.2.2 Texture, motion, and scene complexity

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]. We use results from Jianu et al. to construct diamond textures that follow the tube geometry [13].

Motion is known to be a powerful cue to improve spatial understanding [33]. Sollenberger and Milgram [30] demonstrated the utility of motion parallax in visualizing complex simulated blood vessel structures in the brain. Ware [34] found that motion produced by head-tracking had effects as powerful as stereo viewing on task completion times. Moreover, they reported that these effects were more powerful than stereoscopic viewing on reducing error rate. Related to our study, we also wish to understand how the effects of motion and illumination model are combined.

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 [10]. They found that a stair-step background degraded task performance due to increased scene complexity, and attribute this to increased information access cost. Norman suggested such detrimental effects were caused by distortion from the saddle-shaped background surfaces [22].

2.3 Cue integration

Numerous studies have examined the perception of visual cues. An important question is whether these cues are perceived as separate elements, or as belonging to an integrated, unitary whole object. Knowing so will provide practical design implications for DTI imaging, especially related to multi-variant data mapping. At least three models on non-conflicting cue integration have been proposed, including the additive and multiplicative models, and weak fusion and strong fusion.

The *additive theory* [6] suggests that cues are combined in a weighted additive fashion. The combined effects of multiple depth cues can be determined as the weighted linear summation of the information provided by each individual cue. In contrast, the *multiplicative model* [30] suggests more complicated combinations. For example, multiple cues can interact and foster a "greater-than" or "less-than" additive effect. The weak fusion model [38], similar to the weighted additive model, suggests that the cues are weighted and "additively combined" by a factor reflecting their apparent reliability in that particular visual context; the strong fusion model suggests that cues can interact non-linearly. For example, one cue may work to disambiguate information, thus enabling correct information to be derived from other cues. Weigle and Banks found participants improved task performance especially when GI was combined with perspective viewings [35]. Here, we make use of motion and texture to study if similar effects can be obtained.

3 Experimental Design

The primary purpose of this study was to explore the effects of illumination model, motion, and texture on task performance. Our hypothesis was that use of GI, texture, and motion would 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 and absence), texture (presence and absence), and task (depth-judgment, visual-tracing, and contact-judgment). In each condition, scene complexity was controlled by the number of tubes a visualization contained. Each participant performed 48 tasks, 16 instances of each of the three tasks. The order of the tasks was randomized. Task completion time, accuracy (for the visual-tracing task only), error rate, and participants' comments were recorded.

3.1 Visualization synthesis

3.1.1 tubes, texture, and scene complexity

Visualizations were generated by collecting the tubes contained when intersecting a randomly placed, fixed-size box with the full streamtube model generated from a whole brain dataset. The bounding volume of the full model was 217.6 mm x 217.6 mm x 153 mm. The edge lengths of the boxes used in this study were 8.8 mm, 14.08 mm, and 24.64 mm, each containing 30-50, 100-150, and 300-560 numbers of tubes accordingly. By doing so, we created three levels of scene complexity, either small, medium, or large. UV texture coordinates were generated for the geometry so we could apply diamond textures at render time.

3.1.2 Illumination model

We used two rendering algorithms for this study: GL and GI. For GL we use Maya's hardware renderer to implement fixed-pipeline OpenGL style rendering with per-vertex lighting and Gouraud shading. For GI, we use Maya's Mental Ray plugin with three million simulated photons per image. The resolution of an image was 1600x1600, rendered in perspective.

We used a traditional three-point lighting scheme plus several fill lights. Rim and key lights were placed in relation to a preset camera that used a 35 mm focal length. 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 components, each of which has a visual impact on the final image. Unfortunately, small changes in light placement and intensity can yield very different images.

3.1.3 Motion

Motion was synthesized using a sequence of 23 images, each image sequence accounting for ± 5 degrees of rotation in the local coordinate of the dataset about the vertical direction in the image plane. We did not allow free-form interaction because the high computational costs associated with GI prohibited real-time viewpoint-based rendering.

3.2 Tasks and user interface

Three tasks are 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 screen pixels.

- Depth judgment, e.g., which tube is closer, the blue or the green tube?
- Visual tracing, e.g., where is the endpoint of the tube?
- Contact judgment, e.g., does the blue sphere intersect, touch, or have no contact with the tubes?

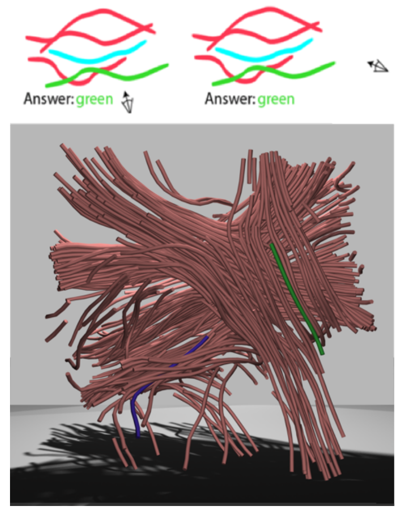


Fig. 2. Depth-judgment task. Participants were asked to report which of the two tubes was closer to their viewpoint. The answer here is green. The dataset rotates automatically when motion is present.

Each task included two (for the visual-tracing task) or three (for the depth-judgment and the contact-judgment task) levels of scene complexity. We did not include all three levels for the visual-tracing task because this task took longest task completion time and could potentially introduce fatigue.

The depth-judgment task includes two Each visual condition also includes two to three levels of scene complexity controlled by the number of tubes ranged from 30 to 560. Dependent variables include task

3.2.1 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 a global visualization task because it may require visual scanning of neighboring tubes in order to make a correct judgment.

3.2.2 Visual-tracing task

This task required participants to trace a randomly selected tube marked with a blue sphere on one end and to mark the other endpoint (Fig. 3). 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 tracing a tube in a dynamic scene could be difficult and impractical. We call this a global visualization task because it also requires visual scanning of neighboring tubes, similar to the depth-judgment task.

3.2.3 Contact-judgment task

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

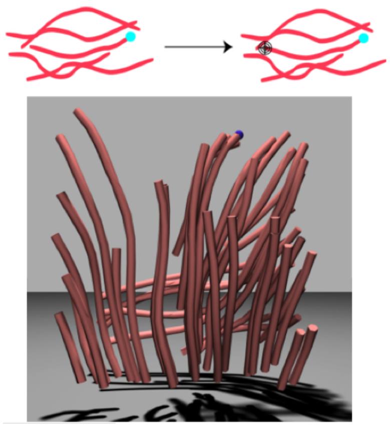


Fig. 3. Visual-tracing task. Participants were asked to trace a tube and mark its endpoint. When motion was enabled, participants could rotate the dataset using arrow keys on the keyboard.

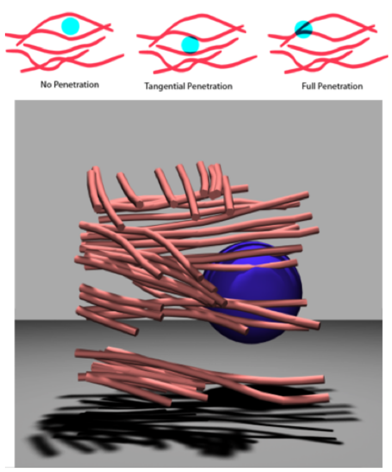


Fig. 4. Contact-judgment task. Participants were asked to judge the relationship between the blue sphere and its surrounding tubes. There are three possible answers: no penetration, tangential penetration and full penetration. The answer here is (3), full penetration.

among the three cases. We call this a local visualization task because participants need to examine only the tubes around the sphere.

3.2.4 User interface

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

3.3 Participants

Twenty-six volunteers (11 males and 15 females) participated in this study. All were xxx *(masked for blind review)* undergraduate and graduate students. Two had prior exposure to medical imaging applications.

3.4 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, were told how to examine the visual cues, and were allowed to iterate until they were comfortable performing the tasks. They were also asked to remember the cues and were told that this document would not be provided

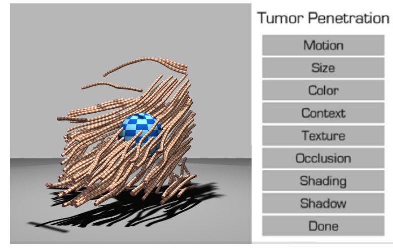


Fig. 5. User interface for cue choice. This display was shown after each task to ask about the cues used in answering the task questions. This is a multiple-choice question.

Table 1. Summary of the F and p Values for the Main Effects on Task Completion Time. "*" indicates the significant main effects.

Task	Illumination	Motion	Texture	Participant	Complexity
Depth judgment	$p = 0.75$ $F = 0.1$	$p<0.0001$* $F = 18.5$	– –	$p<0.0001$* $F = 7.1$	p=0.002* $F = 6.2$
Visual tracing (all)	p=0.13 $F = 3.3$	p=0.003* $F = 8.8$	p=0.005* $F = 7.9$	$p<0.0001$* $F = 18.4$	$p<0.0001$* $F = 90.1$
Visual tracing (correct only)	$p = 0.54$ $F = 0.4$	$p<0.0001$* $F = 17.0$	p=0.12 $F = 2.5$	p=0.002* $F = 2.2$	$p<0.0001$* $F = 59.5$
Contact judgment	$p = 0.5$ $F = 0.5$	$p<0.0001$* $F = 22.8$	$p = 0.51$ $F = 0.4$	$p<0.0001$* $F = 6.4$	$p = 0.34$ $F = 1.1$

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 two sequences of tasks with a short break between them. Each sequence was composed of 24 tasks from each of the three types. A total of 48 unique tasks were displayed in a different order for each participant. There was a total of 12 follow tasks, 16 depth judgment tasks, and 16 tumor tasks. The cue question was answered after each task. Participants were told to finish the task as quickly and accurately as they could. Finally, participants' 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.

4 Results

4.1 Summary statistics and statistical method

We performed within-subject GLM (general linear model) procedure using SAS (statistical analysis software) on illumination models (GL and GI), texture (presence and absence), motion (presence and absence), and task type (depth-judgment, visual-tracing, and contact-judgment), participants, and scene complexity. We found a significant main effect of task type ($p = 0.0005$, $F(15,1184) = 12.07$). We thus report our results based on task type. For each task, we conducted statistical analysis on all answers. In the case when the correctness had significant impact on task completion time, we also did a second analysis on the correct-only answers. When there was a significant main effect, we separated the levels to study the causes of effects. We also combined perceptual cues to learn the accumulated effects. When combined, the three levels of the scene complexity were not separated because real-world applications could include all of them. We also removed outliers at 99% percentile. All error bars represent 95% confidence intervals. The summary of the F and p values are shown in Tables 1 and 2.

4.2 Depth-judgment task

We removed five outliers at 99% percentile, leaving 446 observations. Texture was always present in this condition due to missing data in the experimental design (we had miscoded the case of no texture). As

Table 2. Summary of the F and p values for the main effect on error rate. "*" indicates significant main effect.

Task	Illumination	Motion	Texture	Participant	Complexity
Depth judgment	$p = 0.9$ $F = 0.001$	$p = 0.08$ $F = 3.0$	– –	p=0.9 $F = 0.4$	p<0.0001* $F = 30.9$
Visual tracing (all)	$p = 0.04*$ $F = 4.3$	$p = 0.03*$ $F = 10.2$	$p = 0.07$ $F = 3.34$	$p = 0.9$ $F = 0.43$	p<0.0001* $F = 60.3$
Contact judgment	$p = 0.96$ $F = 0.00$	$p = 0.07$ $F = 3.36$	$p = 0.81$ $F = 0.05$	p=0.002* $F = 2.19$	p<0.0001* $F = 59.5$

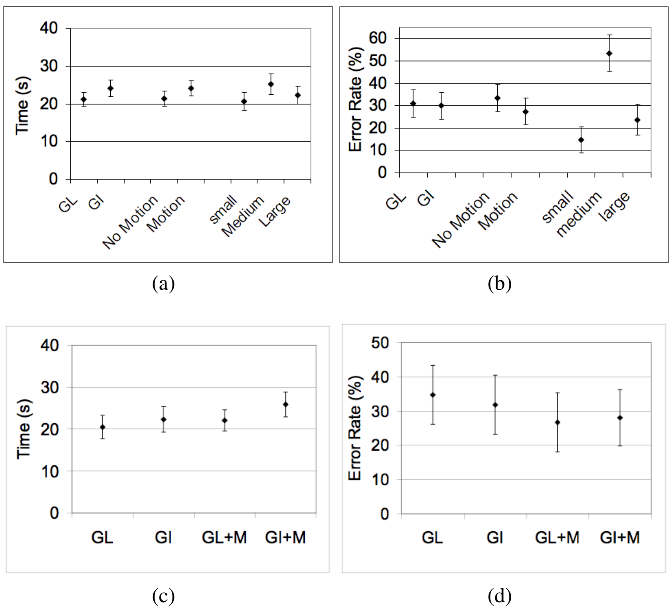


Fig. 6. Depth-judgment task: Effect of illumination model, motion, and scene complexity on task performance. (a) Motion and scene complexity had a significant impact on task completion time. (b) Motion and scene complexity had a significant impact on error rate. (c) Combined effect: GL+M and GL were in different groups. (d) Combined effect: All conditions were in the same group.

a result, no effect of texture can be measured. Correctness had no significant impact on task completion time.

Fig. 6(a) shows the main effects of motion, illumination model on task completion time. The summary statistics including the the effects of participants and scene complexity is shown in Table 1. For the effects on task completion time, motion was significant with motion increasing mean task completion time from 21.4s to 24.09s. Scene complexity was also significant, with the mid-range complexity led to the longest task completion time. Participants were significant. Tukey's Studentized Range (HSD) test for illumination model revealed that GL and GI were in different groups. By separating the presence and absence of motion condition, we found the differences between GL and GI occurred when motion was presented ($F = 4.63$, $p < 0.0001$) with GI having longer task completion time (25.9s vs. 22.1s). When the effects were combined to form the 4 conditions 6(c), GL and GI+M are the only pair in different groups.

The main effects of motion and illumination model on error rate are shown in Fig. 6(b). The summary statistics is shown in Table 2. The only significant effect was scene complexity. As expected, the higher the complexity the higher the error rate. The smallest box size led to 14% error rate compared to 24.8% and 52% for middle and large size boxes. When the effects were combined, all four perceptual cues were in the same group 6(d).

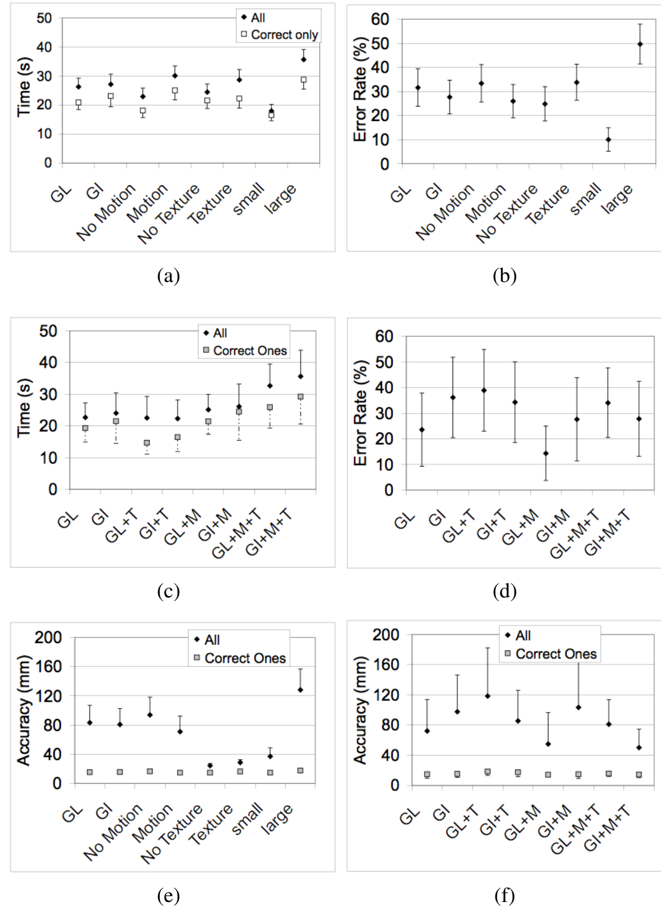


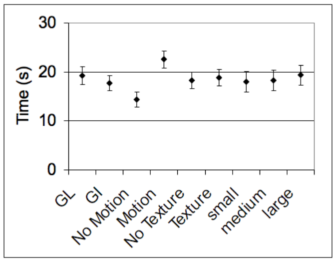
Fig. 7. Visual-tracing task: Effect of illumination model, motion, texture, and scene complexity on task performance. (a) Motion, texture, and scene complexity had a significant impact on time when all observations were used; Motion and scene complexity were significant when the correct-only answers were used. (b) Illumination model, model and scene complexity had a significant impact on effor rate. (c) Combined effect: GI+M+T was in different groups from GL, GI, GL+T, and GI+T accordingly when all answers were used. GL+M+T and GI+M+T were in different groups from GI+T and GL+M when the correct-only answers were used. (d) Combined effect:GL+M led to the lowest error rate. (e) Only scene complexity was significant. (f) GI+M+T led to the most accurate answers especially when all observations were used.

4.3 Visual-tracing task

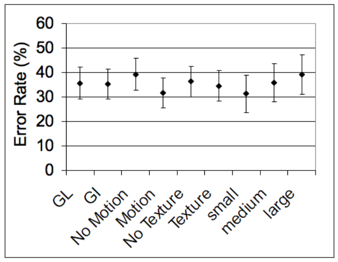
Correct answers were made significantly faster than incorrect ones (mean=22.4s vs. 42.13s, $p < 0.0001$, $F(1,249) = 92.15$). We therefore separately analyze all observations and correct observations. We removed all five outliers beyond the 99% percentile in each case.

The effects of illumination model, motion, texture, scene complexity, and participant on task performance are shown in Fig. 7. The F and p values are shown in Table. 1 When all answers were analyzed, there was a significant main effect of motion, which increased mean task time from 22.9s to 30.1s (Fig. 7(a)). Texture had significant impact on task completion time, with an increase from 24.47s vs. 28.7s. Scene complexity and participants was also significant. There was no significant change of the task completion time for the correct-only answers for texture. The illumination model was not significant; GL showed slightly shorter task completion time (mean=26.3s vs. 27.1s).

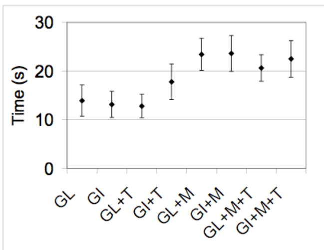
The main effects on error rate are shown in Fig. 7(b). The F and p values are shown in Table. 2. Illumination model, motion, and scene complexity had significant impact on error rate. We again found that the significant effect occured when motion was presented, when we



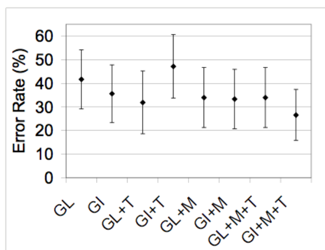
(a)



(b)



(c)



(d)

Fig. 8. Contact-judgment task: Effect of illumination model, motion, texture, and scene complexity on task performance. (a),(b) motion and scene complexity had a significant impact on task completion time and error rate. (c) Combined effect: GI+M+T led to relatively longer task completion time. (d) Combined effect: GI+M+T led to the lowest error rate.

separated the presence and absence of motion. Interesting, participants were not significant.

We measured the combined effect (Fig. 7(b) and Fig. 7(d)). HSD test on task completion time suggested that GI+M+T was in different groups from GL, GI, GL+T, and GI+T accordingly, when all observations were used; GL+M+T and GI+M+T were in different groups from GI+T and GL+M when correct-only observations were used. GL+T led to the shortest task completion time, followed by GI+T (Fig. 7(c)). All eight conditions were in the same group for the error rate, though GL+M led to the lowest error rates(Fig. 7(d)).

Accuracy was measured as the distance between the pointer and the true target location on the image plane: the smaller the distance, the more accurate the participant was for the task response. We found both scene complexity and participants were significant ($p < 0.0001$ for both cases). No other significant main effect (illumination model, motion, or texture) on accuracy was observed. There was also a strong correlation between accuracy and task completion time ($r = 0.25$, $p < 0.0001$) for all observations: participants who were more accurate had shorter task completion time. This trend did not hold for the correct observations: there was no correlation between accuracy and time. Also, motion had significant impact on accuracy, leading to higher accuracy when all observations were used (Fig. 7(e)); however, this trend disappeared when only correct observations were used. The full cue condition led to the longest task execution time, but most accurate answers (Fig. 7(f)).

4.4 Contact-judgment task

Six outliers were removed from the data. Since the correctness of participants' answer did not have a significant correlation with task completion time ($p = 0.18$, $F(1,447) = 1.77$), we analyze the data based on all responses (Fig. 8). One significant main effect was motion ($p < 0.0001$, $F(1,384) = 47.22$); again, no motion led to shorter task completion time (mean=15.0s vs. 23.8s) (Fig. 8(a)) and lower error rate (Fig. 8(b)). The other significant effect was participants. However, HSD test suggested all participants belonged to the same group. The scene complexity did not have an impact on task completion time. None of the main effects was significant on the error rate.

We also measured the combined effect of illumination model, motion, and texture on task completion time (Fig. 8(c)) and error rate (Fig. 8(d)). HSD test suggested that GL+M, GI+M, GL+M+T, and GI+M+T are in different groups from GL, GI and GL+T accordingly. the full-cue condition (GI+M+T) had the lowest error rate, although task completion time was longer. An intermediate visualization method was (GL+T), with intermediate task completion time and intermediate error rate.

4.5 Subjective comments

Participants were asked to rate the usefulness of the illumination model, aesthetics, and the difficulty of the tasks on a post questionnaire. 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., GL (3.27) $<$ GL + Texture (3.78) $<$ GI (4.29) $<$ GI + Texture (5.27). The scores for aesthetic beauty were the reverse of the texture categories: GL + Texture (3.57) $<$ Local (3.62) $<$ GI + Texture (4.19) $<$ GI (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).

5 Discussion

Our data did not agree with the expectations that GI, motion, texture would improve task completion time. Our data did support that motion was a strong cue to reduce the error rates and improve the accuracy of visual tracing. Interesting, the analysis of the accumulated cues conditions suggested that full cue conditions, though led to longer task completion times, often led to the lowest error rates.

5.1 Illumination model effects are task dependent

No effect of illumination model on task completion time was observed. GL and GI were different only when motion was presented for the depth-judgment and the visual-tracing tasks. We suspect that the increase of task completion time by GI was caused by the need to mentally process more visual cues. The significant effect on error rate was observed only for the visual-tracing task. And the significance again occurred when motion was presented. This result could suggest that GI and motion when used together could led to better judgment. In general, GI slightly reduced error rates compared to GL, but at the expense of longer task completion time.

These results only partially support our hypothesis that GI would improve task performance. Due to the small magnitude of the reduced error rate, increased response time trade-offs, and the rendering cost for producing GI, it is problematic to assert that GI assisted visualization for the tasks under study. We expected that GI would more be important for the contact-judgment task because inter-reflection could suggest orientations and distances between the sphere and the tube [3]. However, this effect was not supported in our study.

Theoretical studies on the processing of cues of visual graphics images may explain this result. The result for the contact task can be supported by the weak fusion mechanism for cue combination [38]. When cues are presented, participants evaluate each independently. Separated cues are weighted and combined reflecting their "apparent reliability" in a particular visual context. According to this theory, participants would spend time evaluating the rich cues provided by GI, which may have caused the longer task completion times.

One possible explanation for the lack of significant results on the error rate is that the tubes were dense enough to cause inter-object relationship ambiguity. It was difficult for participants to find which tube casted which shadow and where, especially when large numbers of tubes were presented. The shadows casted on the ground in the GI conditions help recognize large-scale structures such as bundles. However, shadows did not support visualization at a fine level examining individual tubes. Another explanation is that GL shading used in our experiment produced salient depth effects which might be sufficient to help participants perform tasks. The image contrast for the visualizations produced using GL and GI was similar thus did not lend advantages to GI, which could produce higher contrast images than GI. Finally, the added scene complexity of streamtubes blurred the benefit of these cues, even though most participants reported using context cues (57%, 67.8%, and 60.5% for the three task conditions).

5.2 Motion increases accuracy but lengthens time

Our hypothesis that motion would decrease error rate is supported by the results at the cost of lengthened task completion time. We think the time cost with motion is mostly due to the motor actions for changing views. The benefits of motion likely come at the cost of cognitive workload for visually processing more imagery. Participants had to understand the image from multiple views, and had to track views that did not represent continuous motion due to the ± 5 degrees gap between neighboring images.

5.3 Texture

The presence of texture caused higher error rates in the visual-tracing tasks but not in the contact-judgment tasks. We attribute this result to the difference between local and global tasks. Texture did not help the visual-tracing task because when a DTI bundle has many tubes, the entire scene is reduced 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, which causes spatial masking. In addition, the diamond shape could have been poor at conveying orientation because it included more than one edge orientation, making it hard for the human visual system to detect continuity. A way to reduce this difficulty is to make the geometry long and thin, to distinguish the two textures by at least 30 degrees [5]. Another way would be to orient the texture to near-horizontal or near-vertical lines [18] or to follow the principal curvature directions.

Participants also reported confusion about spatial relationships due to texture. Thus, textures must be carefully generated in highly dense environments, and can 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, participants 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 present is the tube and its implicit path, participants reported they used shading (32.5%, 43.4%, and 25% for depth judgment, visual tracking and contact judgments. We would expect shading or shading to be used for the contact-judgment task because both convey relational properties. We did not observe this effect, probably because the sphere had a large round shape that did not require detailed visual examination. Another explanation was that the scene complexity caused less clear inter-object relationships.

5.4 Scene complexity impacts performance for global tasks

Scene complexity is measured by the size of the bounding volume used in the visualization. Three levels were used. Interestingly, the scene complexity had significant impact on both task completion time and error rate for the depth-judgment and the visual-tracing tasks, but not for the contact-judgment task. An intuitive explanation is that contact-judgment was a local task that only requires the viewing of the information directly surrounding the sphere.

Besides scene complexity, our informal observations also suggest that task complexity may have increased as a function of a larger field of view introduced by the display hardware. For example, the visual-tracing task is simple when no occlusion occurs, and when the trace of the tube requires only foveal vision. In these conditions, eyeball movement is sufficient to accomplish the task. However, head movement would be required when the display becomes larger, and the tube crosses a large portion of the screen. The efforts of changing gaze could introduce higher mental and cognitive load, and thus introduce time penalties. The error rates was especially high for the visual-tracing task. The large scene complexity however had lower error rate compared to the medium size ones for the depth-judgment task. This is partially because the distances between the tubes were further thus making size cues more salient in the large conditions.

5.5 Participants had similar error rates

Reducing the error rate is one of the most desirable characteristics for visualization design. Interestingly, HSD test on participants suggested they appear in the same group for all task conditions. This consistency might suggest that illumination model, motion, and texture were not conflicting cues, and participants could reach a reasonable answer if provided appropriate viewing time.

5.6 Differences among tasks

The visual cues had a larger impact on task performance and error rate for global tasks. Our results suggested that special attention needs to be paid to these kinds of tasks compared with local ones. A better design for the contact-judgment task is to explicitly control the location of the spheres. This will allow us to produce non-biased results on what illumination cases work and why not. Another better control of our design is to have a visualization that is sufficiently difficult (i.e., a clearly free-floating sphere is trivially easy) and sufficiently constrained (i.e., the sphere is not hidden by all the tubes). We found that task generation algorithms that afford such results are difficult to implement effectively.

5.7 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.

5.8 Lighting design

Lighting design is critical to user performance. 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" [31]. 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 by-product of insufficient lighting design. Differences in lighting could explain some of the difference between our work and previous studies. It is also likely that more attention to lighting is likely to help in a number of visualization applications.

6 Conclusion

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:

- Numerous psychophysical studies support that GI can aid 3D visualization by providing shadow and inter-reflection to recover 3D form 2D images on the screen. Our study suggested that GI provided such benefits in a limited fashion with visually complex visualizations. Considering the computational cost in read-time rendering, GI could be detrimental to real-world task compared to other stronger 3D cues such as motion.

- 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 participants understand the layout. Ignoring motion completely to shorten task completion time might not be practical because the scene could appear flattened and two-dimensional [11].

- Special care must be taken to identify whether tasks require local or global visual scanning. This determination can aid the design of visualization components, such as texture and illumination paradigms.
- Our results suggest that, in general, the simpler stimuli (ie, those with GL, no motion, and no texture) usually speed task execution, although they also lead to reduced accuracy in some cases. Where more complex stimuli are not needed to increase accuracy, they should be eschewed.
- Motion and scene complexity are dominant over the illumination model or texture. Adding ineffective texture causes task completion time to degrade. The data is consistent with the weak fusion theory.
- Scene complexity needs to be further studied to provide insight into the effective design of visual algorithms that would reduce scene complexity through methods such as interactive filtering.

In conclusion, scene complexity does not directly benefit from global illumination. Care needs to be taken to properly utilize such techniques in order to maximize visualization quality through careful lighting design and rendering parameter exploration. In addition, the usage of visualization components such as illumination model, motion and texture must be designed to correlate with the task's associated global or local scope.

REFERENCES

[1] D. Akers, A. Sherbondy, R. Mackenzie, R. Dougherty, and B. Wandell. Exploration of the brain's white matter pathways with dynamic queries. In *IEEE Visualization*, pages 377–384, 2004.

[2] D. C. Banks. Illumination in diverse codimensions. *SIGGRAPH*, pages 327–334, 1994.

[3] D. C. Banks and C.-F. Westin. Global illumination of white matter fibers from dt-mri data. In *Visualization in Medicine and life sciene*. Springer, 2007.

[4] A. Blake and H. Bulthoff. Shape from specularities: Computation and psychophysics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 331(1260):237–252, 1991.

[5] R. Blake and K. Holopigian. Orientation selectivity in cats and human assessed by masking. *Vision research*, 25(10):1459–1467, 1985.

[6] N. Bruno and J. E. Cutting. Minimodularity and perception of layout.

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

[8] J. A. Ferwerda, S. H. Westin, R. C. Smith, and R. Pawlicki. Effects of rendering on shape perception in automobile design. In *1st Symposium on Applied perception in graphics and visualization*, pages 107–114, 2004.

[9] J. J. Gibson. *The ecological approach to visual perception*. 1986.

[10] G. S. Hubona, P. N. Wheeler, G. W. Shirah, and M. Brandt. The relative contributions of stereo, lighting, and background scenes in promoting 3d depth visualization. *Transactions on Human-Computer Interaction*, 6(3):214–242, 1999.

[11] V. Interrante, H. Fuchs, and S. M. Pizer. Conveying the 3d shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on visualization and computer graphics*, 3(2):98–117, 1997.

[12] V. Interrante and C. Grosch. Strategies for effectively visualizing 3d flow with volume lic. In *IEEE Visualization*, pages 421–424, 1997.

[13] D. Jianu, W. Zhou, C. Demiralp, and D. H. Laidlaw. Visualizing spatial relations between 3d-dti integral curves using texture patterns. In *IEEE Visualization (poster compendium)*, 2007.

[14] J. T. Kajiya. The rendering equation. *SIGGRAPH*, 20(4):143–150, 1986.

[15] D. C. Knill, P. Mamassian, and D. Kersten. Geometry of shadows. *Journal of the Optical Society of America (A)*, 14(12):3216–3232, 1997.

[16] C. Madison, W. Thompson, D. Kersten, P. Shirley, and B. Smits. Use of interreflection and shadow for surface contact. *Perception AND Psychophysics*, 63(2):187–194, 2001.

[17] O. Mallo, R. Peikert, C. Sigg, and F. Sadlo. Illuminated lines revisited. In *IEEE Visualization*, pages 19–26, 2005.

[18] R. Mansfield. Neural basis of orientation perception in primate vision. *Science*, 186(1133-1135), 1974.

[19] Z. Melek, D. Mayerich, C. Yuksel, and J. Keyser. Visualization of fibrous and thread-like data. *Transaction on Visualization and Computer Graphics*, 12(5):1165–1172, 2006.

[20] B. Moberts, A. Vilanova, and J. J. van Wijk. Evaluation of fiber clustering methods for diffusion tensor imaging. *IEEE Visualization*, pages 65–72, 2005.

[21] R. Ni, M. Braunstein, and G. Andersen. Perception of scene layout from optical contact, shadows, and motion. *Perception*, 33(11):1305–1318, 2004.

[22] L. Y.-L. P. F. N. H. F. J. L. R. M. T. R. Norman, J. F.

[23] J. Obert, J. Krivanek, D. Sykora, and S. Pattanaik. Interactive light transport editing for flexible global illumination. In *SIGGRAPH (Sketch)*, 2007.

[24] V. Petrovic, J. Fallon, and 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)*, 13(6):1488–1495, 2007.

[25] V. S. Ramachandran. Perception of shape from shading. *Nature*, 331(6152):163–166, 1988.

[26] K. J. J. Richards, W. A. and D. D. Hoffman. Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America*, A(4):1168–1176, 1987.

[27] F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, and H. O. Peitgen. Real-time illustration of vascular structures. *Transaction on Visualization and Computer Graphics*, pages 877–884, 2006.

[28] J. C. Rodger. Choosing rendering parameters for effective communication of 3d shape. *Computer Graphcis and Applications*, pages 20–28, 2000.

[29] S. A. Shafer and T. Kanade. Using shadows in finding surface orientations. *Computer vision graphics image processing*, 22:145–176, 1983.

[30] R. L. Sollenberger and P. Milgram. Effects of stereoscopic and rotational display of a three-dimensional path-tracing task. *Human Factors*, 35(18):483–499, 1993.

[31] W. B. Thompson, P. Shirley, B. Smits, D. J. Kersten, and C. Madison. Visual glue. *University of Utah Technical Report UUCS-98-007, March*, 12, 1998.

[32] A. Vilanova, S. Zhang, G. Kindlmann, and D. H. Laidlaw. An introduction to visualization of diffusion tensor imaging and its applications. In *Visualization and Processing of Tensor Fields*, pages 121–153. Springer-Verlag, 2006.

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

[34] C. Ware and G. Franck. Evaluating stereo and motion cues for visualizing information nets in three dimensions.

[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*, 14(6):1723–1730, 2008.

[36] C. Weigle and R. Taylor. Visualizing intersecting surfaces with nested-surface techniques. *IEEE Visualization*, pages 503–510, 2005.

[37] A. Wenger, D. Keefe, S. Zhang, and D. H. Laidlaw. Interactive volume rendering of thin thread structures within multivalued scientific datasets. *IEEE Transaction on Visualization and Computer Graphics*, 10(6):664–672, 2004.

[38] L. M. S. Young, M. J. and L. T. Maloney. A perturbation analysis of depth perception from combinations of texture and motion cues. *Visualization Research*, 33(18):2685–2696, 1993.

[39] S. Zhang, C. Demiralp, and D. H. Laidlaw. Visualizing diffusion tensor mr images using streamtubes and streamsurfaces. *IEEE Transaction on Visualization and Computer Graphics*, 9(4):454–462, 2003.

[40] M. Zockler, D. Stalling, and H.-C. Hege. Interactive visualization of 3d-vector fields using illuminated stream lines. In *IEEE Visualization*, pages 434–445, 2004.